Universidade de Aveiro Departamento de Biologia
2011

Sónia Cristina Marques Pascoal

Nucella lapillus: análise do transcriptoma e plasticidade fenotípica

Nucella lapillus: imposex transcriptome analysis and phenotypic plasticity

Universidade de Aveiro Departamento de Biologia
2011

Sónia Cristina Marques Pascoal

Nucella lapillus: análise do transcriptoma e plasticidade fenotípica

Nucella lapillus: imposex transcriptome analysis and phenotypic plasticity

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Biologia, realizada sob a orientação científica da Doutora Sónia Alexandra Leite Velho Mendo Barroso, Professora Auxiliar do Departamento de Biologia da Universidade de Aveiro e do Professor Doutor Gary Robert Carvalho, Professor catedrático do Departamento de Ciências Biológicas da Universidade de Bangor, Reino Unido.

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio através da bolsa SFRH/BD/27711/2006.

o júri

presidente

Prof. Doutor José Joaquim de Almeida Grácio
professor catedrático do Departamento de Engenharia Mecânica da Universidade de Aveiro

Prof. Doutor Octávio Fernando de Sousa Salgueiro Godinho Paulo
professor auxiliar do Departamento de Biologia Animal/ Centro de Biologia Ambiental da Faculdade de Ciências da Universidade de Lisboa

Doutora Susana Galante Correia Pinto de Oliveira
investigadora em Pós-Doutoramento do CESAM, CESAM e Departamento de Biologia da Universidade de Aveiro

Prof. Doutor Amadeu Mortágua Soares
professor catedrático do Departamento de Biologia da Universidade de Aveiro

Doutor Simon Creer
senior research fellow da School of Biological Sciences, Bangor University, Reino Unido

Prof. Doutor Carlos Miguel Miguez Barroso
professor auxiliar do Departamento de Biologia da Universidade de Aveiro

Prof. Doutora Sónia Alexandra Leite Velho Mendo Barroso (orientadora)
professora auxiliar do Departamento de Biologia da Universidade de Aveiro

Prof. Doutor Gary Robert Carvalho (co-orientador)
professor catedrático da School of Biological Sciences, Bangor University, Reino Unido

agradecimentos/ Acknowledgements

I would like to express my gratitude to all the ones that made this journey possible and to Fundação para a Ciência e a Tecnologia (FCT) and Natural Environment Research Council (NERC) for financial support.

I am very grateful for the opportunity of being the link between three excellent institutions that indeed, provided me all the facilities, support, guidance, knowledge and friendship during the last four years. It was a big privilege to belong to such a group of amazing people. My deepest thanks to all of them:

To my supervisors and co-supervisors: Prof. Sonia Mendo, Prof. Gary Carvalho, Prof. Roger Hughes, Prof. Carlos Barroso and Dr. Si Creer;

To my collaborators from the Institute of integrative biology and Centre for genomic research from University of Liverpool: Prof. Andrew Cossins, Dr. Kevin Ashelford; Dr. Yongxiang Fang; Dr. Lisa Olohan and Dr. Olga Vasieva;

To my collaborators from Bangor University: Dr. Delphine Lallias, Dr. Frederico Batista, Dr. Jenny Rock and Dr. Niklas Tysklind;

To my Molecular Ecology and Fisheries Genetics Lab (MEFGL) group of colleagues and friends from Bangor University;

To my Molecular Biotechnology Lab (LBM) friends and to Dr. Susana Galante-Oliveira and Isabel Oliveira from the Marine and Estuarine Ecology Lab (LEME) from University of Aveiro;

To research support staff from School of Biological Sciences and School of Ocean Sciences from Bangor University.

My thanks go also to all my Good friends from Bangor and Aveiro that were always there for me, and to my volleyball team for the great moments of happiness; you made it all much easier.

Finally, but most importantly, I would like to thank my family, specially my parents, my brother and sister for their unconditional support, understanding and affection; I wouldn't be where I am today without you.

Statement of collaborative work

The present Ph.D. thesis was financially supported by FCT (Ph.D. grant SFRH/BD/27711/2006) and by NERC-Molecular Genetics Facility Liverpool (grant 294). Since this grant included staff support, I would like to acknowledge the long-term support of NERC for the NERC Biomolecular Analysis Facility, in particular my collaborators that have contributed to the development of crucial steps of the work:

Professor Andrew Cossins for valuable discussions and for the opportunity to develop collaborative work with his group (Centre for genomic research, University of Liverpool);

Dr. Kevin Ashelford for writing the Perl scripts used for the transcriptome analysis;

Dr. Lisa Olohan for running eArray and for hybridizing the experimental samples on the array;

Dr. Yongxiang Fang for conducting the array data analysis;

Dr. Olga Vasieva for contributing for the network/pathway analysis and biological interpretation of the array data.

palavras-chave

Alteração do meio ambiente, disrupção endócrina, *imposex*, *Nucella lapillus*, tributilestanho, pirosequenciação (454 Roche), transcriptoma, microarray, expressão génica, receptores nucleares, factores de transcrição, “RXR:PPARγ heterodimer”, plasticidade fenotípica, adaptação, morfologia da concha, morfometria, exposição à força das ondas, predação.

resumo

O conhecimento de mecanismos de genómica funcional tem sido maioritariamente adquirido pela utilização de organismos modelo que são mantidos em condições laboratoriais. Contudo, estes organismos não reflectem as respostas a alterações ambientais. Por outro lado, várias espécies, ecologicamente bem estudadas, reflectem bem as interacções entre genes e ambiente mas que, das quais não existem recursos genéticos disponíveis. O *imposex*, caracterizado pela superimposição de caracteres sexuais masculinos em fêmeas, é induzido pelo tributilestanho (TBT) e trifenilestanho (TPT) e representa um dos melhores exemplos de disrupção endócrina com causas antropogénicas no ambiente aquático. Com o intuito de elucidar as bases moleculares deste fenómeno, procedeu-se à combinação das metodologias de pirosequenciação (sequenciação 454 da Roche) e microarrays (Agilent 4*180K) de forma a contribuir para um melhor conhecimento desta interacção gene-ambiente no gastrópode *Nucella lapillus*, uma espécie sentinela para *imposex*. O trancriptoma de *N. lapillus* foi sequenciado, reconstruído e anotado e posteriormente utilizado para a produção de um “array” de nucleótidos. Este array foi então utilizado para explorar níveis de expressão génica em resposta à contaminação por TBT. Os resultados obtidos confirmaram as hipóteses anteriormente propostas (esteróidica, neuroendócrina, retinóica) e adicionalmente revelou a existência de potenciais novos mecanismos envolvidos no fenómeno *imposex*. Evidência para alvos moleculares de disrupção endócrina não relacionados com funções reprodutoras, tais como, sistema imunitário, apoptose e supressores de tumores, foram identificados. Apesar disso, tendo em conta a forte componente reprodutiva do *imposex*, esta componente funcional foi a mais explorada. Assim, factores de transcrição e receptores nucleares lipofílicos, funções mitocondriais e actividade de transporte celular envolvidos na diferenciação de géneros estão na base de potenciais novos mecanismos associados ao *imposex* em *N. lapillus*. Em particular, foi identificado como estando sobre-expresso, um possível homólogo do receptor nuclear “peroxisome proliferator-activated receptor gamma” (PPARγ), cuja função na indução de *imposex* foi confirmada experimentalmente *in vivo* após injecção dos animais com Rosiglitazone, um conhecido ligando de PPARγ em vertebrados. De uma forma geral, os resultados obtidos mostram que o fenómeno *imposex* é um mecanismo complexo, que possivelmente envolve a cascata de sinalização envolvendo o receptor retinoid X (RXR):PPARγ “heterodimer” que, até à data não foi descrito em invertebrados. Adicionalmente, os resultados obtidos apontam para alguma conservação de mecanismos de acção envolvidos na disrupção endócrina em invertebrados e vertebrados.

Resumo (cont)

Finalmente, a informação molecular produzida e as ferramentas moleculares desenvolvidas contribuem de forma significativa para um melhor conhecimento do fenómeno *imposex* e constituem importantes recursos para a continuação da investigação deste fenómeno e, adicionalmente, poderão vir a ser aplicadas no estudo de outras respostas a alterações ambientais usando *N. lapillus* como organismo modelo.

Neste sentido, *N. lapillus* foi também utilizada para explorar a adaptação na morfologia da concha em resposta a alterações naturais induzidas por acção das ondas e pelo risco de predação por caranguejos. O contributo da componente genética, plástica e da sua interacção para a expressão fenotípica é crucial para compreender a evolução de caracteres adaptativos a ambientes heterogéneos. A contribuição destes factores na morfologia da concha de *N. lapillus* foi explorada recorrendo a transplantes recíprocos e experiências laboratoriais em ambiente comum (com e sem influência de predação) e complementada com análises genéticas, utilizando juvenis provenientes de locais representativos de costas expostas e abrigadas da acção das ondas. As populações estudadas são diferentes geneticamente mas possuem o mesmo cariótipo. Adicionalmente, análises morfométricas revelaram plasticidade da morfologia da concha em ambas as direcções dos transplantes recíprocos e também a retenção parcial, em ambiente comum, da forma da concha nos indivíduos da $F_{2,}$ indicando uma correlação positiva (co-gradiente) entre heritabilidade e plasticidade. A presença de estímulos de predação por caranguejos estimulou a produção de conchas com labros mais grossos, de forma mais evidente em animais recolhidos de costas expostas e também provocou alterações na forma da concha em animais desta proveniência. Estes dados sugerem contra-gradiente em alterações provocadas por predação na morfologia da concha, na produção de labros mais grossos e em níveis de crescimento.

O estudo das interacções gene-ambiente descritas acima demonstram a actual possibilidade de produzir recursos e conhecimento genómico numa espécie bem caracterizada ecologicamente mas com limitada informação genómica. Estes recursos permitem um maior conhecimento biológico desta espécie e abrirão novas oportunidades de investigação, que até aqui seriam impossíveis de abordar.

keywords

Environmental change, endocrine disruption, imposex, *Nucella lapillus*, tributyltin, 454 Roche sequencing, transcriptome, microarray, gene expression, nuclear receptors, transcription factors, RXR:PPARγ heterodimer, phenotypic plasticity, adaptation, shell morphology, geometric morphometrics, wave exposure, crab predation.

abstract

Our understanding of functional genomic mechanisms is largely acquired from model organisms through laboratory conditions of exposure. Yet, these laboratory models typically have little environmental relevance. Conversely, there are numerous “ecological” model species that present important gene-environment interactions, but lack genomic resources. Imposex, the superimposition of male sexual characteristics in females, is caused by tributyltin (TBT) and triphenyltin (TPT) and provides among the most widely cited ecological examples of anthropogenically-induced endocrine disruption in aquatic ecosystems. To further elucidate the functional genomic basis of imposex, combinations of 454 Roche pyrosequencing and microarray technologies (Agilent 4*180K) were employed to elucidate the nature and extent of gene-environment interactions in the prosobranch gastropod, *Nucella lapillus*, a recognized sentinel for TBT-induced imposex. Following transcriptome characterization (*de novo* sequencing, assembly and annotation), microarray fabrication and competitive hybridizations, differential gene expression analyses provided support for previously suggested hypotheses underpinning imposex (steroid, neuroendocrine, retinoid), but also revealed potential new mechanisms. Evidence for endocrine disruption (ED) targets such as the immune system, apoptosis and tumour suppressors other than reproduction-related functions were found; however, given the ED nature of imposex, primary focus was on gender-differentiation pathways. Among these, transcription factors and lipophilic nuclear receptors as transducers of TBT toxicity along with mitochondrial functions and deregulation in transport activity suggested new putative mechanisms for the TBT-induced imposex in *N. lapillus*. Particularly, up-regulation of a putative nuclear receptor peroxisome proliferator-activated receptor gamma (PPARγ) homolog was evident, and its role was further confirmed by inducing imposex *in vivo* using Rosiglizatone, a well-known vertebrate PPARγ ligand. Our analyses show that TBT-induced imposex is a complex mechanism, but is likely to act through the retinoid X receptor (RXR):PPARγ heterodimer signalling pathway, hitherto not described in invertebrates. Moreover, collectively, our findings support a commonality of signalling between invertebrate and vertebrate species that has previously been overlooked in the study of endocrine disruption. The genomic resources generated here largely contribute to the molecular understanding of imposex, yielding valuable insights for further examination of responses to TBT contamination exposure. Additionally, we anticipate that the new genomic resources described herein will contribute to the further exploration of adaptive responses of dogwhelks to environmental variation.

Abstract (cont)

N. lapillus was also used to explore adaptive shell shape morphology in response to natural variation in wave-action and crab predation. Knowledge of the contributions of genotype, plasticity and their interaction to phenotypic expression is crucial for understanding the evolution of adaptive character traits in heterogeneous environments. We assessed contributions of the above factors by reciprocal transplantation of snails between two shores differing in exposure to wave action and predation, and rearing snails of the same provenance in a laboratory common garden experiment with crab-predation odour, complemented by genetic analysis. The two target populations are genetically different but maintain the same karyotype. Truss-length and morphometric analyses revealed plasticity of shell shape in reciprocal transplants, but also the partial retention of parental shape by F_2 snails in common garden controls, indicating co-gradient variation between heritable and plasticity components. Crab-predation odour influenced shell shape of snails from exposed-site origin and stimulated the production of thicker shell lips with greater response in snails of exposed-site ancestry. We interpret these data as countergradient variation on predator-induced changes in shell shape and increased thickening of the shell lip as well as on growth rates.
The above exploration of gene-environment interactions demonstrates the feasibility, insights and novel opportunities that can now be addressed in a species that is well characterised ecologically, but hitherto constrained by the general lack of genomic tools and archived resources. Notably, a greater focus on detailed responses of a single species facilitates the comparative approach, as illustrated by the apparent commonality in regulation of endocrine disruption processes in invertebrates and vertebrates.

Table of contents

Chapter 3. Endocrine disruption in the marine gastropod Nucella lapillus: combination of pyrosequencing and microarray technologies as a diagnostic instrument

Chapter 4. Induction of imposex in Nucella lapillus by Rosiglitazone: evidence for a potential involvement of the RXR:PPARγ signalling pathway

Chapter 5. Genetic differentiation and plasticity underlying adaptive shell variation in Nucella lapillus: an assessment using reciprocal transplant and common garden experiments

Chapter 6. General discussion

Supplementary information

List of Figures

Chapter 1. General Introduction

Chapter 6. General discussion

List of Tables

Table 3. Reciprocal transplant experiment. Cross-validated percentage group membership predicted by discriminant function analysis of RW scores. Symbols as in Table 2.

Table 4. Microsatellite analysis: genetic variability measures by locus for each population. Na: number of alleles found per locus; H_E: expected heterozygosity; H_O: observed heterozygosity; F_{IS}: standardised genetic variance within populations at each locus; F_{ST}: standardized genetic variance among populations at each locus; HWE: Hardy-Weinberg P values.

List of abbreviations

9CRA: 9-cis-retinoic acid
%I: Percentage of imposex affected females
AF: Antifouling
Ah: Aryl hydrocarbon
AhR: Aryl hydrocarbon receptor
ANOVA: Analysis of variance
ANCOVA: Analysis of covariance
ARNT: AhR nuclear translocator
ATAT: Acyl coenzyme A: testosterone acyltransferase
ATP: Adenosine Triphosphate
BLAST: Basic Local Alignment Search Tool
bp: Base pair
cDNA: Complementary deoxyribonucleic acid
COI: Cytochrome Oxidase Subunit I
cRNA: Complementary RNA
Cyp: Cytochrome P450
CPA: Cyproterone acetate
CTAB: Hexadecyltrimethylammonium Bromide
DBT: Dibutyltin
DE: Differential Expressed
dNTPs: Deoxynucleoside triphosphate
DNA: Deoxyribonucleic acid
DR: Direct Repeats
dw: Dry weight
EDCs: Endocrine Disrupting Chemicals
ER: Estrogen Receptor
ESRRA: Estrogen-related receptor alpha
ESTs: Expressed Sequence Tags
FPL: Female Penis Length
GM: Geometric Morphometrics
GO: Gene Ontology
hRXR: Human RXR
IMO: International Maritime Organization
Indels: Insertion-deletion
JAMP: Joint Assessment and Monitoring Programme
LBD: Ligand-Binding Domain
LXR: Liver X Receptor
KAAS: Kegg Automated Annotation Server
KEGG: Kyoto Encyclopedia of Genes and Genomes
MBT: Monobutyltin
MDH: Malate dehydrogenase
MgCl2: Magnesium Chloride
mRNA: Messenger RNA
NCBI: National Center for Biotechnology Information
NR: Non-Redundant Protein
NT: Nucleotide
NTC: No-Template Control
OPRK: Opioid Receptor Kappa
ORF: Open Reading Frame
OSPAR: Oslo and Paris Commission
OTs: Organotin Compounds
PCA: Principal Component Analysis
PCR: Polymerase Chain Reaction
PPARs: Peroxisome Proliferator-Activated Receptors
PTGIR: Prostaglandin receptor
PVC: Polyvinylchloride
PXR: Pregnane X Receptor
QC: Quality control
qPCR: Quantitative real-time PCR
QTL: Quantitative Trait Loci
RAD: Restriction-site associated DNA
RAR: Retinoic Acid Receptor

RIN: RNA Integrity Number
RNA: Ribonucleic acid
RNA-Seq: RNA Sequencing
ROR: Retinoic Acid-related Orphan Receptor
Rosi: Rosiglitazone
RPLI: Relative Penis Length Index
RTK: Receptor Tyrosine Kinase
RT-PCR: Reverse Transcriptase-PCR
RXR: Retinoid X Receptor
RWs: Relative Warps
SBW: Soft Body Wet Weight
SD: Standard Deviation
SGS: Second Generation Sequencing
Sn: Tin
SNPs: Single Nucleotide Polimorphisms
SREBP: Sterol Regulatory Element Binding Protein
SRY: Sex determining region Y
SOX9: SRY-box9
TBT: Tributyltin
TBTCl: Tributyltin Chloride
TFs: Transcription Factors
TPT: Triphenyltin
UTR: Untranslated regions
VDSI: Vas Deferens Sequence Index

CHAPTER 1

General introduction

1.1. RESPONSE TO ENVIRONMENTAL CHANGE

Environmental change is a general term that can be defined as a change or disturbance of the environment either by natural variability (e.g. temperature (Cossins et al. 2006), salinity (Kalujnaia et al. 2007), wave action (Etter 1996)), by human activity (e.g. pollution (Alzieu 2000), deforestation (Patino et al. 2010)), or by, at some extent, a combination of both (climate change (Atkins and Travis 2010)). Organism responses to such disturbances can include modifications of biochemical, physiological, morphological, or behavioural traits of adaptive significance. How individuals and ecosystems respond to environmental change is of worldwide concern and the focus of diverse areas of research. Ecological genomics, by its interdisciplinary approach, has been widely used to address the complexity of phenotypic and genetic responses observed.

1.1.1. Ecological genomics

Ecological research groups have been using largely neutral genetic markers, such as microsatellites and mitochondrial DNA to understand population structure and connectivity. However, patterns of variation and divergence in adaptive traits are not well reflected by such markers (Pfrender et al. 2000; Gomez-Mestre and Tejedo 2004).

Further interest in a more direct approach of the dynamics of genes under selection and their underlying mechanisms in response to changing environments is now being addressed in ecological genomics studies and is a fundamental topic in modern evolutionary ecology (Stapley et al. 2010).

Ecological genomics (Fig. 1) is "a scientific discipline that studies the structure and functioning of the genome with the aim of understanding the relationship between the organism and its biotic and abiotic environments" (van Straalen and Roelofs 2006). This is being achieved through the application of functional genomic techniques such as quantitative trait loci (QTL), microarray analyses and proteomics which offer new and detailed insights into mechanisms supporting all biological processes and facilitate the identification and characterization of genes with ecological and evolutionary relevance (Landry and Aubin-Horth 2007; Ungerer et al. 2008).

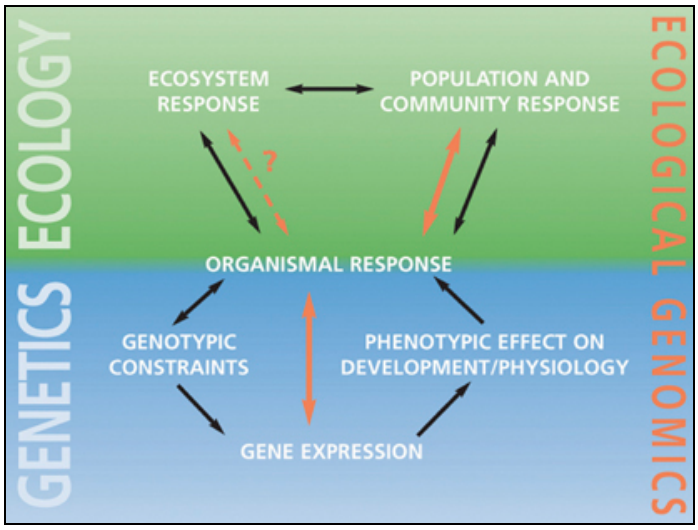


Figure 1. Conceptual framework for Ecological Genomics (Ungerer et al. 2008).

Until recently, the identification and understanding of specific genes in ecology and evolution has been limited to genomic model systems such as Drosophila melanogaster, Mus musculus, Arabidopsis thaliana and Caenorhabditis elegans whose ecology is not well studied. There is usually a lack of information on natural history and responses to environmental change within a natural ecological context (Ungerer et al. 2008). The use of model ecological systems for such approaches have been limited by a combination of high costs, small research communities and a need for truly integrated scientific research programs (Wheat 2008).

However, recent technological advances in high throughput sequencing have greatly reduced the difficulty for genomic tool development which facilitates functional genomics insights (Margulies et al. 2005; Ellegren 2008). Consequently, with the advent of next generation sequencing (described below) it is now relatively easy to generate genomic toolkits for ecological model organisms that display a well-understood ecology, including knowledge of adaptation to different environments (Stapley et al. 2010). Such new technologies and approaches offer the opportunity to perform genomic studies on many additional ecologically interesting species without the requirement of a closely related genetic model organism (although the latter remains useful), (Stapley et al. 2010). Developing genomics tools for ecological organisms is desirable as they will facilitate the

study of a wider range of phenotypic traits over evolutionary timescale and in more populations than possible previously (Wheat 2008).

Gene-environment interactions in non-model organisms using these emerging technologies that enable the integration of genomic and ecological data are expected to revolutionize several areas of biology and contribute to classical questions such as the genetic basis of adaptation, linkages between genotype and phenotype, and the factors determining the rate of genotypic and phenotypic shifts in relation to environmental change.

Consequently, in this thesis the ecological model species, Nucella lapillus (L., 1758), was used in order to explore the molecular basis of endocrine disruption in response to organotin pollution and to further explore phenotypic plasticity in response to wave action and crab predation employing an ecological genomics approach. A brief background on these main subjects and related technologies are presented below.

1.2. NEXT-GENERATION SEQUENCING

The sequencing of DNA molecules began in the 1970s with the development of the Sanger method which involves a reaction where chain-terminator nucleotides are labelled with fluorescent dyes, combined with fragmented DNA, DNA sequencing primers and DNA polymerase (Sanger et al. 1977). Despite continued improvements to this technology, which indeed is still widely used, urgency for sequencing entire genomes of organisms promoted the development of new ("next-generation") sequencing approaches that essentially had to be cheaper and less laborious and time-consuming, facilitating their up-take and application by a wider scientific community.

Next-generation sequencing technologies provide new opportunities for research programs focusing on a phenotype of interest that can quickly move from no genetic resources to having various functional genomic tools for mechanistic understanding (Wheat 2008). Importantly, using these technologies, a range of genomic resources such

as whole genome sequences, transcriptome (which includes the part of the genome that encodes proteins) sequence and genome-wide marker panels can be generated within the scope of a three-year project (Wheat 2008; Stapley et al. 2010). Moreover, their capacity to make existing techniques cheaper and faster is supplemented by the fact that they enable, for the first time, genomic studies to be conducted in any organism (Wheat 2008; Stapley et al. 2010). Therefore, even from a starting point of no genetic resources in the target species and no whole genome sequence in closely related species (also known as reference genome), the tools required to identify genetic mechanisms involved in adaptation can be generated - e.g. genome scans using single nucleotide polymorphisms (SNPs) in Littorina saxatillis (Galindo et al. 2010) and accessing diversity of expressed sequence tags in cichlids (Elmer et al. 2010).

1.2.1. Resources and applications

Recently developed next-generation sequencing technologies include 454 implementation of pyrosequencing (from 454 Life Sciences, introduced in 2005), Solexa reversible terminator technology (from Illumina, introduced in 2006) and SOLiD (from ABI, introduced in 2007). All three technologies share a common workflow, but differ greatly in the type of solid support used and the chemistry used to interrogate the DNA base pairs (see Fig. 2 and Fig. 3). These methods use nanotechnology to generate gigabases of sequence in a single experiment (Hudson 2008). They differ from traditional sequencing methods in two ways. First, rather than sequencing individual DNA clones, hundreds of thousands (the 454 system) to tens of millions (Solexa and SOLiD) of DNA molecules are sequenced in parallel. Second, the sequences obtained are much shorter (up to 125 nucleotides for the Illumina and up to 50 nucleotides for the ABI technology, and 400-500 nucleotides (predicted to be 1,000 by the end of 2010) for the 454 system)) than those generated by traditional sequencing (typically more than 800 nucleotides). Matching these shorter sequences unambiguously to the reference genome is more difficult, but this is a relatively minor trade-off compared with the enormous total sequence generated (Graveley 2008).

The three sequencing systems have already revolutionized the study of chromatin structure, DNA-binding proteins, DNA methylation, genome organization and small RNAs and more recently these technologies proved to be very useful, unbiased and accurate for studying transcriptomes (Jarvie and Harkins 2008).

Long reads: ABI 3730XL
Read length: Up to 1,100 bases
Reads per run: 96
Capacity: Approximately 1 MB per day and machine
Technology: Sanger
First choice for: finishing projects: full length cDNA sequencing: single sample sequencing

Extreme speed: Roche GS FLX
Read length: In average 400 bases
Reads per run: 1 million
Capacity: Approximately 400 MB per run
Technology: Pyrosequencing
Optimal for: de novo sequencing (procaryots and eucaryotes); Metagenomes, transcriptomes/ cDNA/ BAC/ fosmid pools

Huge data quantity: Illumina Genome Analyser
Read length: Up to 125 bases
Reads per run: Up to 3 billion
Capacity: Up to 300 Gb per run
Technology: Sequencing by synthesis

:ABI SOLid
Read length: Up to 50 bases
Reads per run: Up to 1 billion
Capacity: Up to 60 GB per run
Technology: Sequencing by ligation
Highly attractive for: re-sequencing projects, small RNAs, SAGE/CAGE and ChiP-Seq, ultra deep mutation/SNPs

Figure 2. Sanger and next-generation sequencing: options and facilities (adapted from GATC biotech-www.gatc.biz).

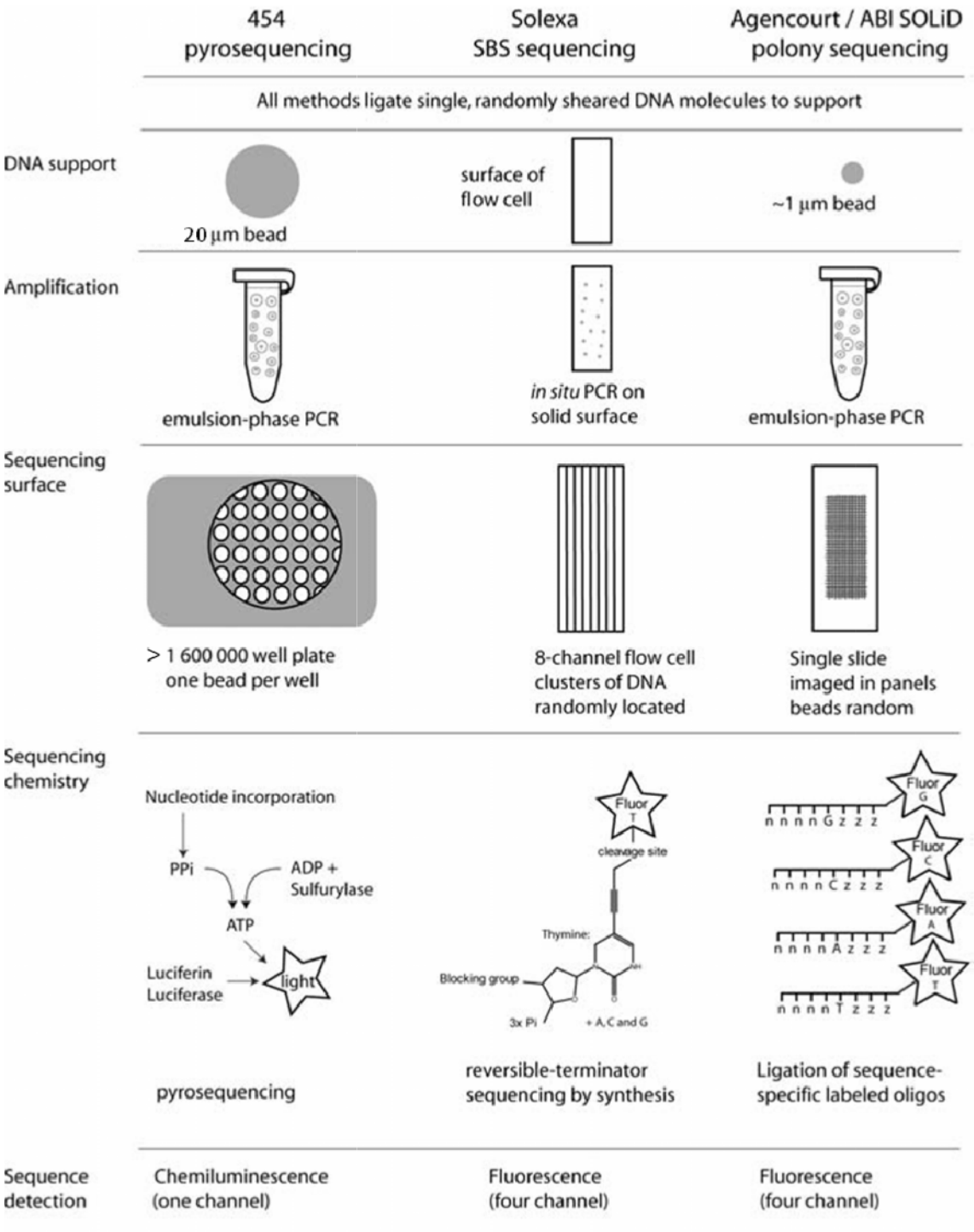


Figure 3. The three commercially available next-generation sequencing methods: key features and differences (adapted from Hudson, 2008).

A new step in sequencing technology is presently being developed – the third generation sequencing – and it is supposed to be released by the end of 2010 with the introduction of the Single Molecular Real Time DNA sequencer from Pacific Biosciences (McCarthy 2010). It uses a very different technology from the ones described above (also called second generation sequencing (SGS)) and is predicted to produce read lengths up to 10,000 bases long and more than 100,000s time faster than the current next generation sequencing (Stapley et al. 2010). As main advantages to SGS, this longer reads are expected to improve de novo genome assemblies, detect epigenetic changes to DNA and observe RNA translation in real time (McCarthy 2010; Stapley et al. 2010).

1.2.2. Transcriptome Sequencing: 454 Roche sequencing approach

The transcriptome is the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition (Wang et al. 2009). Transcriptome sequencing provides direct access to the messenger RNA (mRNA) sequence containing a coding gene sequence as well as both the 5' and 3' flanking untranslated regions (UTR) (the 3' end of the UTR is a long stretch of A's called the poly A tail), (Wheat 2008). Sequence from mRNA can be used for quantifying global gene expression (via designed microarrays) or genome wide coding variation (via sequencing 100's of genes), (Bouck and Vision 2007; Wheat 2008). When the mRNA material is a pool of outbred individuals, transcriptome sequencing can also provide data for finding microsatellite and SNP molecular markers, which can be located in either coding or the more variable UTR regions (Beldade et al. 2006; Bouck and Vision 2007). Recent technological advances (e.g. 454 Roche sequencing) have brought transcriptome sequencing, which was traditionally labor intensive and costly, to within reach of any research group during a normal grant period and funding range (e.g. Vera et al. 2008).

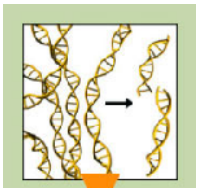
The 454 Roche sequencing technology, a high-throughput DNA sequencing using a novel massively parallel sequencing-by-synthesis approach, experienced rapid growth since its acquisition by Roche Diagnostics and release of the GS20 sequencing machine in 2005 (the first next-generation DNA sequencer on the market). Following some

improvements, the Genome Sequencer FLX instrument was released in 2007 and then updated with the GS FLX Titanium series reagents for use on the current instrument (released in 2008). This new version can sequence 400-600 million base pairs with 400-500 base-pair read-lengths.

The 454 Roche sequencing plate can be physically divided into a maximum of sixteen regions and barcode sample-specific sequence tags (adaptors) can be used, allowing the simultaneous sequencing of different samples. For that purpose, 5'-tagged PCR primers to distinguish amplicon sequences from different sources were initially used (Binladen et al. 2007) and later improved by Meyer et al. (2007) who developed the parallel tagged sequencing that uses a barcoding adaptor and a restriction system that excludes background sequences. Samples are pooled prior to 454 sequencing and are identified after sequencing by their unique sequence tags.

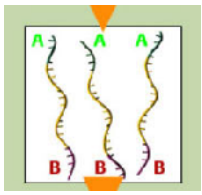
The 454 Roche sequencing approach (Fig. 4) involves breaking up genomic DNA into fragments, placing the individual fragments onto specially designed microbeads where the many copies of each fragment are amplified via emulsion-PCR. The amplified fragments are then loaded onto very small wells on a special plate (PicoTiterPlate™). As the wells are loaded with samples, reagents are pumped across the plate. The addition of the reagents result in an enzymatic reaction between complementary bases in the DNA fragments, and a light signal is created and read by the Genome Sequencer FLX analyzer (Roche 454 website, www.454.com).

Sample Input and Fragmentation



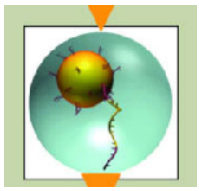
The Genome Sequencer FLX System supports the sequencing of samples from a wide variety of starting materials including genomic DNA, PCR products, BACs, and cDNA. Samples such as genomic DNA and BACs are fractionated into small, 300- to 800-basepair fragments. For smaller samples, such as small non-coding RNA or PCR amplicons, fragmentation is not required. Instead, short PCR products amplified using Genome Sequencer fusion primers can be used for immobilization onto DNA capture beads as shown below under "One Fragment = One Bead".

Library Preparation



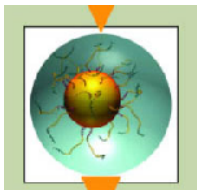
Using a series of standard molecular biology techniques, short adaptors (A and B) - specific for both the 3' and 5' ends - are added to each fragment. The adaptors are used for purification, amplification, and sequencing steps. Single-stranded fragments with A and B adaptors compose the sample library used for subsequent workflow steps.

One Fragment = One Bead



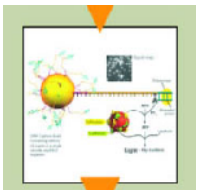
The single-stranded DNA library is immobilized onto specifically designed DNA Capture Beads. Each bead carries a unique single-stranded DNA library fragment. The bead-bound library is emulsified with amplification reagents in a water-in-oil mixture resulting in microreactors containing just one bead with one unique sample-library fragment.

emPCR (Emulsion PCR) Amplification



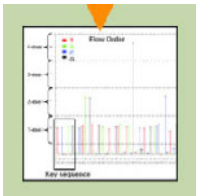
Each unique sample library fragment is amplified within its own microreactor, excluding competing or contaminating sequences. Amplification of the entire fragment collection is done in parallel; for each fragment, this results in a copy number of several million per bead. Subsequently, the emulsion PCR is broken while the amplified fragments remain bound to their specific beads.

One Bead = One Read



The clonally amplified fragments are enriched and loaded onto a PicoTiterPlate device for sequencing. The diameter of the PicoTiterPlate wells allows for only one bead per well. After addition of sequencing enzymes, the fluidics subsystem of the Genome Sequencer FLX Instrument flows individual nucleotides in a fixed order across the hundreds of thousands of wells containing one bead each. Addition of one (or more) nucleotide(s) complementary to the template strand results in a chemiluminescent signal recorded by the CCD camera of the Genome Sequencer FLX Instrument.

Data Analysis



The combination of signal intensity and positional information generated across the PicoTiterPlate device allows the software to determine the sequence of more than 400,000 individual reads per 7.5-hour instrument run simultaneously. For sequencing-data analysis, three different bioinformatics tools are available supporting the following applications: de novo assembly up to 120 megabases; resequencing up to 3 gigabases; and amplicon variant detection by comparison with a known reference sequence.

Figure 4. Detailed sequencing workflow underlying 454 Roche sequencing technology (source: www.454.com).

The benefits of 454 sequencing have been exploited for an increasing number of applications, including genomic sequencing, cDNA sequencing and ultra-deep amplicon sequencing (Meyer et al. 2007), convincing many researchers to shift from traditional Sanger capillary sequencing toward the 454 Roche approach (Ellegren 2008; Wheat 2008).

Transcriptomics is essential to monitoring the genomic activation of cells or organisms in response to environmental signals. The information-containing portions of genomes are transcribed into two RNA classes: messenger RNAs, which are translated into proteins, and non-coding RNAs, which have regulatory and mechanistic roles (van Straalen and Roelofs 2006). Therefore, studying the transcribed portion of the genome (the transcriptome), significantly assists gene identification, providing insights into functional genomics of an organism.

The combination of long, accurate reads and high-throughput makes 454 Roche sequencing analyses a powerful platform for detailed transcriptome investigation. Transcriptome sequencing encompasses studies on mRNA transcript-expression analysis (full-length mRNA, expressed sequence tags (ESTs) and ditags), novel gene discovery and annotation, gene space identification in novel genomes, assembly of full-length genes, SNPs, insertion-deletion (indels) and splice-variant discovery, as well as analysis of allele-specific expression, chromosomal rearrangement, mutational analysis of expressed genes and read-count-based gene expression profiling (digital transcriptomics) (Bainbridge et al. 2006; Barbazuk et al. 2007; Toth et al. 2007; Jarvie and Harkins 2008; Sugarbaker et al. 2008; Vera et al. 2008).

The assembly of 454 Roche sequences from samples that contain large amounts of repetitive DNA, such as eukaryotic genomes, may prove problematic for conventional fragment assembly programs. In contrast, the read-length limitation associated with 454 technology is of no concern for transcriptome sequencing and analysis because transcriptomes are smaller than the genomes from which they derived and typically contain less repetitive DNA (Emrich et al. 2007). Furthermore, a crucial step in assembly, that indeed largely contributed for the success of 454 Roche transcriptome sequencing

was the successful de novo transcriptome assembly without a reference genome, with the Glanville fritillary butterfly being one of the first successful examples (Vera et al. 2008).

1.3. TRANSCRIPTION PROFILING: MICROARRAY TECHNOLOGY APPROACH

Transcription profiling is one of the most informative genomic approaches for addressing ecological questions since it involves the development of a complete overview of all the genes in a genome that are up-regulated or down-regulated in response to some factor(s) of interest (van Straalen and Roelofs 2006).

Several approaches can be adopted for transcription profiling, but microarray technology, particularly high-density oligonucleotide arrays that allow a greater coverage of the genome, has been the most extensively applied for global gene expression analysis (Lockhart et al. 1996). An advantage of microarray analysis is that once the array has been made, albeit at high cost, many measurements can be made quickly, and at relatively low cost. However, only known genes can be spotted on the array and for non-model species, the generation of the underpinning resources have been a problem (Gracey and Cossins 2003).

The emergence of next-generation sequencing (described above), however, is filling this gap providing the necessary datasets for array design for any species of interest and are invaluable genomic resources for a global overview of how the expression of every gene in the organism is responding to a particular stressor (Gracey 2007). The combination of these technologies along with advances and facilities in the microarray design and production (e.g. Agilent custom arrays using eArray for oligoarray design) are enabling the exploration of ecological questions difficult to address so far (e.g. Spade et al. 2010).

In microarray technology (Fig. 5), in order to get an expression overview of all the transcripts, two labelled samples, one experimental, and one control, are competitively

hybridized for a large number of sequences (van Straalen and Roelofs 2006). After hybridization, a scanner records the intensity of the fluorescence emission signals that is proportional to transcript levels in the biological samples. The microarray data are therefore analysed using specific softwares (e.g. R/Bioconductor, Limma) that enable clustering of genes with similar expression patterns, assuming that they share common biological functions.

This method, however, show some limitations such as the already mentioned reliance upon existing knowledge about genome sequence, high background levels owing to cross-hybridization, a limited dynamic range of detection owing to both background and saturation of signals and the often need for complicated normalization methods in order to compare expression levels across different experiments (Wang et al. 2009). In contrast to microarray methods, sequence-based approaches directly determine the cDNA sequence, and the development of novel high-throughput DNA sequencing methods (Illumina, SOLID and Roche 454 sequencing) has provided a new method for both mapping and quantifying transcriptomes. This method, RNA-Seq (RNA sequencing), has clear advantages over existing approaches and is expected to revolutionize the manner in which transcriptomes are analysed (Wang et al. 2009).

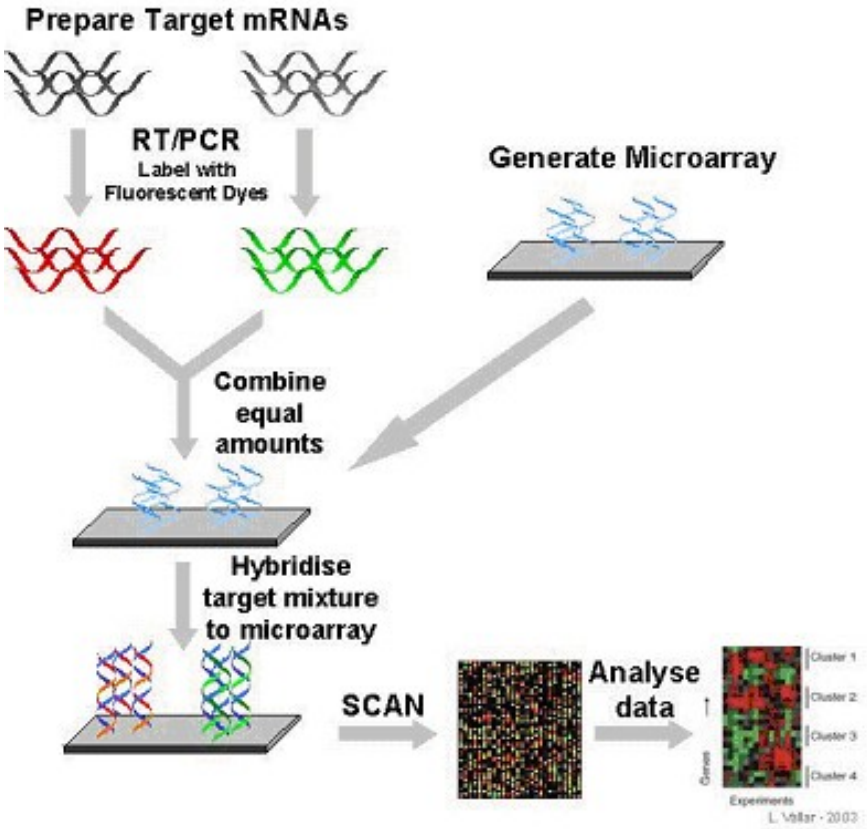


Figure 5. Microarray technology workflow (source: http://www.microarray.lu/en/MICROARRAY_Overview.shtml).

1.4. ENDOCRINE DISRUPTION

The endocrine system is a complex network of glands and hormones that regulates many of the body's functions, including growth, development and maturation, as well as the way various organs operate (Raven et al. 2005).

The endocrine systems of vertebrates largely share molecular mechanisms; however, the physiological consequences of these mechanisms differ in different classes, e.g. sex differentiation - sex is determined by the sry gene in mammals and by the dmy gene in medaka fish, whereas temperature dependent sex determination is common in crocodilians and turtles (Iguchi and Katsu 2008). For invertebrate species, however,

information on the endocrine system and the hormone receptor system is much more limited (Oetken et al. 2004; Porte et al. 2006).

Mainly due to industrialization development, the production of chemicals and their release into the environment has greatly increased, with particular negative impact in aquatic environments.

Interference with the hormonal system exerted by an exogenous substance – xenobiotic - is referred to as endocrine disruption and the substances themselves are known as endocrine disrupting chemicals (EDCs), which have the ability to disturb endogenous hormone signalling pathways (Swedenborg et al. 2009). EDCs encompass a variety of natural and synthetic chemicals, including hormones, pharmaceuticals, pesticides, compounds used in the plastics industry and in consumer products, and other industrial by-products and pollutants. Most EDCs are fat-soluble with a general potential to bioaccumulate and biomagnify up in the food chain (Porte et al. 2006).

Endocrine disruption is of worldwide concern and has been focused primarily on human health, however, it affects a large taxonomic range of invertebrate and vertebrate organisms including marine snails, fish, alligators, frogs, birds and whales (Oetken et al. 2004; Iguchi and Katsu 2008), covering marine, freshwater and terrestrial ecosystems.

Through biological monitoring programs and empirical prove of adverse impacts of EDCs to humans and wildlife, legislation and regulations have been implemented worldwide in order to control or in some cases ban the use of such compounds, e.g. tributyltin (described below). However, considering that more than 100,000 known man-made chemicals are used in everyday life (Thain et al. 2008) and that they end up in the environment, it is very difficult to identify and monitor such a large amount of EDCs, making them of continuous concern.

Endocrine disruption caused by environmental xenobiotics can have profound effects on the organism, including abnormal development of male and female reproductive traits, feminization of males, masculinisation of females, lower sperm counts, disruption of reproductive cycle and reduced fertility, malformations, decreased immunity, carcinogenesis, behavioural changes, etc. (Colborn et al. 1993; Mueller 2004; Guillette and Moore 2006; Tabb and Blumberg 2006), and it may ultimately lead to local population extinction (Bryan et al. 1986; Gibbs et al. 1991; Huet et al. 1996).

EDCs can act at multiple sites through multiple mechanisms of action and recently, several nuclear receptors have been identified as mediators of endocrine disruption as well as steroid hormone receptors (Iguchi and Katsu 2008). Many studies have focused on the effects that EDCs pose on reproductive processes regulated by hormonal signalling primarily mediated by members of the nuclear receptors family, in particular the estrogen and the androgen receptors as well as the thyroid hormone receptors, aryl hydrocarbon receptors, retinoid X receptor (RXR) and peroxisome proliferator-activated receptors (PPARs), (Iguchi and Katsu 2008; Swedenborg et al. 2009).

Although receptor-mediated mechanisms received most attention, other mechanisms, such as hormone synthesis, transport and metabolism, activation of nuclear receptors and gene methylation, have been identified as equally important targets for the EDC action (Fig. 6), (Tabb and Blumberg 2006; Iguchi and Katsu 2008).

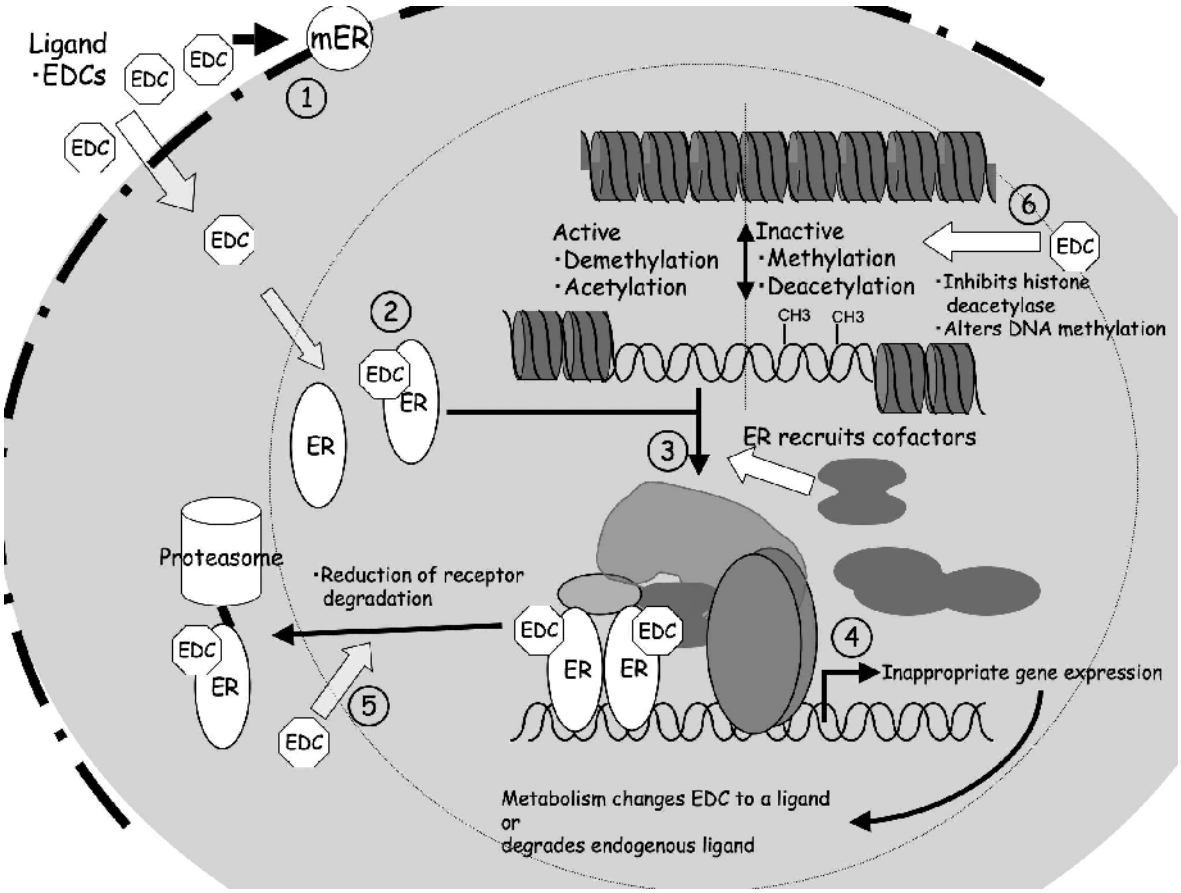


Figure 6. Endocrine-disrupting chemicals (EDCs) may operate by a variety of mechanisms. When EDCs arrive at a cell membrane (top left) they may bind to a membrane estrogen receptor (mER) (1), or pass through the membrane and bind to a nuclear estrogen receptor (ER) in the nucleus (2).When the complex of EDC and ER binds to a gene containing an estrogen-responsive element (3), it recruits molecules that promote gene expression by boosting gene transcription by the RNA polymerase complex (large ovals), and thus may cause gene expression at inappropriate times (4). EDC/ER complexes can also bind to proteasomes, which can lead to a reduction of the normal process of degradation of ER (5). DNA is shown near the top of the figure wrapped around histone complexes, remaining inactive. In the middle of the figure, DNA is shown unwrapped from the histone complexes, exposing it to molecules that boost transcription. EDCs may cause methylation of DNA or deacetylation of histone (as shown by the methyl groups on the right side of the gene), both of which reduce gene expression. Alternatively, EDCs may cause demethylation of DNA or, by inhibiting the enzyme histone deacetylase, lead to the acetylation of histone (left side of the gene). Both of the second two effects induce gene expression (6) (Iguchi and Katsu 2008).

1.4.1. Endocrine disruption in Gastropoda

The hormonal regulation of biological functions such as growth, sexual differentiation and reproduction is a common characteristic for all phyla, including invertebrates (Oetken et al. 2004). Invertebrates account for roughly 95% of all animals representing diverse life cycles such as the presence of metamorphosis, diapause and formation of larval forms, that are not present in vertebrates. For that reason, endocrine systems of invertebrates are considered more diverse than those found in vertebrates; however they are by far less documented (deFur 2004; Oetken et al. 2004).

Among invertebrates, the endocrine system of insects has been the most investigated, due to their economic and ecological significance mainly related with the need for insect pest control (Soin and Smagghe 2007). Yet, molluscs, especially gastropods, are widely studied mainly due to their abundance and use as bioindicators in water pollution monitoring (Ketata et al. 2008). Indeed, TBT-induced imposex and intersex in gastropods is considered one of the best documented examples of endocrine disruption (Matthiessen and Gibbs 1998).

Still, incomplete knowledge of the endocrinology of sexual differentiation and development in gastropods (Sternberg et al. 2010) has constrained progress in understanding the imposex mechanism. Nevertheless, some important advances are being reported (Sternberg et al. 2010):

i) Vertebrate-type sex steroids including estrogens and androgens have been identified in gastropod (Gas-chromatography). However, the detection of sex steroids in gastropods provides no insight as to whether these biomolecules actually function as hormones in these organisms.

ii) Almost all of the enzymatic activities already identified in vertebrate steroidogenesis have been detected in at least one gastropod species, indicating that gastropods have the enzymatic machinery for synthesizing sex steroids.

iii) Metabolic studies indicate that gastropods are capable of biosynthesizing and biotransforming the sex steroids that are present in their tissues.

iv) No successful identification of an androgen-type receptor has been reported, but an estrogen receptor has been identified in gastropods;

v) The molluscan estrogen receptor seems unable to bind estrogens but may be a constitutive transcriptional activator via estrogen response elements.

1.5. ORGANOTIN CONTAMINATION

Organotin compounds (OTs) are characterized by a tin (Sn) atom covalently bound to one or more organic substituents (e.g. methyl, ethyl, butyl, propyl, phenyl, octyl) (Hoch 2001). Organotins have a wide industrial application and consequently, considerable amounts of toxic OTs have entered into various ecosystems. Among OTs, tributyltin (TBT) pollution has been the best documented and pervasive of cases.

Tributyltin compounds are man-made organic substances containing the $(C_4H_9)_3Sn$ functional group (Fig. 7). Tributyltin is the active component of many products that are biocides against a broad range of organisms and is used primarily as an antifouling paint additive on ship and boat hulls, docks, fishnets and buoys to avoid the growth of marine organisms such as barnacles, bacteria, tubeworms, mussels and algae (Sternberg et al. 2010). Additionally, TBT has been used as wood preservative in industry and agriculture, as a stabilizer in PVC plastic manufacturing, and as an antifungal in textiles and in industrial water systems (Hoch 2001).

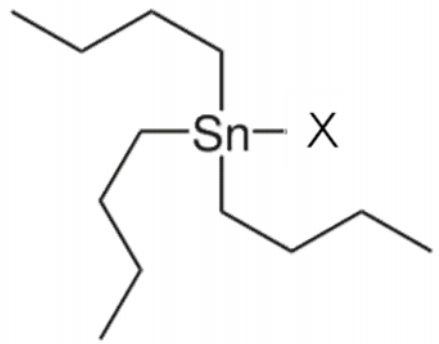


Figure 7. Chemical structure of tributyltin (adapted from http://en.wikipedia.org/wiki/Tributyltin).

Tributyltin compounds are moderately to highly persistent organic pollutants, although TBT itself is unstable and will break down in the environment unless it is combined with a chemical element. Due to its low water solubility, it strongly binds to suspended material and inorganic sediments (Laughlin et al. 1986). Additionally, TBT shows a substantial potential for bioaccumulation that biomagnifies through the food chain being of particular concern in long-lived biota (Murata et al. 2008). TBT is commonly manufactured as TBT oxide, TBT fluoride and TBT chloride and it is environmentally degraded by debutylation to form less toxic compounds such as dibutyltin (DBT), monobutyltin (MBT) and ultimately inorganic tin. Degradation depends on temperature and occurs through biotic processes and the rate of debutylation in sediments is dependent upon microbial activity (Sternberg et al. 2010). Under aerobic conditions, tributyltin takes one to three months to degrade but will persist longer under anoxic conditions: whereas the half–life of TBT in water is about three months it can range from 6 months to 8.7 years in anaerobic sediments (Sternberg et al. 2010).

1.5.1. TBT history and legislation restrictions

Fouling communities on vessels and man-made structures at sea increases drag and, and in the case of vessels, fuel consumption leading to substantial increase of emissions and economic consequences (Santillo 2002). Ships' antifouling paints containing TBT were introduced in the 1960s and soon were recognised as extremely effective and relatively economical antifouling biocides, contributing to a rapid take-up of organotin-based paints by private and commercial users in the 1970s (Evans and Sheppard 2000), and consequent extensive use for about four decades.

At the same time as its explosive increase in use, the first TBT effects on non-target organisms were observed. The first severe biological effects from TBT use as an antifouling agent were described in mid 1970's in oysters from Arcachon Bay (France), which developed shell anomalies and low growth rates that caused near collapse in production (Alzieu 1998, 2000). Almost simultaneously, effects on other marine molluscs were reported and reduction of gastropods densities were described (De Mora and

Pelletier 1997). Besides malformation in oysters and imposex in neogastropods, other deleterious TBT effects were reported to occur in a large taxa range such as microalgae, polychaetes, crustaceans, bryozoans, echinoderms, tunicates, fish (Bryan and Gibbs 1991) and mammals (Grun et al. 2006; Murata et al. 2008). TBT is reported to have only modest adverse effects on mammalian male and female reproductive tracts and does not alter sex ratios (Ogata et al. 2001; Omura et al. 2001). Instead, hepatic-, neuro-, and immunotoxicity appear to be the predominant effects of organotin exposure on mammals (Boyer 1989). Indeed, TBT was considered one of the most toxic compounds deliberately released into marine environments by man (Goldberg 1986; Fent 2004).

Worldwide concern over negative impacts of TBT resulted in the introduction of legislative restrictions for the use of organotin-based paints. Table 1 summarizes the main adopted decisions over time. Since 2008, the use of organotin paints from the global shipping fleet, the main source of contamination, has been abolished. Yet persistence of TBT in sediments and long-lived biota as well as new inputs mainly through their use as additives in a range of consumer products (e.g. wood preservatives, antifungal in textiles and industrial water systems) still remain (Santillo 2002). Therefore, a continuing negative impact on aquatic wildlife must be expected.

Table 1. Historical legislation restrictions over organotin-based paints; adapted from Santillo (2002).

Date	Actions
1982	France introduces legislation prohibiting the use of TBT paints on small vessels.
1985	First controls introduced in United Kingdom limiting concentrations of TBT in paints.
Jan 1987	United kingdom announces further restrictions on TBT content of applied antifouling paint.
May 1987	United Kingdom introduces ban on retail sale of TBT paint for use on vessels < 25 m and on fish cages.
Jun 1987	Paris Commission Recommendation 87/1 calls for similar ban over entire convention area (Northeast Atlantic)
1988	United States introduces restrictions.
1989	Restrictions introduced in Canada, Australia and New Zealand.
1991	Harmonised ban on retail sale of TBT paint introduced at European Union level.
1995	Ministerial declaration of fourth North Sea conference (Esbjerg) commits to working for global phase-out of TBT paint within International Maritime Organization.
1997	Concept of global phase out of organotin containing paints agreed at Marine Environmental Protection Committee's 40th session.
1998	Draft mandatory regulations aimed at such a phase-out adopted. OSPAR prioritises organotins for action to cease all releases. Cessation of all releases of organotins to marine environment, under OSPAR's hazardous substances strategy in 2020.
1999	Deadlines for phase-out adopted under International Maritime Organization Assembly Resolution A.895(21)
2001	Text of International Conventional on the Control of Harmful Anti-fouling Systems (AFS convention) to be finalised. In 2003 worldwide prohibition on new application of organotin antifoulants to all vessels and in 2008 the existing organotin antifouling coatings should be replaced on all vessels worldwide.
2003	Due to the slow ratification process and to promote the implementation of the AFS Convention in the member states, the European Union moved to the adoption of Directive 2002/62/EC and Regulation 782/2003 imposing the total interdiction of OTs AF paints application on EU ships after the 1st of July 2003 and the presence of TBT AF paints on ships' hulls from the 1st of January 2008.

The Oslo and Paris Commission (OSPAR) adopted the imposex response as a component of their international Joint Assessment and Monitoring Programme (JAMP) and recommended the dogwhelk N. lapillus as a sensitive bioindicator of TBT pollution (OSPAR 1998). Since legislation restrictions have started, an evident decrease in TBT pollution and imposex levels as well as populations recovery have been observed over time (Colson et al. 2006; Sousa et al. 2007; Oliveira et al. 2009).

1.5.2. Imposex: a bio-monitoring tool for TBT pollution

The imposex phenomenon (Fig. 8) provides one of the best ecological examples of anthropogenically-induced endocrine disruption in aquatic ecosystems (Matthiessen and Gibbs 1998). This phenomenon is mainly caused by exposure to TBT and it is characterized by the superimposition of male sexual characteristics (vas deferens, penis) on female gastropods (Smith 1971).

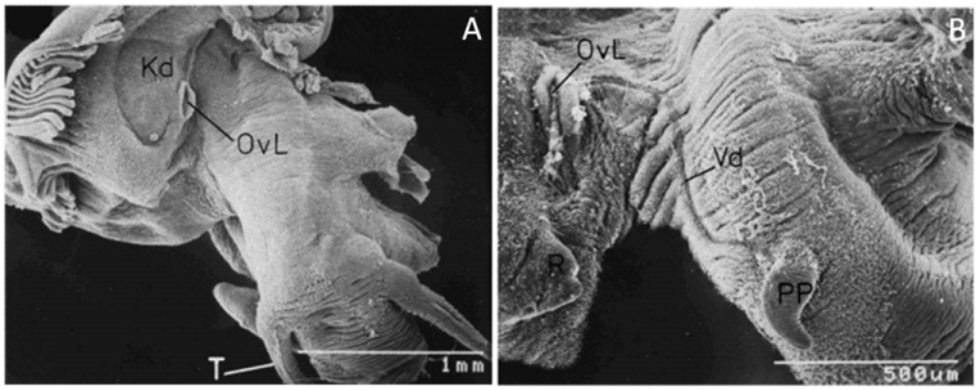
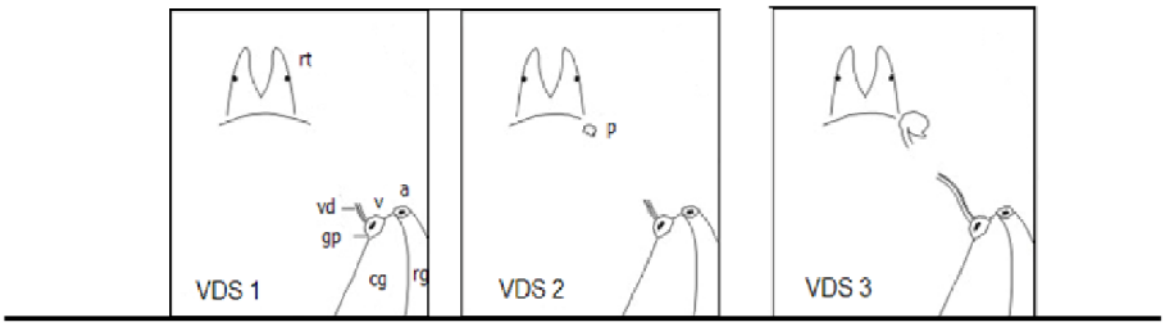


Figure 8. Scanning electron micrographs of Hydrobia ulvae females. A) Normal female without imposex; B) Sterilised female in the final stage of imposex with blocked oviduct. Abreviations: Kd: capsule gland; OvL: Oosparous opening of oviduct (open in A, closed in B); PP: penis; T: tentacles; Vd: vas deferens (Oetken et al. 2004).

In N. lapillus, females affected with imposex develop a penis and vas deferens that in an advanced stage can occlude the genital papillae, blocking the egg duct and preventing the release of egg capsules (Fig. 9), so leading to female sterilization. Additionally, the aborted capsules may build up and eventually rupture the capsule gland, which may kill the individual. Both processes had, in some cases, lead to population decline (Bryan et al. 1986). Imposex has been widely used to monitor trends in TBT contamination in the marine environment (Oliveira et al. 2009; Rato et al. 2009) and it has been reported in more than 150 gastropod species worldwide (Tillmann et al. 2001).



VDS	Imposex development
0	The unaffected female with no apparent male characters. The genital papilla and the vulva are clearly visible with no development of the vas deferens tissue.
1	An infolding of the mantle cavity epithelium in the ventral area next to the genital papilla marks the first development of the vas deferens.
2	Initiation of penis development by the formation of a ridge behind the right tentacle.
3	The ridge becomes recognisable as a small penis and development of the vas deferens from the base of the penis begins.
4	The two sections of vas deferens, started at the genital papilla and at the base of the penis, become fused into a continuous tube. The penis is larger and closer to the male penis size and shape.
5	The proliferation of the vas deferens tissue on the genital papilla (hyperplasia) results in the occlusion of the vulva, and the genital papilla becomes displaced, constricted or no longer visible.
6	The lumen of the capsule gland contains the material of aborted capsules; this material may comprise a single capsule or several to many that are compressed together to form a translucent or light to dark brown mass.

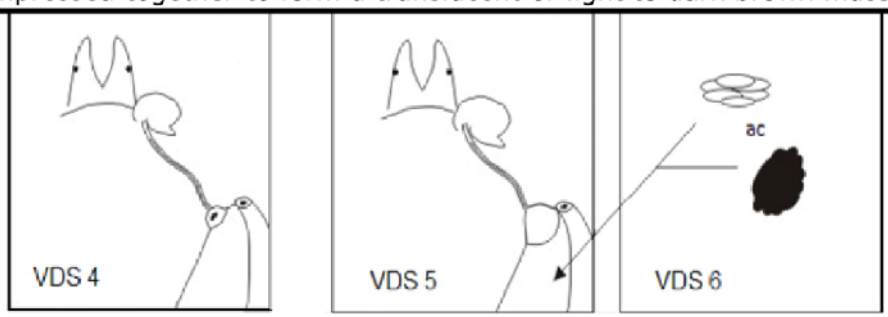


Figure 9. Stages in the development of imposex in Nucella lapillus based on Vas Deferens Sequence (VDS). Abbreviations: a – anus; ac – aborted capsules; cg – capsule gland; gp – genital papillia; p – penis; rg – rectal gland; rt – right tentacle; v – vulva; vd – vas deferense; adapted from (Gibbs et al. 1987).

1.5.2.1. Suggested pathways for imposex induction

Although providing a key ecological example of anthropogenically-induced endocrine disruption in aquatic ecosystems (Matthiessen and Gibbs 1998), the molecular and biochemical mechanisms underlying imposex remain unclear. However, several hypotheses for how TBT causes imposex have been proposed, generating new insights into our understanding of the endocrine system and reproductive physiology of gastropods (Sternberg et al. 2010). To date, three main possible pathways have been suggested: the steroid, the neuroendocrine and the retinoid. The interplay between the suggested pathways, though, is still poorly understood (Castro et al. 2007). An overview of each pathway is given below.

i) Steroid pathway

The steroid hypothesis proposes that TBT increases free testosterone levels in exposed females and that this imbalance initiates a biochemical cascade causing the imposex phenomenon. Several studies indicate that TBT disrupts the steroid signalling and physiological balance (Spooner et al. 1991; Bettin et al. 1996; Santos et al. 2005). Besides support that testosterone administration increased the incidence of imposex (Spooner et al. 1991; Bettin et al. 1996), exposure to the androgen receptor antagonist cyproterone acetate (CPA) alone reduced accessory sex organs and penis length in male N. lapillus and attenuated TBT's ability to induce imposex in females (Tillmann et al. 2001).

The elevation of free testosterone by TBT may involve the inhibition of enzymes that metabolize steroids and the main proposed targets to inhibition are the Cytochrome P450 aromatase (Spooner et al. 1991; Bettin et al. 1996; Santos et al. 2002), sulfotransferase (Ronis and Mason 1996) and acyl coenzyme A: steroid acyltransferase (Gooding et al. 2003; Janer et al. 2005).

Cytochrome P450 aromatase: TBT might inhibit the cytochrome P450-dependent aromatase which converts androgens to estrogens. TBT was shown to be an in-vitro

inhibitor of cytochrome P450 aromatase in humans (Heidrich et al. 2001) and periwinkles Littorina littorea (Ronis and Mason 1996). Additionally, aromatase inhibitors were able to induce imposex under laboratory conditions (Bettin et al. 1996; Santos et al. 2005) and a depression of aromatase activity in wild populations of Buccinum undatum affected by imposex was observed (Santos et al. 2002). However, the overall contribution of aromatase activity to testosterone metabolism in gastropods may be minor and therefore its inhibition would not result in a significant increase in testosterone levels (Sternberg et al. 2010).

Sulfotransferases: TBT might inhibit the action of sulfotransferases in metabolising testosterone to more excretable metabolites, decreasing the organism's ability to eliminate androgens. This theory, however, presents some weaknesses since the impaired androgen elimination has been reported in L. littorea (Ronis and Mason 1996), a species affected by intersex and not imposex and a direct relationship between TBT and sulfotransferases has not been shown. Moreover, even in TBT-induced imposex species, sulfurotransferases have not been shown to significantly contributed to the metabolism of testosterone (Gooding and LeBlanc 2001; Janer et al. 2006).

Acyl coenzyme A: testosterone acyltransferase (ATAT): TBT might inhibit the fatty acid esterification of testosterone. It has been found that the fatty acid esterification of testosterone is the main regulatory process for maintenance of free testosterone homeostasis in the mud snail Ilyanassa obsoleta and that the ATAT enzyme biotransforms free testosterone to testosterone-fatty acid esters that are retained by the organism (Gooding and LeBlanc 2001, 2004). Additionally, it was suggested that TBT suppresses the ability of I. obsoleta for testosterone-fatty acid esters production or accumulation (Gooding et al. 2003). So far, the ATAT hypothesis is the most likely explanation for how TBT elevates free testosterone in imposex-affected females (Sternberg et al. 2010).

ii) Neuroendocrine pathway

The neuroendocrine hypothesis proposes that TBT acts as a neurotoxicant in gastropods, causing the aberrant secretion of neurohormones that contributes to male sexual differentiation. It was suggested by Féral et al. (1983), followed by Oberdorster and McClellan-Green (2000, 2002) that i) excised ganglia from female Ocenebra erinacea were activated by exposure to TBT; ii) the activated ganglia induced the development of a penis in the excised tissues of the penis-forming area of Crepidula fornicate and iii) APGWamide, a putative penis morphogenic factor (PMF), significantly induced imposex in Ilyanassa obsoleta. More recently, however, APGWamide has been shown not to promote imposex in the purple dye murex Bolinus brandaris (Santos et al. 2006) and a causal relationship between TBT exposure, abnormal APGWamide release and imposex has not yet been established (Sternberg et al. 2010).

iii) Retinoid pathway

The retinoid hypothesis is the most recent, and proposes that the retinoid X receptor (RXR), a nuclear receptor, plays an important role in inducing the development of imposex. The hypothesis was proposed by Nishikawa et al. (2004) after showing that Thais clavigera, a close relative of N. lapillus, has RXR similar to that in humans and other vertebrates and invertebrates (Horiguchi et al. 2007), and that the rock shell RXR binds to both 9-cis-retinoic acid (9CRA), the natural ligand of RXR, and organotins. Additionally, it was shown that injection of 9CRA or organotins into normal T. clavigera (Nishikawa et al. 2004) and N. lapillus (Castro et al. 2007) females stimulated the development of imposex and that TBT inhibited the binding of 9CRA to the RXR of T. clavigera (Nishikawa 2006). Therefore, organotins would mimic the endogenous ligand of RXR disrupting the signalling pathway which are retinoic acid dependent (Castro et al. 2007).

More recently (Sternberg et al. 2008), it has been suggested that RXR-mediated signalling may have an important role in sex differentiation in gastropods - 9CRA or some other RXR ligand being the strongest candidate for the ultimate regulator of male

reproductive tract development - and that TBT would induce female imposex by initiating RXR signalling prematurely (Fig. 10).

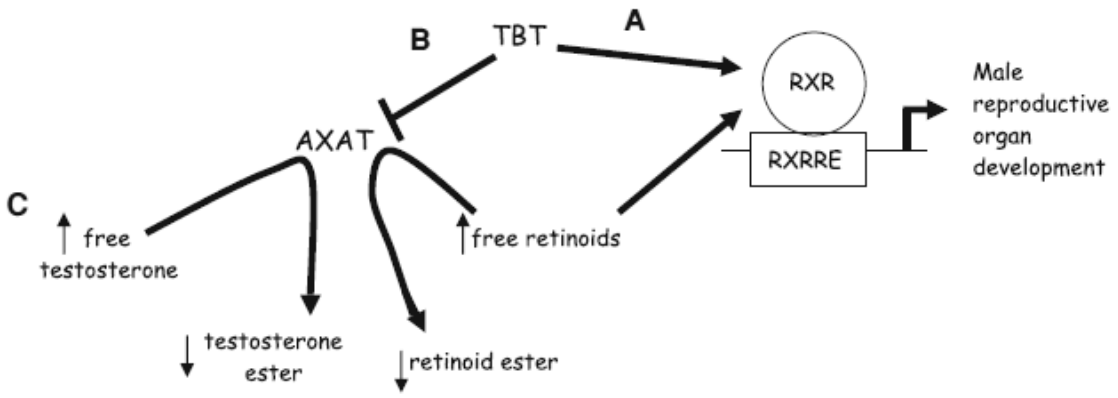


Figure 10. Proposed mechanism for TBT-induced imposex. TBT activates the retinoid X receptor (RXR) signalling pathway to initiate the transcription of genes necessary for male reproductive system development (a) directly, by binding to and activating RXR; or (b) indirectly, by inhibiting acyl coenzyme A:acyltransferase (AXAT), resulting in an increase in endogenous retinoid (and testosterone) levels. RXR is then activated by the endogenous free retinoid. RXR stimulates gene transcription through interaction with RXR response elements (RXRRE). Similarly, exogenously-administered testosterone (c) competitively inhibits retinoid esterification resulting in elevated free, endogenous retinoid levels that are capable of activating the RXR signalling pathway, leading to male reproductive organ development (Sternberg et al. 2010).

To date, it has been reported that the environmental contaminant TBT is a ligand for RXR in Thais clavigera (Nishikawa et al. 2004; Horiguchi et al. 2007) and N. lapillus (Castro et al. 2007) and it acts as a ligand for both the RXR and the PPARγ in the frog Xenopus laevis, mice and in humans (Grun et al. 2006). Therefore, TBT, which induces imposex in marine snails and promotes adipogenesis in X. laevis and in mice, is an example of an environmental endocrine disrupter that promotes adverse effects, from the snail to mammals, through common signalling (Iguchi and Katsu 2008).

1.6. PHENOTYPIC PLASTICITY

The ability of populations to adapt to a particular habitat is a key topic in evolutionary biology as the explotation of new niches is a key component of the speciation process (Etter 1988; Nussey et al. 2007). In order to survive in heterogeneous habitats, organisms can adopt three main evolutionary strategies i) fix a generalist genotype more or less suitable over a broad range of environmental conditions; ii) adapt to a particular environmental condition by heritable variation in a particular trait (direct genetic determination); iii) control the sensitivity of the genotype to changes in the environment (phenotypic plasticity), (Etter 1988; Martinez-Fernandez et al. 2010). The first and third one are expected to be typical of organisms with relatively high dispersal abilities, while the second is preferentially expected for species with restricted dispersal ability, although different exceptions are known (Hollander 2008).

Phenotypic plasticity is the development of different phenotypes from a single genotype, depending on the environment and it is a major theme in studies of ecology and evolution (Pigliucci 2005). Proximate mechanisms underlying plastic responses have, however received less attention, but it has long been recognized that phenotypic variation reflects both genetic and environmental influences (Bradshaw 1965; Etter 1988). Knowledge of the contributions of genotype, plasticity and their interaction to phenotypic expression is crucial for understanding the evolution of adaptive character traits in heterogeneous environments (Gould 1966; Endler 1986; Stearns 1989; Travis 1994; Via et al. 1995; Schlichting and Pigliucci 1998). Yet it remains pertinent to ask what conditions favour plasticity, local genetic adaptation, or both (Pigliucci 2001). It has also been suggested that genetic variation is favoured in stable environments (Hori 1993; Smith 1993), whereas phenotypic plasticity is favoured in unstable and fluctuating environments (Stearns 1989; Scheiner 1993; Svanback et al. 2009). Additionally, it is often argued to be the result of natural selection and it is currently seen as a primary mechanism by which organisms can respond adaptively to environmental change (Day et al. 1994; Ellers and Stuefer 2010).

Plasticity can influence the evolution and adaptive responses of organisms by altering the relationship between the phenotype, which is the target of selection, and the genotype and it has been hypothesized that plastic individuals would be favoured over nonplastic ones by natural selection. Moreover, it provides one mechanism for maintaining genetic variation because it reduces the probability that a single genotype will be optimal in all environments (Trussell and Etter 2001). Plasticity can, therefore influence which genotypes are favoured in particular environments, how populations respond to selection, the rate at which genotypes are fixed and the maintenance of genetic variation (Trussell and Etter 2001). By it nature, phenotypic plasticity plays a crucial role on the direction and rate of evolution and has consequently been focus of many evolutionary studies.

Several plastic responses have been documented in different taxa such as plants (Callaway et al. 2003), fish (Domenici et al. 2008), barnacles (Lively 1986) and rocky intertidal snails (Appleton and Palmer 1988). These responses can be expressed as morphological modifications, changes in physiology, life history and behaviour (Pigliucci 2001; Aubin-Horth and Renn 2009). A plastic response, however, usually involves costs and constraints such as reduction in feeding and growth (DeWitt 1998).

It is now known that environmental signals are important modulators of the transcriptional activity of genes, and subsequently, altered gene expression has been linked to environmentally-induced phenotypes (Sumner et al. 2006; Kent et al. 2009).

Consequently, despite the extensive empirical and theoretical research, the causes and consequences of plasticity are still poorly understood and an interdisciplinary approach with application of new techniques such as microarrays, next generation sequencing and proteomics is expected to contribute to this knowledge (Aubin-Horth and Renn 2009; Ellers and Stuefer 2010). Although these post-genomic techniques have been widely used in genetic model organisms, substantial further progress awaits their successful application to non-model organisms commonly used as ecological targets for studying the processes and mechanisms of phenotypic plasticity (Aubin-Horth and Renn 2009).

1.6.1. Response to wave action

Wave energy on rocky intertidal shores (Fig. 11) creates an extremely heterogeneous environment (Trussell 1997). Because of the strong correlation between readily measured phenotype and distinct environmental gradient, the gastropod-wave exposure system has been widely used for studying the evolution of phenotypic plasticity (Kitching et al. 1966; Trussell & Etter 2001).

Shell shape phenotypic variation in N. lapillus is influenced by genotype-environment interactions (Hughes and Taylor 1997; Trussell and Etter 2001). The shell of N. lapillus is more globular at sites exposed to wave action and more elongated at sheltered sites (Kitching et al. 1966). The exposed-site shape possibly offers less drag (Hughes and Taylor 1997) and is characterized by a relatively larger, more rounded aperture (Kirby et al. 1994) that accommodates a larger foot, affording stronger attachment to the rock and therefore greater resistance to dislodgement by waves (Kitching et al. 1966; Etter 1988). The sheltered-site shape is associated with slower evaporation by having a relatively smaller aperture (Coombs 1973) and with greater capacity for evaporative cooling through holding a relatively greater volume of extra-corporeal water within the basal whorl (Kirby et al. 1994).

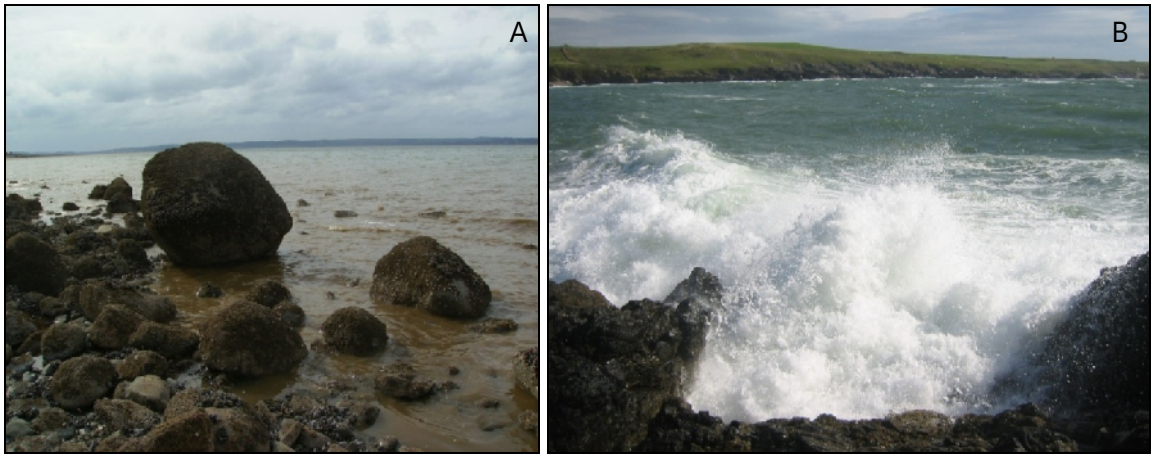


Figure 11. View of a sheltered shore (A: Llanfairfechan, North Wales, UK) and an exposed shore (B: Cable Bay, Anglesey, UK). Note: These were the experimental sites selected for the present work.

1.6.2. Response to crab predation

Some of the most frequently studied examples of phenotypic plasticity are predator-induced defences (DeWitt and Scheiner 2004), mainly focused on single traits and single predators but also evaluating responses to combined impacts of multiple predators (Bourdeau 2009).

Many marine invertebrates respond to water-borne alarm substances, either released by the predator or by damaged prey, showing morphological or behaviour alterations in order to reduce vulnerability to predation (Wisenden 2000; Griffiths and Richardson 2006). Examples of such responses are the production of defensive spines in bryozoans (Harvell 1984), extension and opening of pedicellaria in sea urchins (Phillips 1978), changes in shell morphology in barnacles (Lively 1986), increased byssus thread production in mussels (Fassler and Kaiser 2008), increased burial depths in burrowing bivalves (Griffiths and Richardson 2006) and reduced growth and thickening of the shell lip in whelks (Appleton and Palmer 1988; Palmer 1990; Rawlings 1994). In N. lapillus, a relatively narrow aperture of the elongated sheltered shells in combination with thickened shell walls (Currey and Hughes 1982) hinders attacks by crabs, which tend to be abundant at sheltered sites (Fig. 12) but rare at exposed (Hughes and Elner 1979).

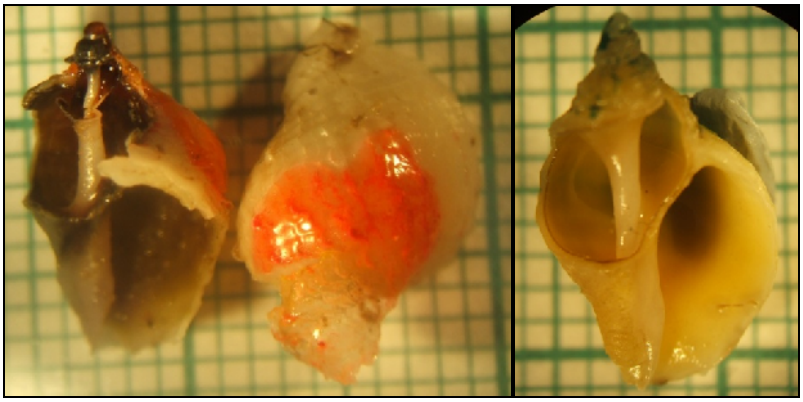


Figure 12. Evidence of crab predation over N. lapillus from a sheltered shore in North Wales, UK.

The use of anti-predator defences significantly reduce predator attack success and therefore directly increases survival, and ultimately, fitness (Smith and Jennings 2000; Reimer and Harms-Ringdahl 2001).

1.7. NUCELLA LAPILLUS: AN ECOLOGICAL MODEL SPECIES

The dogwhelk N. lapillus (Fig. 13), family Muricidae, is a predatory snail - mainly feeding on mussels and barnacles - with limited dispersal ability owing to non-planktonic larvae and a restricted adult ambit size with aggregative behaviour throughout most of their life cycle (Crothers 1985). N. lapillus is a gonochoristic (unisexual) neogastropod with internal fertilization and direct development within egg capsules (Spight 1975). The young emerge at the crawling stage complete with shells, like an adult in miniature (Crothers 1977; Etter 1996). A newly deposited capsule may contain up to 600 eggs but only a few (15-30) will complete its development while the others may act as food for embryos.

Species with planktonic larvae are expected to show higher levels of gene flow and less genetic structure than direct developers (Chambers et al. 1996; De Wolf et al. 2000; Féral 2002). Consequently, it is expected that N. lapillus has limited dispersal ability and restricted gene flow with evidence for significant population differentiation. Indeed, depending on local hydrography and long-shore distribution of suitable habitat, N. lapillus may show significant population differentiation on a scale of kilometres (Day and Bayne 1988; Kirby et al. 1997; Kirby 2000; McInerney et al. 2009) or even metres (Day 1990; Goudet et al. 1994; Guerra-Varela et al. 2009). However, contrary to expectations for a slowly moving developer (Chambers et al. 1996; De Wolf et al. 2000; Féral 2002), previous genetic studies have shown N. lapillus to be capable of dispersal over several kilometres or more, probably by early juveniles drifting while attached to buoyant mucous threads or debris (Colson and Hughes 2004; Colson et al. 2006).

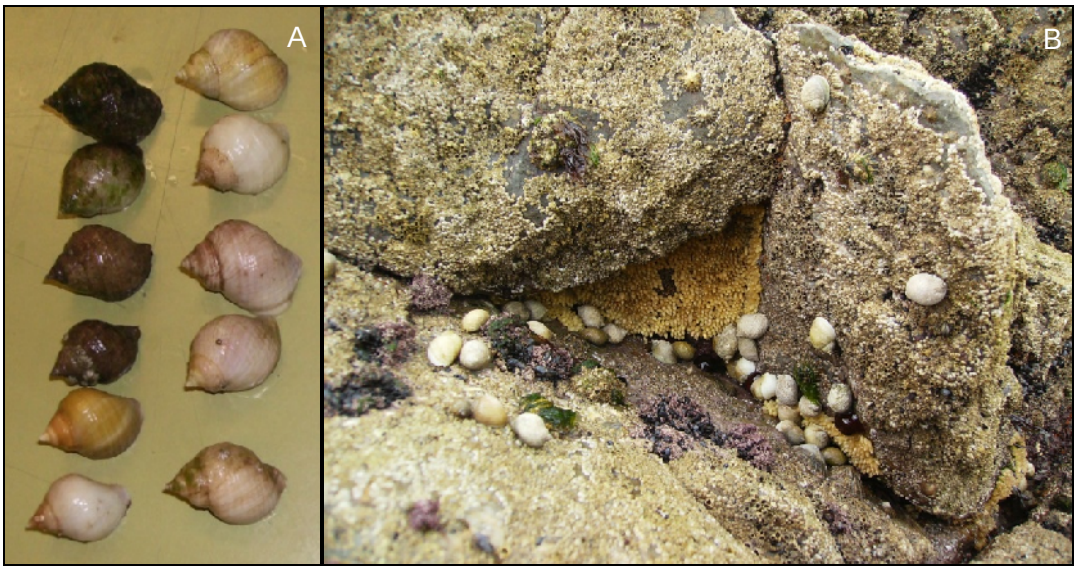


Figure 13. Nucella lapillus A) Adult shell polymorphism; B) Example of N. lapillus aggregative behaviour showing adults and egg capsules (photographs taken at Cable Bay, Anglesey, UK).

N. lapillus is commonly found on rocky shores and estuaries of the North Atlantic (Fig. 14), ranging from the most exposed to wave action to the most sheltered (Crothers 1985).

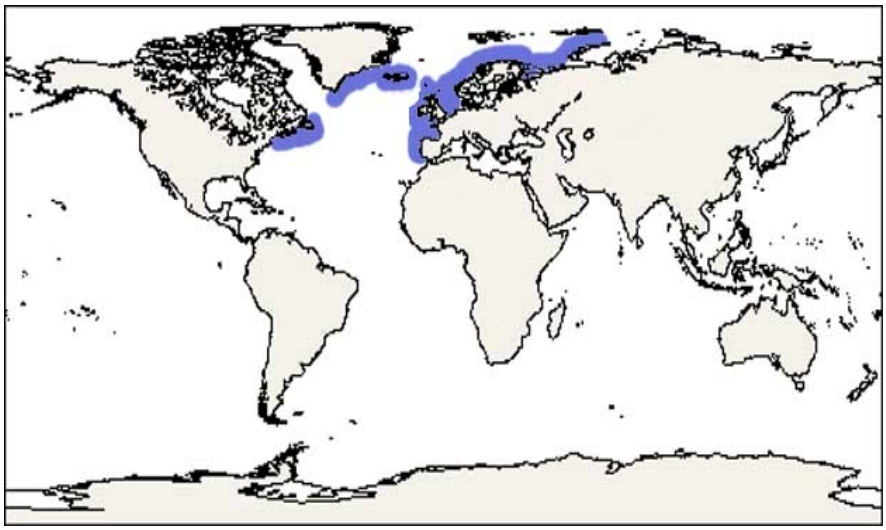


Figure 14. Nucella lapillus geographical distribution (source: http://www.marlin.ac.uk).

Dogwhelks show marked intraspecific variation in shell colour, shape and thickness, presenting more globular shells at sites exposed to wave action and more elongated shells at sheltered sites (Kitching et al. 1966). With marked evidence of

polymorphism in shell shape and the presumed low rates of gene flow promoting fine-scale adaptive differentiation, N. lapillus provides a suitable system for studying the evolution of phenotypic plasticity.

N. lapillus is not only a key species for ecological and evolutionary studies, but it is widely recognised as a sentinel organism to TBT pollution and is widely used for biomonitoring via the imposex response.

As a result, N. lapillus was the selected species for the present study as it is an ecological model organism for both TBT contamination and wave action responses. Additionally, its wide distribution and abundance in nature, easy maintenance under laboratory conditions, and lack of commercial importance made N. lapillus the best candidate for the present study.

1.8. THESIS SCOPE AND OBJECTIVES

The present PhD thesis is written under the overarching framework of ecological genomics, where the genetic mechanisms underlying responses of organisms to their natural environment are explored.

One example of response to anthropogenic environmental change (imposex) and one example of response to natural environmental change (shell shape phenotypic plasticity) were explored.

Despite the extensive empirical and theoretical research on both endocrine disruption (imposex) and phenotypic plasticity subjects, their mechanisms are still poorly understood. Thus, in this thesis, an interdisciplinary approach using ecological analysis such as determination of imposex levels, reciprocal transplants and morphometrics, alongside with the application of new genomic techniques such as next-generation sequencing technologies applied to transcriptome and gene expression analysis using microarrays, were employed aiming to contribute to a better understanding of both of these phenomena.

The main aim of this thesis is to study alterations induced in dogwhelks by natural and man-made environmental change by:

i) Apply combinations of next-generation sequencing (454 Roche sequencing) and microarray technologies to develop N. lapillus genomic resources and to apply these tools for exploration of N. lapillus response to TBT pollution at the genomic level;

ii) Explore N. lapillus shell shape phenotypic plasticity in response to wave action and crab predation using insights from reciprocal transplants and common garden experiments.

To achieve the main goal, the following work pipelines (Fig. 15 and Fig. 16) were followed:

i) Investigate the genetic mechanism behind imposex development induced by tributyltin.

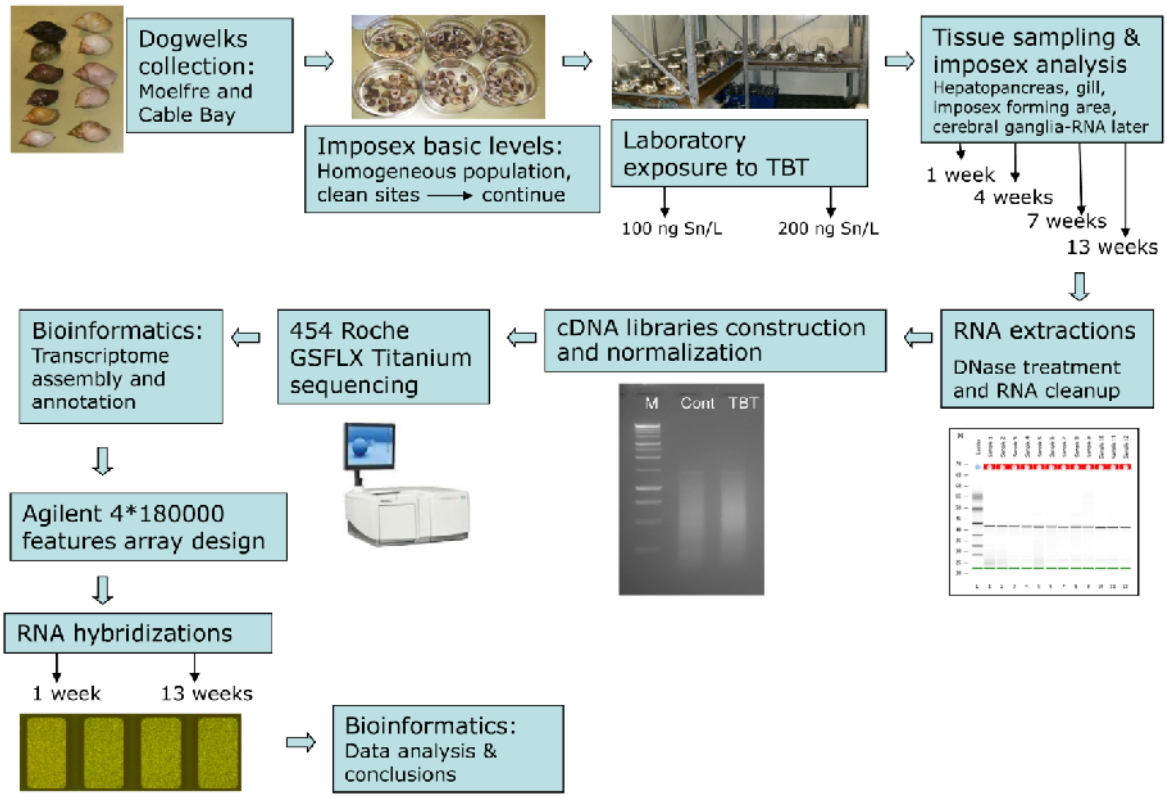


Figure 15. Molecular pipeline adopted for imposex mechanism investigation.

ii) Explore shell shape phenotypic plasticity of N. lapillus in response to wave exposure and crab predation.

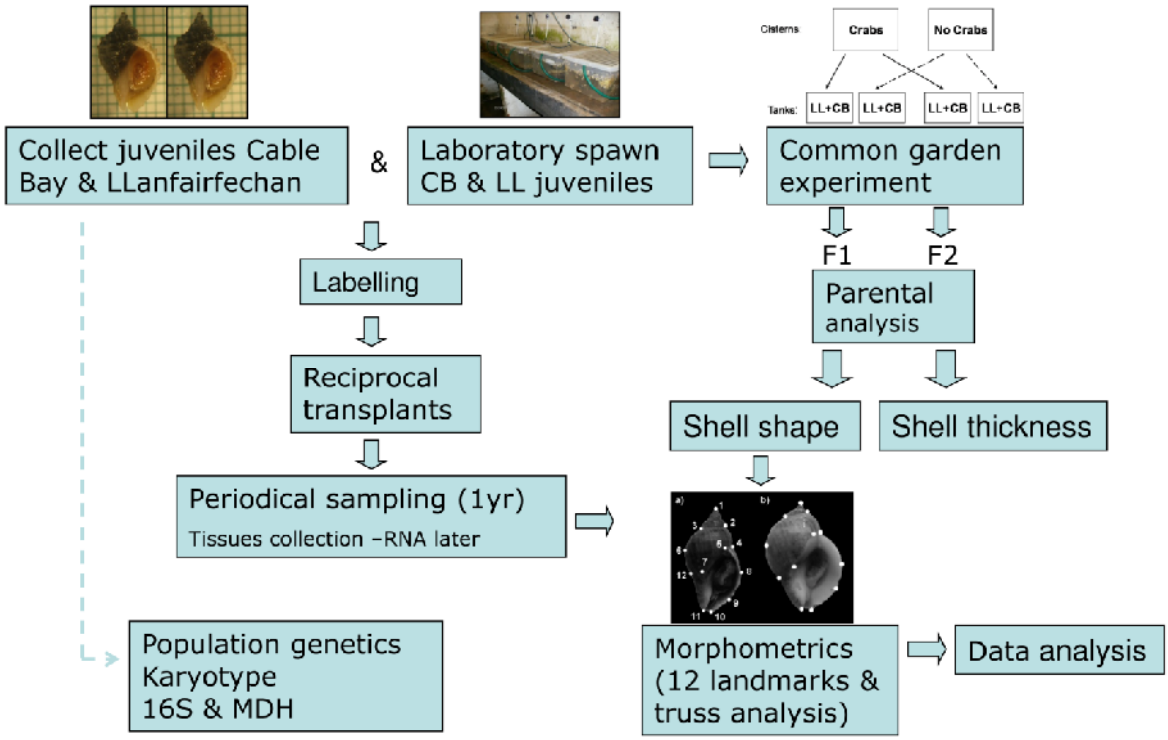


Figure 16. Shell shape phenotypic plasticity adopted work pipeline.

1.8.1. Thesis organization:

Chapter 1 provides a brief theoretical overview concerning the response to environmental change, particularly endocrine disruption and phenotypic plasticity. A brief summary on next-generation sequencing and microarray technologies is also introduced.

In chapter 2 a de novo N. lapillus transcriptome sequencing analysis using 454 Roche sequencing is described, followed by transcriptome assembly, BLAST and gene ontology annotations. Further development of about 2000 EST-molecular markers and quantification of genome similarity between N. lapillus and Lottia gigantea is also explored.

Chapter 3 describes how the transcriptomic data obtained in chapter 2 was used to design an N. lapillus 180000 features oligoarray and how it was used to test differential gene expression levels in response to TBT contamination.

Following molecular finding from chapter 3, in chapter 4 a new hypothesis for the induction of imposex is tested.

In chapter 5 environmental and genetic correlation of N. lapillus shell shape phenotypic plasticity was explored through reciprocal transplants and common garden experiments (F1 and F2 generations) with and without risk of predation. Population genetics of the studied populations using microsatellite, nuclear and mitochondrial markers, and karyotype analysis are described. Morphometrics as a tool to explore shell shape and parental analysis of common garden snails are also explored.

Chapter 6 provides a general discussion of all the results obtained along this study.

1.9. REFERENCES

Alzieu, C. 1998. Tributyltin: case study of a chronic contaminant in the coastal environment. Ocean & Coastal Management 40:23-36.

Alzieu, C. 2000. Impact of tributyltin on marine invertebrates. Ecotoxicology 9:71-76.

Appleton, R. D., and A. R. Palmer. 1988. Water-Borne Stimuli Released by Predatory Crabs and Damaged Prey Induce More Predator-Resistant Shells in a Marine Gastropod. Proceedings of the National Academy of Sciences of the United States of America 85:4387-4391.

Atkins, K. E., and J. M. J. Travis. 2010. Local adaptation and the evolution of species' ranges under climate change. Journal of Theoretical Biology 266:449-457.

Aubin-Horth, N., and S. C. P. Renn. 2009. Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. Molecular Ecology 18:3763-3780.

Bainbridge, M. N., R. L. Warren, M. Hirst, T. Romanuik, T. Zeng, A. Go, A. Delaney, M. Griffith, M. Hickenbotham, V. Magrini, E. R. Mardis, M. D. Sadar, A. S. Siddiqui, M. A. Marra, and S. J. M. Jones. 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. Bmc Genomics 7.

Barbazuk, W. B., S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable. 2007. SNP discovery via 454 transcriptome sequencing. Plant Journal 51:910-918.

Beldade, P., S. Rudd, J. D. Gruber, and A. D. Long. 2006. A wing expressed sequence tag resource for Bicyclus anynana butterflies, an evo-devo model. Bmc Genomics 7.

Bettin, C., J. Oehlmann, and E. Stroben. 1996. TBT-induced imposex in marine neogastropods is mediated by an increasing androgen level. Helgolander Meeresuntersuchungen 50:299-317.

Binladen, J., M. T. P. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev. 2007. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. Plos One 2.

Bouck, A., and T. Vision. 2007. The molecular ecologist's guide to expressed sequence tags. Molecular Ecology 16:907-924.

Bourdeau, P. E. 2009. Prioritized phenotypic responses to combined predators in a marine snail. Ecology 90:1659-1669.

Boyer, I. J. 1989. Toxicity of dibutyltin, tributyltin and other organotin compounds to humans and to experimental animals. Toxicology 55:253-298.

Bradshaw, A. D. 1965. Evolutionary significance of phenotypic plasticity in plants. Advances in Genetics 13:115-155.

Bryan, G. W., and P. E. Gibbs. 1991. Impact of low concentrations of Tributyltin (TBT) on Marine Organisms: a Review. In Newman MC, McIntosh AW (eds) Metal ecotoxicology: concepts and applications. Lewis Publishers, Ann Arbor, Boston, pp 323 - 361.

Bryan, G. W., P. E. Gibbs, L. G. Hummerstone, and G. R. Burt. 1986. The decline of the gastropod Nucella lapillus around Southwest England - evidence for the effect of tributyltin from antifouling paints. Journal of the Marine Biological Association of the United Kingdom 66:611-640.

Callaway, R. M., S. C. Pennings, and C. L. Richards. 2003. Phenotypic plasticity and interactions among plants. Ecology 84:1115-1128.

Castro, L. F. C., D. Lima, A. Machado, C. Melo, Y. Hiromori, J. Nishikawa, T. Nakanishi, M. A. Reis-Henriques, and M. M. Santos. 2007. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod Nucella lapillus. Aquatic Toxicology 85:57-66.

Chambers, R. J., C. D. McQuaid, and R. Kirby. 1996. Determination of genetic diversity of South African intertidal limpets (Gastropoda: Siphonaria) with different reproductive modes using polyacrylamide gel electrophoresis of total cellular proteins. Journal of Experimental Marine Biology and Ecology 201:1-11.

Colborn, T., F. S. V. Saal, and A. M. Soto. 1993. Developmental effects of endocrine disrupting chemicals in wildlife and humans. Environmental Health Perspectives 101:378-384.

Colson, I., J. Guerra-Varela, R. N. Hughes, and E. Rolan-Alvarez. 2006. Using molecular and quantitative variation for assessing genetic impacts on Nucella lapillus populations after local extinction and recolonization. Integrative Zoology 1:104-107.

Colson, I., and R. N. Hughes. 2004. Rapid recovery of genetic diversity of dogwhelk (Nucella lapillus L.) populations after local extinction and recolonization contradicts predictions from life-history characteristics. Molecular Ecology 13:2223-2233.

Coombs, V. A. 1973. Desiccation and age as factors in vertical distribution of dog-whelk, Nucella-lapillus. Journal of Zoology 171:57-66.

Cossins, A., J. Fraser, M. Hughes, and A. Gracey. 2006. Post-genomic approaches to understanding the mechanisms of environmentally induced phenotypic plasticity. Journal of Experimental Biology 209:2328-2336.

Crothers, J. H. 1977. Some observations on growth of common dog-whelk, Nucella lapillus (Prosobranchia-Muricacea) in laboratory. Journal of Conchology 29:157-162.

Crothers, J. H. 1985. Dog-whelks an introduction to the biology of Nucella-lapillus. Field Studies 6:291-360.

Currey, J. D., and R. N. Hughes. 1982. Strength of the dogwhelk Nucella-lapillus and the winkle Littorina-littorea from different habitats. Journal of Animal Ecology 51:47-56.

Day, T., J. Pritchard, and D. Schluter. 1994. A comparison of 2 sticklebacks. Evolution 48:1723-1734.

Day, A. J. 1990. Microgeographic variation in allozyme frequencies in relation to the degree of exposure to wave action in the dogwelk Nucella lapillus (L) (Prosobranchia, Muricacea). Biological Journal of the Linnean Society 40:245-261.

Day, A. J., and B. L. Bayne. 1988. Allozyme variation of the dos-whelk Nucella lapillus (Prosobranchia, Muricacea) from the South West peninsula of England. Marine Biology 99:93-100.

De Mora, S. J., and E. Pelletier. 1997. Environmental tributyltin research: Past, present, future. Environmental Technology 18:1169-1177.

De Wolf, H., R. Verhagen, and T. Backeljau. 2000. Large scale population structure and gene flow in the planktonic developing periwinkle, Littorina striata, in Macaronesia (Mollusca : Gastropoda). Journal of Experimental Marine Biology and Ecology 246:69-83.

deFur, P. L. 2004. Use and role of invertebrate models in endocrine disruptor research and testing. Ilar Journal 45:484-493.

DeWitt, T. J. 1998. Costs and limits of phenotypic plasticity: Tests with predator-induced morphology and life history in a freshwater snail. Journal of Evolutionary Biology 11:465-480.

DeWitt, T. J., and S. M. Scheiner. 2004. Phenotypic plasticity: functional and conceptual approaches.Oxford University Press 1-247.

Domenici, P., H. Turesson, J. Brodersen, and C. Bronmark. 2008. Predator-induced morphology enhances escape locomotion in crucian carp. Proceedings of the Royal Society B-Biological Sciences 275:195-201.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17:1629-1631.

Ellers, J., and J. F. Stuefer. 2010. Frontiers in phenotypic plasticity research: new questions about mechanisms, induced responses and ecological impacts. Evolutionary Ecology 24:523-526.

Elmer, K. R., S. Fan, H. M. Gunter, J. C. Jones, S. Boekhoff, S. Kuraku, and A. Meyer. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Molecular Ecology 19:197-211.

Emrich, S. J., W. B. Barbazuk, L. Li, and P. S. Schnable. 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research 17:69-73.

Endler, J. A. 1986. Monographs in population biology 21. Natural selection in the wild. Princeton University Press: Princeton, N.J., USA. Illus:XIII+336P.

Etter, R. J. 1988. Asymmetrical Developmental Plasticity in an Intertidal Snail. Evolution 42:322-334.

Etter, R. J. 1996. The effect of wave action, prey type, and foraging time on growth of the predatory snail Nucella lapillus (L). Journal of Experimental Marine Biology and Ecology 196:341-356.

Evans, S. M., and C. R. C. Sheppard. 2000. Marine antifoulants. Seas at the millennium: an environmental evaluation. Volume 3: global issues and processes.:247-256.

Fassler, S. M. M., and M. J. Kaiser. 2008. Genetically mediated anti-predator responses in bivalve molluscs. Marine Ecology-Progress Series 363:217-225.

Fent, K. 2004. Ecotoxicological effects at contaminated sites. Toxicology 205:223-240.

Féral, C., C. le Gall, J. Lever, and H. H. Boer. 1983. The influence of a pollutant factor (tributylin) on the neuroendocrine mechanism responsible for the occurrence of a penis in the females of Ocenebra erinacea. Molluscan neuro-endocrinology. Proceedings of the International Minisymposium on Molluscan Endocrinology, held in the Department of Biology, Free University, Amsterdam, the Netherlands, August 16-20, 1982.:173-175.

Féral, J. P. 2002. How useful are the genetic markers in attempts to understand and manage marine biodiversity? Journal of Experimental Marine Biology and Ecology 268:121-145.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail Littorina saxatilis. Journal of Evolutionary Biology 23:2004-2016.

Gibbs, P. E., G. W. Bryan, and P. L. Pascoe. 1991. Tbt-Induced Imposex in the Dogwhelk, Nucella-Lapillus - Geographical Uniformity of the Response and Effects. Marine Environmental Research 32:79-87.

Gibbs, P. E., G. W. Bryan, P. L. Pascoe, and G. R. Burt. 1987. The use of the dog-whelk, Nucella lapillus, as an indicator of tributyltin (TBT) contamination. Journal of the Marine Biological Association of the United Kingdom 67:507-523.

Goldberg, E. D. 1986. TBT – An environmental dilemma. Environment 28:17-22.

Gomez-Mestre, I., and M. Tejedo. 2004. Contrasting patterns of quantitative and neutral genetic variation in locally adapted populations of the natterjack toad, Bufo calamita. Evolution 58:2343-2352.

Gooding, M. P., and G. A. LeBlanc. 2001. Biotransformation and disposition of testosterone in the eastern mud snail Ilyanassa obsoleta. General and comparative endocrinology 122:172-180.

Gooding, M. P., and G. A. LeBlanc. 2004. Seasonal variation in the regulation of testosterone levels in the eastern mud snail (Ilyanassa obsoleta). Invertebrate Biology 123:237-243.

Gooding, M. P., V. S. Wilson, L. C. Folmar, D. T. Marcovich, and G. A. LeBlanc. 2003. The biocide tributyltin reduces the accumulation of testosterone as fatty acid esters in the mud snail (Ilyanassa obsoleta). Environmental Health Perspectives 111:426-430.

Goudet, J., T. Demeeus, A. J. Day, and C. J. Gliddon. 1994. The different levels of populations structuring of the dogwhelk, Nucella lapillus, along the South Devon Coast. Genetics and Evolution of Aquatic Organisms:81-95.

Gould, S. J. 1966. Allometry and size in ontogeny and phylogeny. Biological Reviews of the Cambridge Philosophical Society 41:587-640.

Gracey, A. Y. 2007. Interpreting physiological responses to environmental change through gene expression profiling. Journal of Experimental Biology 210:1584-1592.

Gracey, A. Y., and A. R. Cossins. 2003. Application of microarray technology in environmental and comparative physiology. Annual Review of Physiology 65:231-259.

Graveley, B. R. 2008. Molecular biology - Power sequencing. Nature 453:1197-1198.

Griffiths, C. L., and C. A. Richardson. 2006. Chemically induced predator avoidance behaviour in the burrowing bivalve Macoma balthica. Journal of Experimental Marine Biology and Ecology 331:91-98.

Grun, F., H. Watanabe, Z. Zamanian, L. Maeda, K. Arima, R. Cubacha, D. M. Gardiner, J. Kanno, T. Iguchi, and B. Blumberg. 2006. Endocrine-disrupting organotin compounds are potent inducers of adipogenesis in vertebrates. Molecular Endocrinology 20:2141-2155.

Guerra-Varela, J., I. Colson, T. Backeljau, K. Breugelmans, R. N. Hughes, and E. Rolan-Alvarez. 2009. The evolutionary mechanism maintaining shell shape and molecular differentiation between two ecotypes of the dogwhelk Nucella lapillus. Evolutionary Ecology 23:261-280.

Guillette, L. J., and B. C. Moore. 2006. Environmental contaminants, fertility, and multioocytic follicles: A lesson from wildlife? Seminars in Reproductive Medicine 24:134-141.

Harvell, C. D. 1984. Predator-induced defense in a marine bryozoan. Science 224:1357-1359.

Heidrich, D. D., S. Steckelbroeck, and D. Klingmuller. 2001. Inhibition of human cytochrome P450 aromatase activity by butyltins. Steroids 66:763-769.

Hoch, M. 2001. Organotin compounds in the environment - an overview. Applied Geochemistry 16:719-743.

Hollander, J. 2008. Testing the grain-size model for the evolution of phenotypic plasticity. Evolution 62:1381-1389.

Hori, M. 1993. Frequency-dependent natural-selection in the handedness of scale-eating cichlid fish. Science 260:216-219.

Horiguchi, T., T. Nishikawa, Y. Ohta, H. Shiraishi, and M. Morita. 2007. Retinoid X receptor gene expression and protein content in tissues of the rock shell Thais clavigera. Aquatic Toxicology 84:379-388.

Hudson, M. E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular Ecology Resources 8:3-17.

Huet, M., Y. M. Paulet, and M. Glemarec. 1996. Tributyltin (TBT) pollution in the coastal waters of west Brittany as indicated by imposex in Nucella lapillus. Marine Environmental Research 41:157-167.

Hughes, R. N., and R. W. Elner. 1979. Tactics of a predator, Carcinus-maenas, and morphological responses of the prey, Nucella-lapillus. Journal of Animal Ecology 48:65-78.

Hughes, R. N., and M. J. Taylor. 1997. Genotype-environment interaction expressed in the foraging behaviour of dogwhelks, Nucella lapillus (L), under simulated environmental hazard. Proceedings of the Royal Society of London Series B-Biological Sciences 264:417-422.

Iguchi, T., and Y. Katsu. 2008. Commonality in Signaling of Endocrine Disruption from Snail to Human. Bioscience 58:1061-1067.

Janer, G., J. Bachmann, J. Oehlmann, U. Schulte-Oehlmann, and C. Porte. 2006. The effect of organotin compounds on gender specific androstenedione metabolism in the freshwater ramshorn snail Marisa cornuarietis. The Journal of Steroid Biochemistry and Molecular Biology 99:147-156.

Janer, G., R. M. Sternberg, G. A. LeBlanc, and C. Porte. 2005. Testosterone conjugating activities in invertebrates: are they targets for endocrine disruptors? Aquatic Toxicology 71:273-282.

Jarvie, T., and T. Harkins. 2008. De novo assembly and genomic structural variation analysis with genome sequencer FLX 3K long-tag paired end reads. Biotechniques 44:829-831.

Kalujnaia, S., I. S. McWilliam, V. A. Zaguinaiko, A. L. Feilen, J. Nicholson, N. Hazon, C. P. Cutler, R. J. Balment, A. R. Cossins, M. Hughes, and G. Cramb. 2007. Salinity adaptation and gene profiling analysis in the European eel (Anguilla anguilla) using microarray technology. General and Comparative. Endocrinology 152:274-280.

Kent, C. F., T. Daskalchuk, L. Cook, M. B. Sokolowski, and R. J. Greenspan. 2009. The Drosophila foraging Gene Mediates Adult Plasticity and Gene-Environment Interactions in Behaviour, Metabolites, and Gene Expression in Response to Food Deprivation. Plos Genetics 5.

Ketata, I., X. Denier, A. Hamza-Chaffai, and C. Minier. 2008. Endocrine-related reproductive effects in molluscs. Comparative Biochemistry and Physiology, Part C-Toxicology and Pharmacology 147:261-270.

Kirby, R. R. 2000. An ancient transpecific polymorphism shows extreme divergence in a multitrait cline in an intertidal snail (Nucella lapillus (L.)). Molecular Biology and Evolution 17:1816-1825.

Kirby, R. R., B. L. Bayne, and R. J. Berry. 1994. Penotypic variation along a cline in allozyme and karyotype frequencies, and its relationship with habitat, in the dog-welk Nucella-lapillus, L. Biological Journal of the Linnean Society 53:255-275.

Kirby, R. R., R. J. Berry, and D. A. Powers. 1997. Variation in mitochondrial DNA in a cline of allele frequencies and shell phenotype in the dog-whelk Nucella lapillus (L.). Biological Journal of the Linnean Society 62:299-312.

Kitching, J. A., L. Muntz, and F. J. Ebling. 1966. Ecology of Lough Ine. 15. Ecological significance of shell and body forms in Nucella. Journal of Animal Ecology 35:113-126.

Landry, C. R., and N. Aubin-Horth. 2007. Ecological annotation of genes and genomes through ecological genomics. Molecular Ecology 16:4419-4421.

Laughlin, R. B., W. French, and H. E. Guard. 1986. Accumulation of bis(tributyltin) oxide by the marine mussel Mytillus edulis. Environmental Science & Technology 20:884-890.

Lively, C. M. 1986. Predator-induced shell dimorphism in the acorn barnacle Chthamalus anisopoma. Evolution 40:232-242.

Lockhart, D. J., H. L. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. W. Wang, M. Kobayashi, H. Horton, and E. L. Brown. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology 14:1675-1680.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Martinez-Fernandez, M., M. de la Cadena, and E. Rolan-Alvarez. 2010. The role of phenotypic plasticity on the proteome differences between two sympatric marine snail ecotypes adapted to distinct micro-habitats Bmc Evolutionary Biology 10.

Matthiessen, P., and P. E. Gibbs. 1998. Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. Environmental Toxicology and Chemistry 17:37-43.

McCarthy, A. 2010. Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology. Chemistry & Biology 17:675-676.

McInerney, C. E., A. L. Allcock, M. P. Johnson, and P. A. Prodohl. 2009. Understanding marine reserve function in a seascape genetics context: Nucella lapillus in Strangford Lough (Northern Ireland) as an example. Aquatic Biology 7:45-58.

Meyer, M., U. Stenzel, S. Myles, K. Prufer, and M. Hofreiter. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. Nucleic Acids Research 35.

Mueller, S. O. 2004. Xenoestrogens: mechanisms of action and detection methods. Analytical and Bioanalytical Chemistry 378:582-587.

Murata, S., S. Takahashi, T. Agusa, N. J. Thomas, K. Kannan, and S. Tanabe. 2008. Contamination status and accumulation profiles of organotins in sea otters (Enhydra lutris) found dead along the coasts of California, Washington, Alaska (USA), and Kamchatka (Russia). Marine Pollution Bulletin 56:641-649.

Nishikawa, J. 2006. Imposex in marine gastropods may be caused by binding of organotins to retinoid X receptor. Marine Biology 149:117-124.

Nishikawa, J., S. Mamiya, T. Kanayama, T. Nishikawa, F. Shiraishi, and T. Horiguchi. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Environmental Science and Technology 38:6271-6276.

Nussey, D., A. Wilson, and J. Brommer. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20:831-844.

Oberdorster, E., and P. McClellan-Green. 2000. The neuropeptide APGWamide induces imposex in the mud snail, Ilyanassa obsoleta. Peptides 21:1323-1330.

Oberdorster, E., and P. McClellan-Green. 2002. Mechanisms of imposex induction in the mud snail, Ilyanassa obsoleta: TBT as a neurotoxin and aromatase inhibitor. Marine Environmental Research 54:715-718.

Oetken, M., J. Bachmann, U. Schulte-Oehlmann, and J. Oehlmann. 2004. Evidence for endocrine disruption in invertebrates. International Review of Cytology 236:1-44.

Ogata, R., M. Omura, Y. Shimasaki, K. Kubo, Y. Oshima, S. Aou, and N. Inoue. 2001. Two-generation reproductive toxicity study of tributyltin chloride in female rats. Journal of Toxicology and Environmental Health-Part A 63:127-144.

Oliveira, I. B., C. A. Richardson, A. C. Sousa, S. Takahashi, S. Tanabe, and C. M. Barroso. 2009. Spatial and temporal evolution of imposex in dogwhelk Nucella lapillus (L.) populations from North Wales, UK. Journal of Environmental Monitoring 11:1462-1468.

Omura, M., R. Ogata, K. Kubo, Y. Shimasaki, S. Aou, Y. Oshima, A. Tanaka, M. Hirata, Y. Makita, and N. Inoue. 2001. Two-generation reproductive toxicity study of tributyltin chloride in male rats. Toxicological Sciences 64:224-232.

OSPAR. 1998. Joint Assessment and Monitoring Programme (JAMP), Guidelines for contaminant specific biological effects monitoring, Reference Number 1998–3, OSPAR Commision, London, UK.

Palmer, A. R. 1990. Effect of Crab Effluent and Scent of Damaged Conspecifics on Feeding, Growth, and Shell Morphology of the Atlantic Dogwhelk Nucella-Lapillus (L). Hydrobiologia 193:155-182.

Patino, J., K. Hylander, and J. M. Gonzalez-Mancebo. 2010. Effect of forest clear-cutting on subtropical bryophyte communities in waterfalls, on dripping walls, and along streams. Ecological Applications 20:1648-1663.

Pfrender, M. E., K. Spitze, J. Hicks, K. Morgan, L. Latta, and M. Lynch. 2000. Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. Conservation Genetics 1:263-269.

Phillips, D. W. 1978. Chemical mediation of invertebrate defensive behaviors and ability to distinguish between foraging and inactive predators. Marine Biology 49:237-243.

Pigliucci, M. 2001. Syntheses in Ecology and Evolution. Phenotypic plasticity: Beyond nature and nuture. 1-328.

Pigliucci, M. 2005. Evolution of phenotypic plasticity: where are we going now? Trends in Ecology & Evolution 20:481-486.

Porte, C., G. Janer, L. C. Lorusso, M. Ortiz-Zarragoitia, M. P. Cajaraville, M. C. Fossi, and L. Canesi. 2006. Endocrine disruptors in marine organisms: approaches and perspectives. Comparative Biochemestry and Physiology C- Toxicology and Pharmacology 143:303-315.

Rato, M., N. Ferreira, J. Santos, and C. Barroso. 2009. Temporal evolution of imposex in Nassarius reticulatus (L.) along the Portuguese coast: the efficacy of EC regulation 782/2003. Journal of Environmental Monitoring 11:100-107.

Raven, P., G. Johnson, J. Losos, and S. Singer. 2005. Biology. McGraw-Hill

Rawlings, T. A. 1994. Effect of elevated predation risk on the metabolic-rate and spawning intensity of a rocky shore marine gastropod. Journal of Experimental Marine Biology and Ecology 181:67-79.

Reimer, O., and S. Harms-Ringdahl. 2001. Predator-inducible changes in blue mussels from the predator-free Baltic Sea. Marine Biology 139:959-965.

Ronis, M. J. J., and A. Z. Mason. 1996. The metabolism of testosterone by the periwinkle (Littorina littorea) in vitro and in vivo: Effects of tributyl tin. Marine Environmental Research 42:161-166.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA Sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74:5463-5467.

Santillo, D., Johnston, P., Langston, W.J. 2002. Tributyltin (TBT) antifoulants: a tale of ships, snails and imposex. In: Late lessons from early warnings: the precautionary principle 1896-2000, Harremoës, P., Gee, D., MacGarvin, M., Stirling, A., Keys, J., Wynne, B., Guedes Vaz, S., eds., European Environmental Agency, Copenhagen, 135-148.

Santos, M. M., L. F. C. Castro, M. N. Vieira, J. Micael, R. Morabito, P. Massanisso, and M. A. Reis-Henriques. 2005. New insights into the mechanism of imposex induction in the dogwhelk

Nucella lapillus. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 141:101-109.

Santos, M. M., M. A. Reis-Henriques, M. N. Vieira, and M. Sole. 2006. Triphenyltin and tributyltin, single and in combination, promote imposex in the gastropod Bolinus brandaris. Ecotoxicology and Environmental Safety 64:155-162.

Santos, M. M., C. C. ten Hallers-Tjabbes, N. Vieira, J. P. Boon, and C. Porte. 2002. Cytochrome P450 differences in normal and imposex-affected female whelk Buccinum undatum from the open North Sea. Marine Environmental Research 54:661-665.

Scheiner, S. M. 1993. Genetics and evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 24:35-68.

Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: A reaction norm perspective. Phenotypic evolution: A reaction norm perspective:xii+387p.

Smith, B. S. 1971. Sexuality in the American mud snail Nassarius obsoletus. Proceedings of the Malacological Society of London 39:377-378.

Smith, L. D., and J. A. Jennings. 2000. Induced defensive responses by the bivalve Mytilus edulis to predators with different attack modes. Marine Biology 136:461-469.

Smith, T. B. 1993. Disruptive selection and the genetic-basis of bill size polymorphism in the African finch pyrenestes. Nature 363:618-620.

Soin, T., and G. Smagghe. 2007. Endocrine disruption in aquatic insects: a review. Ecotoxicology 16:83-93.

Sousa, A., C. Matsudaira, S. Takahashi, S. Tanabe, and C. Barroso. 2007. Integrative assessment of organotin contamination in a southern European estuarine system (Ria de Aveiro, NW Portugal): tracking temporal trends in order to evaluate the effectiveness of the EU ban. Marine Pollution Bulletin 54:1645-1653.

Spade, D. J., R. J. Griffitt, L. Liu, N. J. Brown-Peterson, K. J. Kroll, A. Feswick, R. A. Glazer, D. S. Barber, and N. D. Denslow. 2010. Queen Conch (Strombus gigas) Testis Regresses during the Reproductive Season at Nearshore Sites in the Florida Keys. Plos One 5.

Spight, T. M. 1975. Factors extending gastropod embryonic development and their selective cost. Oecologia 21:1-16.

Spooner, N., P. E. Gibbs, G. W. Bryan, and L. J. Goad. 1991. The Effect of Tributyltin Upon Steroid Titers in the Female Dogwhelk, Nucella-Lapillus, and the Development of Imposex. Marine Environmental Research 32:37-49.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25:705-712.

Stearns, S. C. 1989. The evolutionary significance of phenotypic plasticity - phenotypic sources of variation among organisms can be described by developmental switches and reaction norms. Bioscience 39:436-445.

Sternberg, R. M., M. P. Gooding, A. K. Hotchkiss, and G. A. LeBlanc. 2010. Environmental-endocrine control of reproductive maturation in gastropods: implications for the mechanism of tributyltin-induced imposex in prosobranchs. Ecotoxicology 19:4-23.

Sternberg, R. M., A. K. Hotchkiss, and G. A. Leblanc. 2008. Synchronized expression of retinoid X receptor mRNA with reproductive tract recrudescence in an imposex-susceptible mollusc. Environmental Science and Technology 42:1345-1351.

Sugarbaker, D. J., W. G. Richards, G. J. Gordon, L. Dong, A. De Rienzo, G. Maulik, J. N. Glickman, L. R. Chirieac, M. L. Hartman, B. E. Taillon, L. Du, P. Bouffard, S. F. Kingsmore, N. A. Miller, A. D. Farmer, R. V. Jensen, S. R. Gullans, and R. Bueno. 2008. Transcriptome sequencing of malignant pleural mesothelioma tumors. Proceedings of the National Academy of Sciences of the United States of America 105:3521-3526.

Sumner, S., J. J. M. Pereboom, and W. C. Jordan. 2006. Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, Polistes canadensis. Proceedings of the Royal Society B-Biological Sciences 273:19-26.

Svanback, R., M. Pineda-Krch, and M. Doebeli. 2009. Fluctuating Population Dynamics Promotes the Evolution of Phenotypic Plasticity. American Naturalist 174:176-189.

Swedenborg, E., J. Ruegg, S. Makela, and I. Pongratz. 2009. Endocrine disruptive chemicals: mechanisms of action and involvement in metabolic disorders. Journal of Molecular Endocrinology 43:1-10.

Tabb, M. M., and B. Blumberg. 2006. New modes of action for endocrine-disrupting chemicals. Molecular Endocrinology 20:475-482.

Thain, J., A. Vethaak, and K. Hylland. 2008. Contaminants in marine ecosystems: developing an integrated indicator framework using biological-effect techniques. International Council for the Exploration of the Sea (ICES) 65: 1508-1514.

Tillmann, M., U. Schulte-Oehlmann, M. Duft, B. Markert, and J. Oehlmann. 2001. Effects of endocrine disruptors on prosobranch snails (Mollusca: Gastropoda) in the laboratory. Part III: Cyproterone acetate and vinclozolin as antiandrogens. Ecotoxicology 10:373-388.

Toth, A. L., K. Varala, T. C. Newman, F. E. Miguez, S. K. Hutchison, D. A. Willoughby, J. F. Simons, M. Egholm, J. H. Hunt, M. E. Hudson, and G. E. Robinson. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. Science 318:441-444.

Travis, J. 1994. Evaluating the adaptive role of morphological plasticity. Ecological morphology: Integrative organismal biology:99-122.

Trussell, G. C. 1997. Phenotypic plasticity in the foot size of an intertidal snail. Ecology 78:1033-1048.

Trussell, G. C., and R. J. Etter. 2001. Integrating genetic and environmental forces that shape the evolution of geographic variation in a marine snail. Genetica 112:321-337.

Ungerer, M. C., L. C. Johnson, and M. A. Herman. 2008. Ecological genomics: understanding gene and genome function in the natural environment. Heredity 100:178-183.

van Straalen, N. M., and D. Roelofs. 2006. An introduction to ecological genomics. An introduction to ecological genomics. 1-307.

Vera, J. C., C. W. Wheat, H. W. Fescemyer, M. J. Frilander, D. L. Crawford, I. Hanski, and J. H. Marden. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636-1647.

Via, S., R. Gomulkiewicz, G. Dejong, S. M. Scheiner, C. D. Schlichting, and P. H. Vantienderen. 1995. Adaptive phenotypic plasticity - consensus and controversy. Trends in Ecology & Evolution 10:212-217.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57-63.

Wheat, C. W. 2008. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138:433-451.

Wisenden, B. D. 2000. Olfactory assessment of predation risk in the aquatic environment. Royal Society Philosophical Transactions Biological Sciences 355:1205-1208.

CHAPTER 2

De novo sequencing, assembly and annotation of a Nucella lapillus transcriptome: an EST resource for the investigation of the imposex mechanism

Chapter submitted as part of an original article:

Pascoal S, Carvalho G, Vasieva O, Hughes R, Cossins A, Fang Y, Ashelford K, Olohan L, Barroso C, Mendo S, Creer S. A Transcriptomic Perspective of Imposex Endocrine Disruption in an Ecological Sentinel (Nucella lapillus) Suggests Involvement of RXR:PPARγ Signalling. PloS Biology

2.1. ABSTRACT

Second-generation sequencing technology has enabled substantial advances in the genomic analysis of non-model organisms. Herein, 454 Roche GS-FLX Titanium pyrosequencing was employed to create a transcriptomic resource to facilitate the investigation of one of the best ecological examples of anthropogenically-induced endocrine disruption (imposex) in aquatic ecosystems. The dogwhelk, Nucella lapillus, is a recognized sentinel for tributyltin (TBT)-induced imposex and although the phenomenon is well documented ecologically, less is known at the genomic and biochemical level. Therefore a partial N. lapillus transcriptome was sequenced, assembled and annotated to create a contemporary genomic resource that can be used to elucidate the transcriptional mechanisms underpinning a poorly understood, but widespread gene-environment interaction. Pyrosequencing generated 899,283 expressed sequence tags (mean length = 242 nucleotides), that were reduced to 866,308 after size and adaptor quality trimming. Cleaned data were co-assembled into 29,645 contigs with 18X coverage and an average length of 447 bp and 141,994 singletons. Approximately 97% of the sequences showed no significant similarity with the NCBI nucleotide database, highlighting the lack of gene annotation in gastropods. However, combining NCBI NT and NCBI NR protein similarity results, 11,393 unique gene elements were identified (e-value≤10^{-6}) representing 7,412 independent proteins, and over 3,200 sequences were assigned with a functional gene ontology annotation. Additionally, approximately 2,000 potential microsatellite markers were identified within the EST sequences. This study generated a large EST resource for N. lapillus enabling reconstruction and annotation of a Nucella partial transcriptome. The transcriptomic data and functional genomic tools that can be derived from this study will provide a valuable resource to unravel the imposex mechanism and explore additional responses of dogwhelks to environmental variation.

2.2. INTRODUCTION

Since the 1960s, organotin pollutants such as tributyltin (TBT) have been deliberately introduced by man into the environment - mainly through the use of TBT as biocide in vessel's antifouling paints - and have been recognized to induce endocrine disruption in a diverse range of aquatic animals (Matthiessen and Gibbs 1998; Alzieu 2000; Barroso et al. 2000; Murata et al. 2008). In marine gastropods, TBT induces reproductive abnormalities that can lead to population decline and/or mass extinction (Gibbs and Bryan 1986; Bryan et al. 1990; Gibbs 1996). The phenomenon is referred to as imposex and it is characterized by the development of additional male sex organs (penis and/or vas deferens) in females (Gibbs et al. 1987; Morcillo and Porte 1999). The dogwhelk, Nucella lapillus, is not only a target species for ecological and evolutionary studies (Crothers 1985; Kirby 2000; Trussell and Etter 2001; Colson and Hughes 2007), but it is widely recognised as a sentinel organism for TBT pollution and is widely used for biomonitoring via the imposex response (Gibbs et al. 1987; Morcillo et al. 1999). The imposex phenomenon provides among the best ecological examples of anthropogenically-induced endocrine disruption in aquatic ecosystems, although the underlying molecular and biochemical mechanisms remain unclear (Horiguchi 2006; Castro et al. 2007).

Molluscs represent one of the largest and most diverse animal phyla, comprising marine, freshwater and terrestrial taxa, many of which are of broad phylogenetical, ecological and evolutionary interest, yet genomic resources remain very limited. Until now, mainly due to their importance in aquaculture and fisheries, most genomic data in molluscs have been focused on bivalves (Saavedra and Bachere 2006; Tanguy et al. 2008). For gastropods, transcriptomic studies or expressed sequence tags (ESTs) projects are largely restricted to freshwater snails Lymnaea stagnalis (Davison and Blaxter 2005; Feng et al. 2009) and Biomphalaria glabrata (Mitta et al. 2005; Lockyer et al. 2007, 2008) and the sea hare Aplysia (Moroz et al. 2006; Lee et al. 2008). Recently, the sequencing phases for the first gastropod whole-genome-sequencing projects have been completed for the limpet Lottia gigantea (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) and

Biomphalaria glabrata (http://www.ncbi.nlm.nih.gov/sra?term=Biomphalaria%20glabratta). Although the genomes are unpublished, these publicly accessible data facilitate comparative studies.

Following advances in second-generation sequencing, several genomic analyses such as whole-genome-sequencing (e.g. Wicker et al. 2006; Korbel et al. 2007; Quinn et al. 2008; Libants et al. 2009), metagenomics (e.g. Woyke et al. 2009; Creer et al. 2010), molecular marker development (e.g. Barbazuk et al. 2007; Satkoski et al. 2008; Tangphatsornruang et al. 2009) and transcriptome characterization (e.g. Cheung et al. 2006; Hahn et al. 2009; Meyer et al. 2009) are now possible on a scale that was hitherto impractical for most laboratories performing functional genomic research. Moreover, with its high accuracy (of base substitutions), low cost, and long reads, 454 Roche sequencing (Margulies et al. 2005) has become the sequencing platform of choice for the de novo analysis of non-genome-enabled organisms (Vera et al. 2008; Parchman et al. 2010), and is now associated with hundreds of peer-reviewed studies in diverse research fields such as cancer and disease research (Bainbridge et al. 2006; Sugarbaker et al. 2008), ecology and evolution (Wheeler et al. 2009), marine biology (Kristiansson et al. 2009), botany and agricultural biotechnology (Novaes et al. 2008; Wicker et al. 2009). Of particular relevance to the current study is the ability to sequence and assemble de novo non-model organismal transcriptomes using massively parallel pyrosequencing (Vera et al. 2008; Meyer et al. 2009; Parchman et al. 2010).

Acknowledging the increasing need for genomic resources of ecologically important non-model organisms, this large-scale EST project was performed to provide an ecotoxicology relevant transcriptomic resource for N. lapillus. The 454 Roche GS-FLX Titanium platform was used to sequence two normalised N. lapillus cDNA libraries; a control library and a TBT exposed library resulting from a three month laboratory exposure to this contaminant. The obtained EST collection was co-assembled to reconstruct and annotate the N. lapillus transcriptome. Additionally, comparative analyses with the Lottia gigantea whole genome sequencing project were performed and potential microsatellite markers were identified by in silico bioinformatic screening.

Transcriptomic data and the subsequent genomic tools will be valuable resources for exploring the genomic basis of the imposex mechanism and potentially other environmental and evolutionary processes in N. lapillus.

2.3. MATERIAL AND METHODS

2.3.1. Sampling, laboratory TBT exposure and RNA extractions

In order to capture an organismal-level and temporally representative transcriptomic response to TBT contamination, N. lapillus was exposed to TBTCl for 3 months in the laboratory. Based on a previous study of imposex levels in North Wales populations (Oliveira et al. 2009), two reference sites showing the lowest levels of TBT pollution (Cable Bay and Moelfre - Anglesey, UK) were selected for the collection of experimental dogwhelks. Accordingly, adults of N. lapillus were collected by hand from the rocky shores in January 2008. Animals were briefly narcotized in a 7% $MgCl_2$ solution to allow identification of gender and imposex levels under a stereo microscope. A total of 150 female dogwhelks (presenting a sperm ingesting gland (Gibbs et al. 1987)) from each population were analysed for physical evidence of imposex; the penis length was measured to the nearest 0.01 mm with a stereo microscope and eyepiece graticule, Vas Deferens Sequence Index (VDSI) and percentage of affected females (%I) were determined following Gibbs et al. (1987). Additionally, the whole body tissues from 10 females from each site were freeze-dried and analysed by Gas Chromatography (Harino et al. 2002; Harino et al. 2005) to access the organotin content in the tissues (Marine Biological Association, Plymouth). After imposex analysis, females were placed into several 1L glass flasks with seawater from the Menai Strait and maintained at 13 $^{o}C \pm 1$ in an acclimatized room under a 12 h light: 12 h dark photoperiod with continuous aeration. Animals were allowed to acclimatize for one week. After this period, females were exposed to two different TBTCl concentrations (100 and 200 ng Sn/L) for up to three months. Two control situations - just seawater and seawater with absolute ethanol (TBT

solvent) - were kept under the same conditions. Water was replaced weekly to maintain constant test concentrations. Four replicate flasks per population and per treatment were performed, each with 6 animals. The experiment was conducted with no food supply.

The following target tissues from control (with and without ethanol) and exposed animals were sampled after 1, 4, 7 and 13 weeks and further preserved in RNA later (Qiagen): digestive gland, gonads, gill, penis-forming area and cerebral ganglia. Tissues from the replicate animals (3 for the controls and 2 for the exposed, N=32) per treatment, time and population were pooled for downstream analysis. Total RNA was extracted with the RNeasy Midi RNA extraction kit (Qiagen) following the manufacturer´s instructions for animal tissue, followed by DNase treatment. Prior to RNA extraction all the sampled animals (N=80) were examined, using a stereo microscope to assess the level of imposex.

2.3.2. cDNA library construction and screening

To evaluate the quality and integrity of extracted RNA and cDNA, two cDNA libraries (control library: pool of 16 control RNA extractions and exposed library: pool of 16 (eight 100 and eight 200 ng Sn/L) exposed RNA extractions) were constructed following the Creator SMART cDNA library construction Kit (Clontech) following the manufacturer's protocol. One aliquot of each library was cloned and used for PCR library screening, using the provided M13 primers. Clones (N=45) from each library were sequenced by Macrogen using Sanger sequencing, and further analysed combining Trace2dEST (Parkinson et al. 2004) and manual NCBI BLAST searches.

2.3.3. cDNA normalization

Normalization is predicted to enhance the gene discovery rate within the cDNA libraries and facilitate the identification and analysis of rare transcripts, thus allowing a better representation of the transcribed genes (Zhulidov et al. 2004). Thus, the N. lapillus libraries were normalized prior to sequencing. Two total RNA pools (2 µg each): control and exposed (16 RNA extractions each, individually extracted and combined equally) were

sent to Evrogen (Russia Federation) where double stranded cDNA suitable for non-directional cloning was synthesized using the SMART approach (Zhu et al. 2001). The SMART amplified cDNAs were then normalized (2 µg) using a duplex-specific nuclease (DSN) normalization method (Zhulidov et al. 2004). Briefly, this method included cDNA denaturation/reassociation followed by the duplex-specific nuclease treatment (Shagin et al. 2002) and amplification of the normalized fraction by PCR.

2.3.4. 454 sequencing and data analysis

Approximately 2 µg of each normalized cDNA library was used for a half-plate run on the 454 Roche GS-FLX Titanium series sequencer at the Centre for Genomic Research, Liverpool University, UK. Prior to sequencing, TBT and control cDNAs were differentially tagged in order to identify reads derived from the control and experimental treatments. However, given that the cDNA libraries were normalised, differential patterns of gene expression were not analysed here. The TBT exposed library was tagged with MID6-A adapter, primer GSMID6 A1: 5'C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGATATCG*C*G*A*G3' and Primer GSMID6 Aprime1: 5'C*T*C*G*CGATAT CTGAGTCG*G*A*G*A3' and the Control library was tagged with MID7-A adapter, primer GSMID7 A1: 5'C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGCGTGTC*T*C*T*A3' and Primer GSMID7 Aprime1: 5'T*A*G*A*GACACGCTGAGTCG*G*A*G*A3'. Both libraries had the B adapter; Primer B1: 5'/5BioTEG/C*C*T*A*TCCCCTGTGTGCCTTGGCAGTC*T*C*A*G3' and Primer Bprime1: 5'C*T*G*A*GACT*G*C*C*A3'.

Initial quality filtering of the 454 Roche sequences was performed at the machine level and then in-house Perl scripts (Kevin Ashelford, unpublished data; please see appendix 2.1) were used to trim normalization and sequencing adaptors and to size select the ESTs. All sequences with less than 50 bp were discarded. Cleaned data were then submitted to the 454's Roche Newbler assembly software version 2.0.01.12 (Roche) using default settings; firstly assembling the two independent normalized libraries and

finally co-assembling both libraries data. Pyrosequencing reads derived from 454 Roche sequencers may contain artificial duplicates, which might lead to misleading conclusions (Niu et al. 2010). Therefore, in order to evaluate and discard redundant (or highly similar) sequences, all singletons were firstly clustered using CDHIT-EST software (Li and Godzik 2006), with a final threshold identity of 95% and a word size of 8. All contigs and non-redundant singletons were annotated by BLAST search against the NCBI non-redundant protein (NR) and nucleotide (NT) databases where the e-value threshold was set at e-value≤10^{-6}, and top-query sequence was selected by an in-house Perl script based on a higher percentage homology assignment versus alignment length. Gene ontology (GO) annotations were also obtained based on sequence similarity using the BLAST2GO online software using a threshold of e-value≤10^{-6} (Conesa et al. 2005; Gotz et al. 2008) (http://www.blast2go.de/). Given the lack of genomic annotation in the gastropods and the lack of a Nucella genome sequence, BLASTx and tBLASTx (significance threshold of e-value≤10^{-6}) searches of the dogwhelk sequences against the predicted peptides for the limpet (Lottia gigantea) genome (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) were performed, since it is the most closely related organism with a fully sequenced genome. These comparisons facilitated the quantification of the transcriptome similarity between the two gastropod species and the estimation of levels of confidence in the dogwhelk data.

In order to check sequence accuracy and reproducibility between chain termination sequencing and pyrosequencing, the Sanger sequences obtained in the quality control (QC) libraries were searched within the 454 Roche data. For this, the Sanger sequences were added to the NCBI NT database and then BLASTn (significance threshold of e-value≤10^{-6}) searches of the 454 Roche data to the augmented database (Way Sung, personal communication) were performed.

Finally, given the abundance of sequences with repeated motifs (microsatellites) within the ESTs and regarding the wide usage of N. lapillus in various ecological and evolutionary studies (e.g. Colson and Hughes 2004; Guerra-Varela et al. 2009), microsatellite sequences were searched for within the 454 Roche obtained sequences

using the program MSATCOMMANDER (Faircloth 2008), applying a repeat threshold of seven to dinucleotides and five to tri, tetra, penta or hexanucleotides.

2.4. RESULTS AND DISCUSSION

2.4.1. Imposex background levels, organotin content in tissues and TBT laboratory exposures

The imposex background levels from the two target populations (Cable Bay, Moelfre) and from the population (Menai Bridge) representing the water source used in the laboratory experiments together with the organotin contents in the N. lapillus tissues are presented in Table 1. Females with imposex were observed in all populations; however, the levels of imposex and organotin tissue contamination were very low comparing to reported values in the past for the same sampling area, other regions of the U.K. and in other European countries (Oliveira et al. 2009). Using dogwhelks collected in 2006, Oliveira et al. (2009) have reported low level of imposex (Cable Bay: VDSI=0.9; Moelfre: VDSI=0.8; Menai Bridge: VDSI=1.6) that accounted for a VDSI decrease of up to 83% since 1987. The present data (Table 1), confirm a decreasing tendency of TBT pollution levels and a population recovery in the sampling area leading to legislation accuracy.

Table 1. Nucella lapillus imposex basic levels and organotin tissue concentration. SL: mean shell length, FPL: Female Penis Length (average (standard deviation)); VDSI: Vas Deferens Sequence Index (average (standard deviation)); (%I): percentage of females affected by imposex; and for the whole female tissues, concentrations of MBT: Monobutyltin, DBT: Dibutyltin, TBT: Tributyltin; TPT: Triphenyltin (ng Sn/g dry wt).

	Imposex				**Organotin concentrations**			
Population	SL (mm)	FPL (mm)	VDSI	%I	MBT	DBT	TBT	TPT
Cable bay	24.21	0.00 (0.00)	0.36 (0.48)	31.03	1.2	1.5	1.1	5.3
Moelfre	30.00	0.05 (0.18)	0.61 (0.63)	53.70	5.9	1.0	2.4	4.1
Menai Bridge	32.53	0.22 (0.46)	1.00 (0.71)	80.88	3.2	0.8	2.5	6.4

The three month laboratory exposure to TBT induced imposex in N. lapillus (Fig.1), confirming that the ecotoxicological stimulus had instigated a comprehensive endocrine disruption, accompanied by suites of differentially expressed genes.

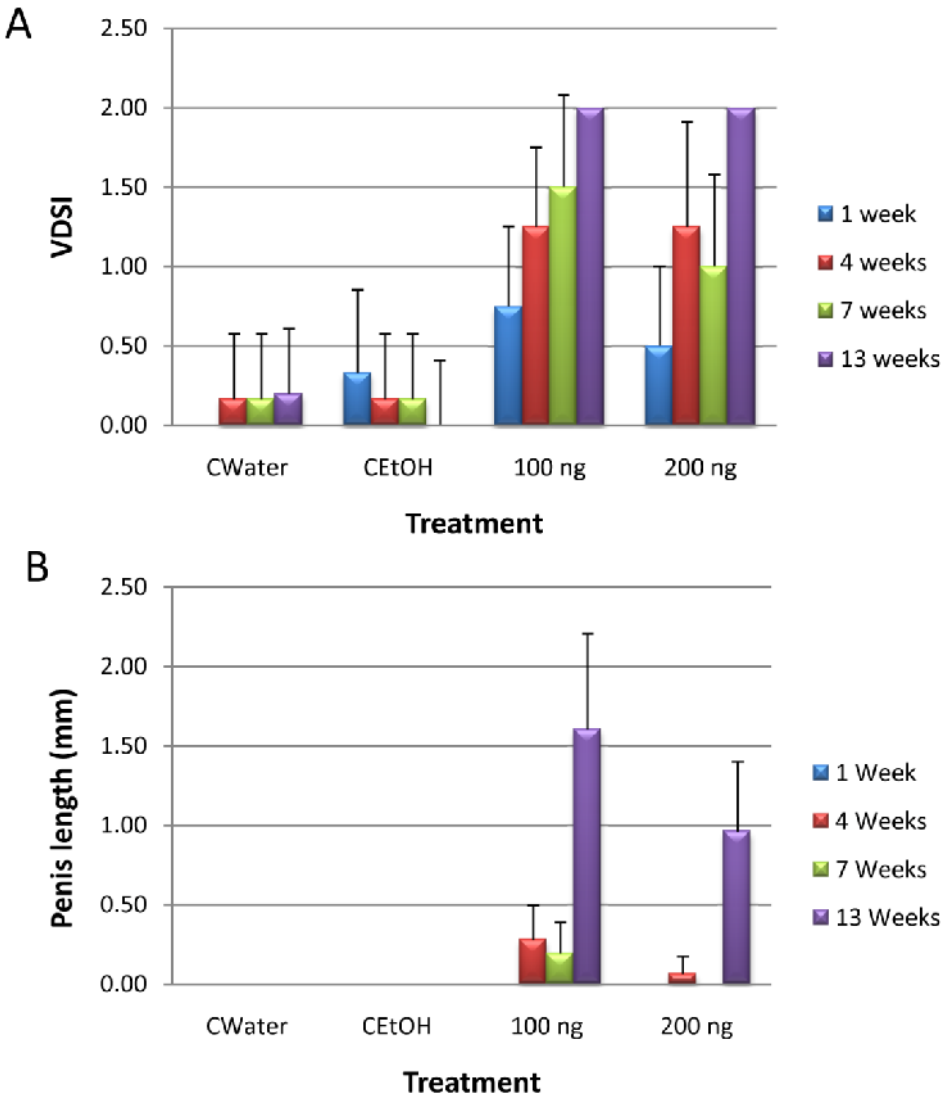


Figure 1. Nucella lapillus imposex resulting from laboratory exposure to TBT. VDSI: Vas Deferens Sequence Index; Cwater: control just water, CEtOH: control with ethanol; 100 ng: 100 ng Sn/L, 200 ng: 200 ng Sn/L. Data correspond to mean values (N=80) and respective standard errors.

2.4.2. Sequencing and data assembly

The control library sequencing yielded 275,523 sequences with an average length of 225 bp and maximum length of 605 bp; the TBT-exposed library produced 623,760 sequences with an average length of 260 bp and a maximum length of 636 bp (Table 2, Fig. 2A). Together, the pyrosequencing generated a total of 223,643,710 bases from 899,283 expressed sequence tags (ESTs) with an average length of 242 bases (Table 2). The sequencing reads produced in this study have been deposited in NCBI's Short Read Archive database (accession number: SRA021021).

Table 2. Summary of Nucella lapillus sequencing and assembly results.

	Sequences (n)	Bases (Mb)
Control library raw reads	275,523	62.0
TBT library raw reads	623,760	161.6
All raw sequencing reads	899,283	223.6
Trimmed & size-selected	866,308	203.4
Co-assembled contigs	29,645	13.3
Average length (bp)	447.08	–
Range length (bp)	92 to 3,296	–
Singletons after co-assembly	141,994	31.6
Average length (bp)	222.33	–
Range length (bp)	50 to 636	–
Total assembled sequences	171,639	44.9

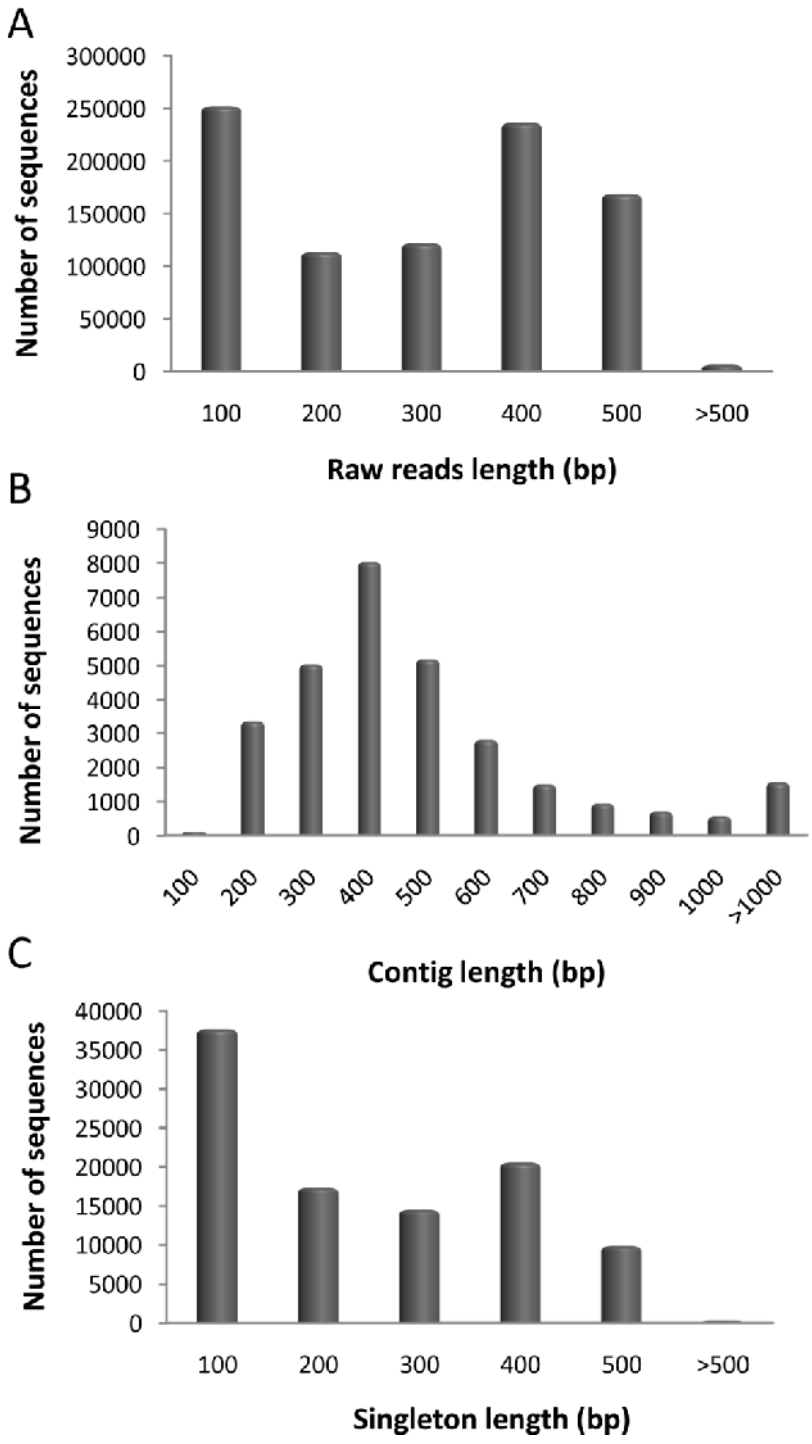


Figure 2. Summary statistics for N. lapillus transcriptome sequencing and assembly. A) Raw sequence read length distribution; B) Contig sequence length distribution; C) Singleton length distribution.

Size (<50 bp) and adaptor quality trimming discarded 3.6 % of the original reads, reducing the number of high quality ESTs to 866,308. The assembly was initially performed separately for the two cDNA libraries, followed by assembly of the combined data (Fig. 3). Co-assembly revealed a higher number of longer contigs and therefore all further analyses were performed based on the co-assembled dataset.

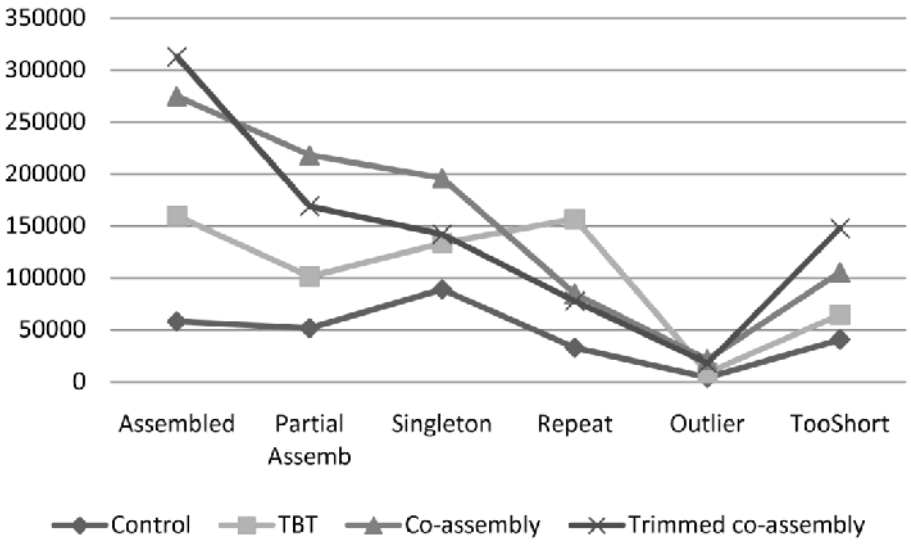


Figure 3. Newbler assembly optimization. Number of assembled and partial assembled sequences, singletons, outliers and too short sequences for each one of the assembly simulations; control: just control library sequencing data; TBT: just exposed to TBT library sequencing data; Co-assembly: assembly performed with both libraries data together; Trimmed co-assembly: assembly performed with both libraries data together after adaptors cleaning.

The quality controlled data was co-assembled into 171,639 sequences; 29,645 contigs (i.e. a set of overlapping DNA segments derived from a single genetic source) with an average length of 447 bp (Table 2, Fig. 2B) and 141,994 singletons (i.e. a single read that does not contain enough overlap in length to be combined with other reads from the same transcribed gene) with an average length of 222 bp (Table 2, Fig. 2C). The data revealed 7,965 large contigs with an average length of 792 bp. As expected for a randomly fragmented transcriptome (Meyer et al. 2009), the length of the contigs generally increased with the number of sequences assembled into them (Fig. 4). Consequently, a mean contig coverage of 18X (mean=18.16; SD=35.59) was obtained.

Following clustering, 31,926 putatively redundant singletons were identified as being putatively homologous to non-redundant sequences and so were removed from downstream analyses. Consequently, all further analyses were performed using a total of 129,713 assembled contigs and non-redundant singleton sequences.

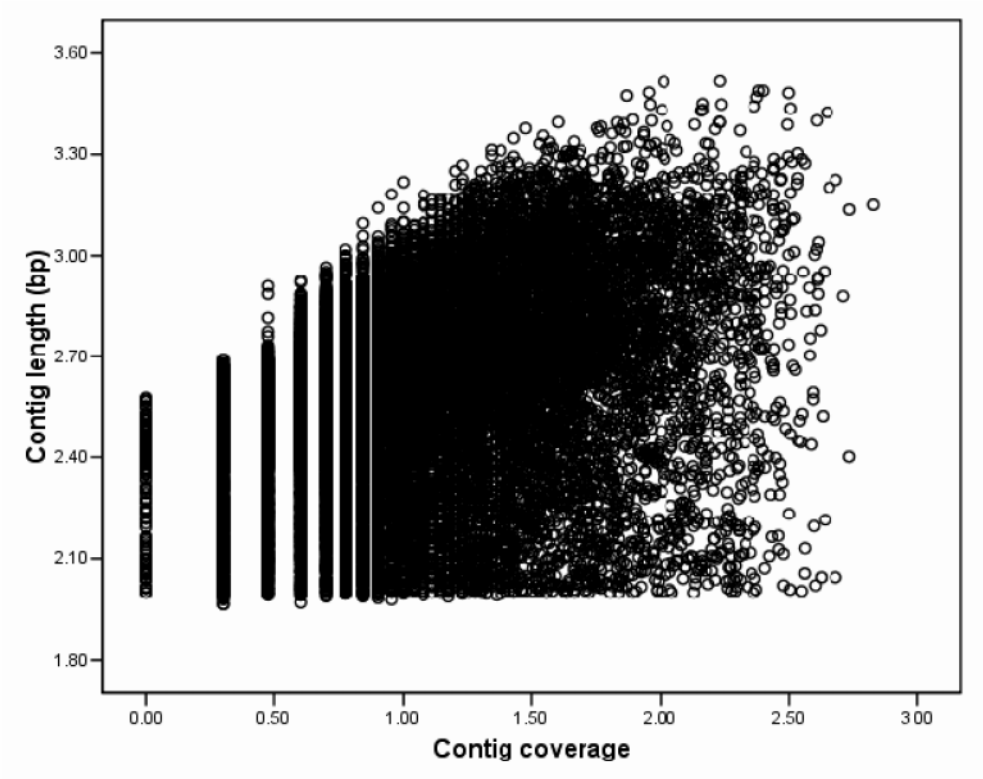


Figure 4. Contig coverage: log-log plot showing contig length as a function of the number of sequences assembled into each contig.

2.4.3. BLAST annotation

All sequences were assigned to gene names based on the gene product and gene name annotation of the best BLAST match for that sequence. Approximately 95% of the contigs and 98% of the singletons showed no significant similarity (e-value≤10^{-6}) with the NCBI nucleotide database, highlighting the lack of gene annotation in gastropods. Analysis of the taxa represented in the annotated sequences (Fig. 5A) shows that gastropods were the most represented (49% of the annotations) featuring approximately 30 different species, but dominated by hits to Nucella. Within the gastropod annotated subset of

genes, all previously described Nucella hits (except microsatellites) were represented (12S, cytochrome b, cytochrome oxidase subunit I (COI), malate dehydrogenase precursor (Mdh), actin, estrogen receptor, retinoid x receptor (RXR) and 28S), and a significant proportion are potentially related to the imposex mechanism (Fig. 5B). However, the dominance of genes encoding estrogen receptors and microsatellites sequences in this figure may be also related to the over-representation of these sequences among the existing gastropod annotations. Along with gastropods, mainly model invertebrates species, were well represented showing a closer genomic similarity between Nucella and taxonomically related species. Additionally, all Sanger sequences obtained from the initial cDNA libraries were found among the 454 Roche sequences with an average percentage of homology of 97% and an average alignment length of 298 bp, confirming reproducibility between protocols and assembly accuracy. Overall, the results provided confidence in the integrity of data, and additionally highlight the representation of genes that may be involved in the imposex response.

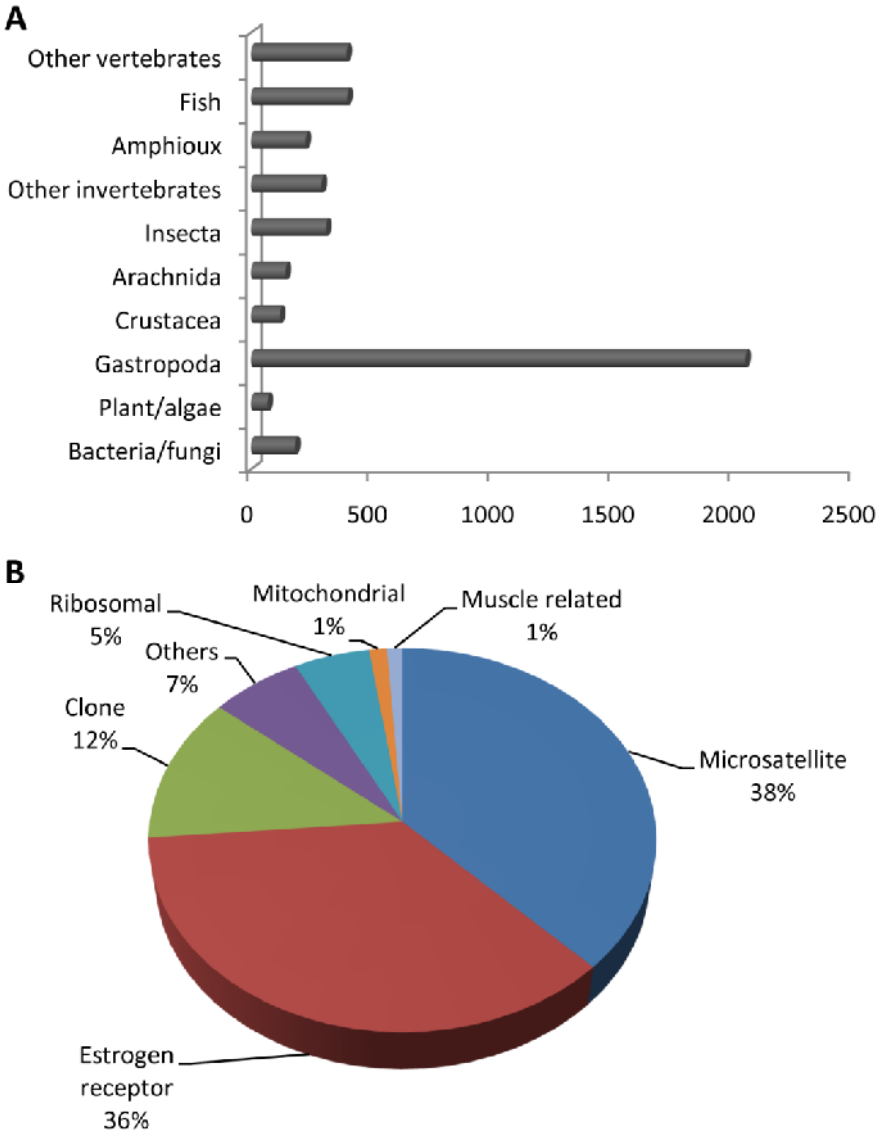


Figure 5. BLAST annotation. Contigs and singletons BLASTn results. A) The number of top scoring BLASTn (e-value≤10^{-6}) homology matches, separated into taxon frequency; B) The percentage of gene annotations obtained via BLASTn (e-value≤10^{-6}) within Gastropoda.

A BLASTx search of all contigs and singletons against the NCBI-NR protein database revealed 4,191 contig and 8,578 singletons as well-identified sequences with at least a single hit e-value≤10^{-6}. Additionally, in order to optimize BLAST searches for gene finding (strategy adopted to serve as a "capture all" search, facilitating the downstream scrutiny of putative annotations via reference to biology, physiology and homology) the e-value was reduced to e-value=10^{2} and subsequently 13,143 contigs and 42,801 singletons

sequences were shown to share homologies with GenBank accession numbers (between e-value≤10^{-6} and e-value=10^{1}), and 9,625 contigs and 36,459 singleton sequences showed no useful hits (e-value>10^{1}). By exploring higher and lower stringency BLAST search parameters 24,588 contigs (83% of the contigs) and 52,805 singletons (53% of the singletons) (Table 3) were annotated.

Table 3. Summary of Nucella lapillus BLAST results against the NCBI NT and NR databases. Annotation: number of different N. lapillus sequences with annotation; unique gi: Number of different annotations in the blasted sequences; All: Contigs and singletons after cluster results.

			BLASTn		BLASTx	
Contigs 29,645	e-6	annotation	1,519	5.12%	4,191	14.14%
		unique gi	1,006		3,173	
	e=10^{2}	annotation	---	---	24,588	82.94%
		unique gi			21,916	
Singletons 100,068	e-6	annotation	2,248	2.25%	4,835	4.83%
		unique gi	947		2,762	
	e=10^{2}	annotation	---	---	52,805	52.77
		unique gi			43,385	
All 129,713	e-6	annotation	11,393			
		unique gi	7,412			
	ne-6&xe=10^{2}	annotation	77,393			
		unique gi	68,094			

As documented before (Meyer et al. 2009), longer assembled sequences showed a higher percentage annotation than shorter sequences (Fig. 6).

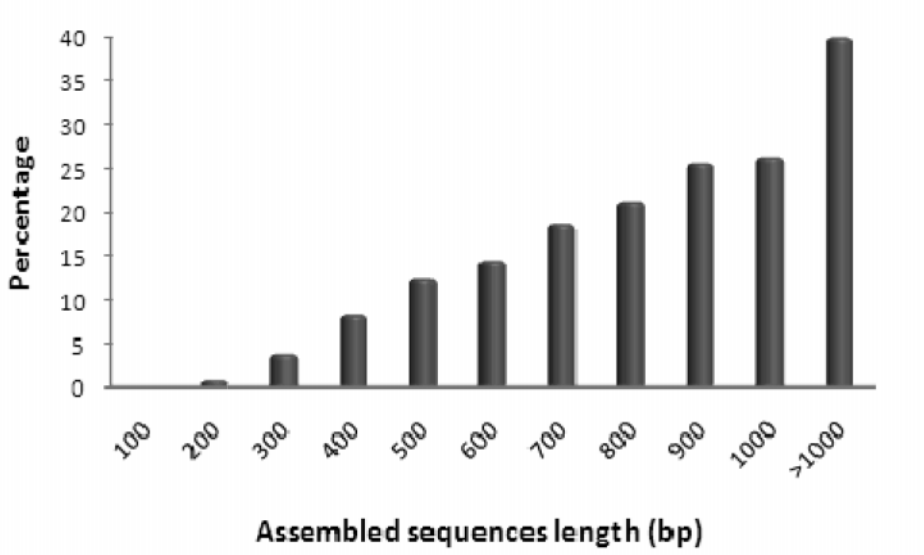


Figure 6. Percentage of annotated sequences with length x.

Combining all the well identified sequences (e-value≤10^{-6}) resulting from BLASTn and BLASTx analysis (17,934), and discarding the occasional duplicate hit per sequence, 11,393 unique well annotated Nucella gene elements representing 7,412 different gene annotations (accession numbers), (Table 3) were obtained. Although the well-identified transcripts represented just a small proportion of the entire sequenced transcriptome, they make a substantial contribution towards the interpretation of Nucella genomic data, which was almost inexistent prior to the current study. Moreover, despite the low number of matches (about 14% of the N. lapillus contigs), the percentage of N. lapillus annotations is in the same range of other non-model marine molluscs, e.g. the blue mussel Mytilus galloprovinciallis (12%), (Craft et al. 2010), and the Antarctic bivalve Laternula elliptica (17%), (Clark et al. 2010).

2.4.4. Functional classification based on Gene Ontology annotation

Using BLAST2GO, a total of 6,962 BLAST-annotated sequences were mapped into 16,994 GO annotation terms (primarily derived from the UniprotKB main database) representing 3,226 assembled sequences with GO. The sequences with an attributed

function were similarly distributed in the main GO categories, molecular function (2,907 sequences), biological process (2,234 sequences) and fewer in the cellular component category (1,986 sequences), (Fig. 7). Sequences mapping to different GO categories were equally represented within the categories. The specific annotated terms were mapped to the more general parent terms (GO level 2) to provide a more general overview of the represented functional group of genes present in the Nucella transcriptome. The most prevalent cellular component assignments were for genes encoding cellular proteins (44%) and for genes encoding organelle proteins (30%), (Fig. 7A). In the molecular function categories, the largest proportion was assigned to binding (47%) and catalytic activities (35%), (Fig. 7B). Biological processes were assigned into 14 main categories with 29% of the assignments representing cellular processes, 23% metabolic processes and 11% biological regulation with the remaining assignments distributed into less well represented group functions (Fig. 7C). In the latter group, reproduction and response to stimulus, with 133 and 303 homologous sequences respectively, are functional groups of great potential relevance to the present study. These two sub-categories were therefore investigated further by exploring the hierarchical structure of the ontology vocabularies that permits the selection of sets of genes involved in a specific process at a desired level of detail. The reproduction functional group was represented by genes involved in the reproductive process, gamete and gonad formation and sex differentiation (Table 4). The group corresponding to the biological process, response to stimulus, was mainly represented by genes involved in the response to external and internal stimulus and either to biotic or abiotic stimuli, as well as detection of that stimulus (Table 4). Overall, the main GOs category distribution of the Lymnaea (Feng et al. 2009) and Biomphalaria (Lockyer et al. 2007) gastropod transcriptomes were similar to those of Nucella.

The GO annotations (e-value$\leq 10^{-6}$) provided a valuable resource for the investigation of specific processes, functions or cellular structures involved in the imposex response of Nucella following exposure to levels of organotin that induce endocrine disruption. Additionally, the broad distribution of Nucella sequences into different functional categories is likely to reflect a comprehensive representation of the dogwhelk transcriptome.

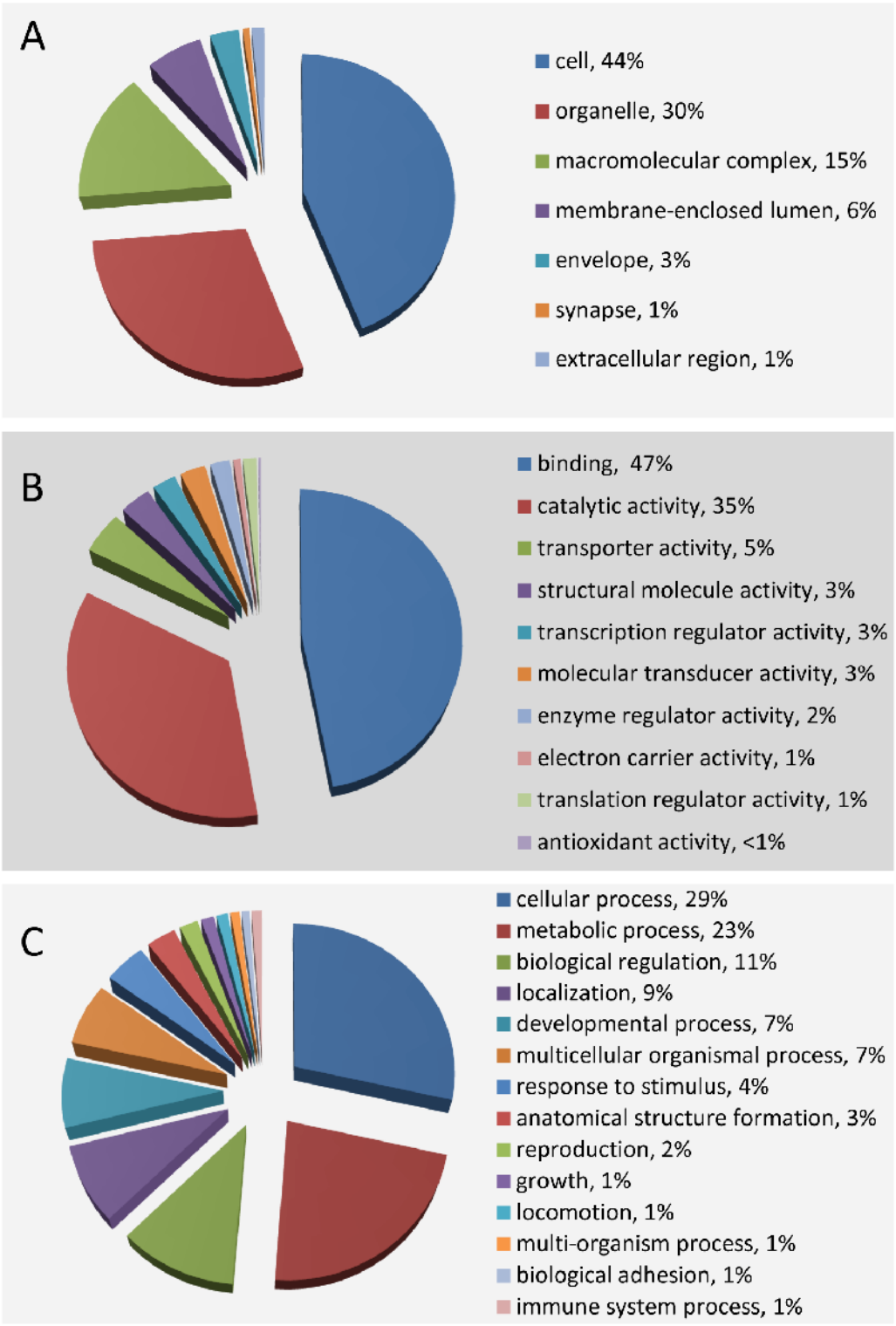


Figure 7. Gene ontology. Percentage of gene ontology annotations for Nucella lapillus sequences; A) Cellular component, B) Molecular function, and C) Biological process.

Table 4. Hierarchical details (until level 5) for the biological process sub-categories response to stimulus and reproduction.

Level	GO sub-category	# seq.	Parents
2	Response to stimulus	303	Biological process
3	Response to stress	170	Response to stimulus
	Response to chemical stimulus	128	Response to stimulus
	Cellular response to stimulus	70	Response to stimulus
	Response to external stimulus	60	Response to stimulus
	Response to abiotic stimulus	43	Response to stimulus
	Response to biotic stimulus	31	Response to stimulus
	Immune response	31	Response to stimulus
	Response to endogenous stimulus	29	Response to stimulus
	Detection of stimulus	11	Response to stimulus
4	Cellular response to stress	66	Response to stress, Cellular response to stimulus
	Response to DNA damage stimulus	64	Response to stress
	Response to organic substances	31	Response to chemical stimulus
	Response to inorganic substance	29	Response to chemical stimulus
	Response to wounding	27	Response to stress, Response to external stimulus
	Defense response	25	Response to stress
	Response to hormone stimulus	24	Response to endogenous stimulus, Response to chemical stimulus
	Response to drug	23	Response to chemical stimulus
	Response to other organism	22	Response to biotic stimulus
	Response to radiation	22	Response to abiotic stress
	Response to oxidative stress	19	Response to stress, Response to chemical stimulus
	Response to extracellular stimulus	18	Response to endogenous stimulus
	Regulation of response to stimulus	15	Response to stimulus
	Response to protein stimulus	14	Response to chemical stimulus
5	Cellular response to DNA damage stimulus	49	Cellular response to stress
	Response to metal ion	27	Response to inorganic substance
	Response to nutrient levels	18	Response to extracellular stimulus
	Response to light stimulus	16	Response to radiation
	Response to organic cyclic substances	12	Response to organic substances
	Wound healing	12	Response to wounding
	Response to bacterium	11	Response to other organism
2	Reproduction	133	Biological process
3	Reproductive process	112	Reproduction
	Sexual reproduction	68	Reproduction
	Multicellular organism reproduction	21	Reproduction
4	Gamete generation	63	Sexual reproduction, Reproductive process
	Reproductive developmental process	24	Reproductive process
	Reproductive process in a multicellular organism	21	Multicellular organism reproduction, Reproductive process
	Viral reproductive process	15	Reproductive process
5	Female gamete generation	36	Gamete generation
	Male gamete generation	29	Gamete generation
	Sex differentiation	24	Reproductive developmental process
	Viral infection cycle	14	Viral reproductive process
	Reproductive structure development	11	Reproductive developmental process

2.4.5. Comparison with the limpet (Lottia gigantia) genomic data

Initially, BLASTn analyses between the two datasets using a threshold of e-value≤10^{-6} were conducted but the levels of similarity between the two genomes were very low (0.6%). Since the low hit rate may reflect differences between the two genetic codes (Feng et al. 2009), tBLASTx analyses (i.e. searching a translated nucleotide database using a translated nucleotide query) were performed with an e-value≤10^{-6} threshold, yielding 8,231 (4.8%) sequences with a BLAST result (Table 5). A total of 4,040 different limpet genes, representing 17% of all limpet genes used for the comparisons, were well represented in the Nucella transcriptome. Finally, BLASTx analysis using an adjusted threshold of e-value≤10^{-6} revealed 7,350 (4.3%) sequences (Table 5) that had significant matches among the predicted limpet proteins. From the limpet data, 3,803 different predicted proteins were well represented in the N. lapillus transcriptome. In total, 9,152 sequences that represent 4,348 different transcripts with high similarity between Lottia and Nucella EST sequences (Table 4) were identified. Individual analysis of the contigs showed slightly higher levels of similarity (approximately 12%), which is within the normal range of homology assignments uncovered between other Gastropoda-Gastropoda comparisons (Feng et al. 2009). Overall, the comparison of the EST datasets revealed a low level of similarity between the transcriptomes of Lottia gigantia and N. lapillus, but nevertheless data collation revealed over 4,000 putitatively orthologous genes between the two species.

Table 5. Summary of BLAST analysis against the Lottia gigantia dataset. Num. sequences: number of different Nucella lapillus sequences with a positive blast result with the limpet database; Unique gene ids: number of different gene identifications represented in the blast results.

		BLASTn		tBLASTx		BLASTx		All
Contigs	Num. sequences	296	1.00%	3,565	12.00%	3,353	11.30%	3,780
	Unique gene ids	246		2,405		2,339		2,523
Singletons	Num. sequences	633	0.50%	4,666	3.30%	3,997	2.80%	5,372
	Unique gene ids	177		2,274		2,042		2,372
All	Num. sequences	929	0.54%	8,231	4.8%	7,350	4.3%	9,152
	Unique gene ids	390		4,040		3,803		4,348

The genbank non-redundant database is one of the best annotated sources for comparative in silico gene analysis. However, there are other larger but less annotated molluscan datasets such as the sequenced genome of L. gigantea (used herein) and more recently 454 data for Mytilus species (Craft et al. 2010), Strombus gigas (Spade et al. 2010), Littorina saxatilis (Galindo et al. 2010) and Laternulla elliptica (Clark et al. 2010) that are useful in terms of EST verification and gene mining. However, most datasets were released very recently and were therefore not used for comparative genomics in this study.

Searching the Nucella dataset against the NCBI and limpet databases yielded a large proportion of sequences that have no similarity to published data. This is likely to be due to a combination of the lack of annotation in closely related species, the incomplete gene sequence nature of EST itself and the presence of specific N. lapillus sequences within the present dataset (Vera et al. 2008). Using the same sequencing approach, similar results have been obtained for the above mollusc species (Clark et al. 2010; Craft et al. 2010; Galindo et al. 2010; Spade et al. 2010) and indeed, non-model species without a close genomic reference species in general (e.g. corals (Meyer et al. 2009) and butterflies (Vera et al. 2008)). Although a reference genome from a related species is not essential, when they are available, the analysis and interpretation of the data is further improved because these genomes provide valuable comparative resources for genome assembly, candidate gene discovery and subsequent analyses of sequence divergence rates and patterns (Toth et al. 2007; Stapley et al. 2010).

Although gene annotation and functional characterization of sequence variation in non-model organisms remains a challenge (Stapley et al. 2010), recent high-throughput sequencing projects such as the N. lapillus are making a significant contribution for the gastropod genomic resources growing field in particular and for non-model organisms biology understanding in general.

2.4.6. Microsatellite development

By screening all the contigs and singletons for microsatellite motifs, 15,162 sequences were recovered containing at least seven dinucleotide repeats, and at least five tri, tetra, penta or hexanucleotide repeats. However, only 12% had both 5' and 3' flanking regions that facilitated primer design (more than 50 bp) for genotyping, whereas some sequences permitted the development of more than one microsatellite marker. Accordingly, in silico 1,955 potential molecular markers (EST-derived microsatellites) for N. lapillus (Table 6) were identified. From this analysis, it was evident that the developed EST database contains an abundance of repetitive regions and hence provides a potentially valuable resource for molecular marker development (Vera et al. 2008; Hahn et al. 2009). Following optimization and polymorphism testing, these markers may enable exploration of diverse ecological and evolutionary questions concerning the intertidal environment (e.g. local adaptation and selective sweeps).

Table 6. Summary of microsatellite identification.

	Singletons	Contigs	Total
Sequences searched for repeats	100,068	29,645	129,713
Sequences containing repeats	9,847	2,569	12,416
Total number of repeats found	12,838	32,214	45,052
Dinucleotide	7,106	1,239	8,345
Trinucleotide	3,462	1,129	4,591
Tetranucleotide	1,743	568	2,311
Pentanucleotide	492	193	685
Hexanucleotide	35	11	46
Potential markers	1,390	565	1,955

2.5. CONCLUSION

Herein, a large EST resource for N. lapillus was generated that facilitated the reconstruction of a de novo Nucella partial transcriptome, with an estimated coverage of 18X. From the general NCBI database and more specific limpet BLAST searches and subsequent functional gene classification, the first group of gene annotations for the Nucella transcriptome was documented identifying a large group of genes of interest. The transcriptomic data and functional genomic tools that can be derived from this study (e.g. candidate genes development (Ellegren 2008), large scale expression analysis using oligoarrays (Vera et al. 2008), genomic scans (Wheat 2008)) will provide a valuable resource for elucidating the functional genomic basis of the imposex mechanism and exploring environmental and evolutionary responses of this sentinel organism to challenges of the intertidal environment. Moreover, a significant resource of potential microsatellite molecular markers for a gastropod species was produced. The produced data will also be a valuable resource for comparative genome analysis and significantly augment transcriptomic knowledge within the class Gastropoda that hitherto has been poorly studied at the genomic level.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) Ph.D. grant SFRH/BD/27711/2006 awarded to Sonia Pascoal and by Natural Environment Research Council (NERC) MGF grant 294 awarded to Simon Creer. I would like to thank Prof. Neil Hall and Dr Margaret Hughes for valuable advice during 454 sample preparation and 454 Roche sequencing and acknowledge the long-term support of NERC for supporting the Liverpool node of the NERC Biomolecular Analysis Facility.

2.6. REFERENCES

Alzieu, C. 2000. Impact of tributyltin on marine invertebrates. Ecotoxicology 9:71-76.

Bainbridge, M. N., R. L. Warren, M. Hirst, T. Romanuik, T. Zeng, A. Go, A. Delaney, M. Griffith, M. Hickenbotham, V. Magrini, E. R. Mardis, M. D. Sadar, A. S. Siddiqui, M. A. Marra, and S. J. M. Jones. 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. Bmc Genomics 7.

Barbazuk, W. B., S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable. 2007. SNP discovery via 454 transcriptome sequencing. Plant Journal 51:910-918.

Barroso, C. M., M. H. Moreira, and P. E. Gibbs. 2000. Comparison of imposer and intersex development in four prosobranch species for TBT monitoring of a southern European estuarine system (Ria de Aveiro, NW Portugal). Marine Ecology-Progress Series 201:221-232.

Clark, M. S., M. A. S. Thorne, F. A. Vieira, J. C. R. Cardoso, D. M. Power, and L. S. Peck. 2010. Insights into shell deposition in the Antarctic bivalve Laternula elliptica: gene discovery in the mantle transcriptome using 454 pyrosequencing. Bmc Genomics 11.

Craft, J. A., J. A. Gilbert, B. Temperton, K. E. Dempsey, K. Ashelford, B. Tiwari, T. H. Hutchinson, and J. K. Chipman. 2010. Pyrosequencing of Mytilus galloprovincialis cDNAs: Tissue-Specific Expression Patterns. Plos One 5.

Bryan, G. W., P. E. Gibbs, P. L. Pascoe, G. R. Burt, and L. G. Hummerstone. 1990. Decline of the dogwhelk Nucella lapillus in response to tributyltin TBT pollution. Journal of the Marine Biological Association of the United Kingdom 70:669-670.

Castro, L. F., D. Lima, A. Machado, C. Melo, Y. Hiromori, J. Nishikawa, T. Nakanishi, M. A. Reis-Henriques, and M. M. Santos. 2007. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod Nucella lapillus. Aquatic Toxicology 85:57-66.

Cheung, F., B. J. Haas, S. M. D. Goldberg, G. D. May, Y. L. Xiao, and C. D. Town. 2006. Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. Bmc Genomics 7.

Clark, M. S., M. A. S. Thorne, F. A. Vieira, J. C. R. Cardoso, D. M. Power, and L. S. Peck. 2010. Insights into shell deposition in the Antarctic bivalve Laternula elliptica: gene discovery in the mantle transcriptome using 454 pyrosequencing. Bmc Genomics 11.

Colson, I., and R. N. Hughes. 2004. Rapid recovery of genetic diversity of dogwhelk (Nucella lapillus L.) populations after local extinction and recolonization contradicts predictions from life-history characteristics. Molecular Ecology 13:2223-2233.

Colson, I., and R. N. Hughes. 2007. Contrasted patterns of genetic variation in the dogwhelk Nucella lapillus along two putative post-glacial expansion routes. Marine Ecology-Progress Series 343:183-191.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Craft, J. A., J. A. Gilbert, B. Temperton, K. E. Dempsey, K. Ashelford, B. Tiwari, T. H. Hutchinson, and J. K. Chipman. 2010. Pyrosequencing of Mytilus galloprovincialis cDNAs: Tissue-Specific Expression Patterns. Plos One 5.

Creer, S., V. G. Fonseca, D. L. Porazinska, R. M. Giblin-Davis, W. Sung, D. M. Power, M. Packer, G. R. Carvalho, M. L. Blaxter, P. J. D. Lambshead, and W. K. Thomas. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology 19:4-20.

Crothers, J. H. 1985. Dog-whelks an introduction to the biology of Nucella-lapillus. Field Studies 6:291-360.

Davison, A., and M. L. Blaxter. 2005. An expressed sequence tag survey of gene expression in the pond snail Lymnaea stagnalis, an intermediate vector of Fasciola hepatica (vol 130, pg 539, 2005). Parasitology 131:723-723.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17:1629-1631.

Faircloth, B. C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8:92-94.

Feng, Z. P., Z. Zhang, R. E. van Kesteren, V. A. Straub, P. van Nierop, K. Jin, N. Nejatbakhsh, J. I. Goldberg, G. E. Spencer, M. S. Yeoman, W. Wildering, J. R. Coorssen, R. P. Croll, L. T. Buck, N. I. Syed, and A. B. Smit. 2009. Transcriptome analysis of the central nervous system of the mollusc Lymnaea stagnalis. Bmc Genomics 10.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail Littorina saxatilis. Journal of Evolutionary Biology 23:2004-2016.

Gibbs, P. E. 1996. Oviduct malformation as a sterilising effect of tributyltin (TBT)-induced imposex in Ocenebra erinacea (Gastropoda: Muricidae). Journal of Molluscan Studies 62:403-413.

Gibbs, P. E., and G. W. Bryan. 1986. Reproductive Failure in Populations of the Dog-Whelk, Nucella-Lapillus, Caused by Imposex Induced by Tributyltin from Antifouling Paints. Journal of the Marine Biological Association of the United Kingdom 66:767-777.

Gibbs, P. E., G. W. Bryan, P. L. Pascoe, and G. R. Burt. 1987. The use of the dog-whelk, Nucella lapillus, as an indicator of tributyltin (TBT) contamination. Journal of the Marine Biological Association of the United Kingdom 67:507-523.

Gotz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talon, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36:3420-3435.

Guerra-Varela, J., I. Colson, T. Backeljau, K. Breugelmans, R. N. Hughes, and E. Rolan-Alvarez. 2009. The evolutionary mechanism maintaining shell shape and molecular differentiation between two ecotypes of the dogwhelk Nucella lapillus. Evolutionary Ecology 23:261-280.

Hahn, D. A., G. J. Ragland, D. D. Shoemaker, and D. L. Denlinger. 2009. Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly Sarcophaga crassipalpis. Bmc Genomics 10.

Harino, H., S. C. M. O'Hara, G. R. Burt, B. S. Chesman, and W. J. Langston. 2005. Distribution of organotin compounds in tissues of mussels Mytilus edulis and clams Mya arenariai. Chemosphere 58:877-881.

Harino, H., S. C. M. O'Hara, G. R. Burt, N. D. Pope, B. S. Chesman, and W. J. Langston. 2002. Butyltin and phenyltin compounds in eels (Anguilla anguilla). Journal of the Marine Biological Association of the United Kingdom 82:893-901.

Horiguchi, T. 2006. Masculinization of female gastropod mollusks induced by organotin compounds, focusing on mechanism of actions of tributyltin and triphenyltin for development of imposex. Environmental Sciences 13:77-87.

Kirby, R. R. 2000. An ancient transpecific polymorphism shows extreme divergence in a multitrait cline in an intertidal snail (Nucella lapillus (L.)). Molecular Biology and Evolution 17:1816-1825.

Korbel, J. O., A. E. Urban, J. P. Affourtit, B. Godwin, F. Grubert, J. F. Simons, P. M. Kim, D. Palejev, N. J. Carriero, L. Du, B. E. Taillon, Z. T. Chen, A. Tanzer, A. C. E. Saunders, J. X. Chi, F. T. Yang, N. P. Carter, M. E. Hurles, S. M. Weissman, T. T. Harkins, M. B. Gerstein, M. Egholm, and M. Snyder. 2007. Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420-426.

Kristiansson, E., N. Asker, L. Forlin, and D. G. J. Larsson. 2009. Characterization of the Zoarces viviparus liver transcriptome using massively parallel pyrosequencing. Bmc Genomics 10.

Lee, Y. S., S. L. Choi, T. H. Kim, J. A. Lee, H. K. Kim, H. Kim, D. J. Jang, J. J. Lee, S. Lee, G. S. Sin, C. B. Kim, Y. Suzuki, S. Sugano, T. Kubo, L. L. Moroz, E. R. Kandel, J. Bhak, and B. K. Kaang. 2008. Transcriptome analysis and identification of regulators for long-term plasticity in Aplysia kurodai. Proceedings of the National Academy of Sciences of the United States of America 105:18602-18607.

Li, W. Z., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Libants, S., K. Carr, H. Wu, J. H. Teeter, Y. W. Chung-Davidson, Z. P. Zhang, C. Wilkerson, and W. M. Li. 2009. The sea lamprey Petromyzon marinus genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. Bmc Evolutionary Biology 9.

Lockyer, A. E., J. Spinks, R. A. Kane, K. F. Hoffmann, J. M. Fitzpatrick, D. Rollinson, L. R. Noble, and C. S. Jones. 2008. Biomphalaria glabrata transcriptome: cDNA microarray profiling identifies resistant- and susceptible-specific gene expression in haemocytes from snail strains exposed to Schistosoma mansoni. Bmc Genomics 9.

Lockyer, A. E., J. N. Spinks, A. J. Walker, R. A. Kane, L. R. Noble, D. Rollinson, E. Dias-Neto, and C. S. Jones. 2007. Biomphalaria glabrata transcriptome: Identification of cell-signalling, transcriptional control and immune-related genes from open reading frame expressed sequence tags (ORESTES). Developmental and Comparative Immunology 31:763-782.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Matthiessen, P., and P. E. Gibbs. 1998. Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. Environmental Toxicology and Chemistry 17:37-43.

Meyer, E., G. V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. K. Colbourne, B. L. Willis, and M. V. Matz. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. Bmc Genomics 10.

Mitta, G., R. Galinier, P. Tisseyre, J. F. Allienne, Y. Girerd-Chambaz, F. Guillou, A. Bouchut, and C. Coustau. 2005. Gene discovery and expression analysis of immune-relevant genes from Biomphalaria glabrata hemocytes. Developmental and Comparative Immunology 29:393-407.

Morcillo, Y., A. Albalat, and C. Porte. 1999. Mussels as sentinels of organotin pollution: Bioaccumulation and effects on P450-mediated aromatase activity. Environmental Toxicology and Chemistry 18:1203-1208.

Morcillo, Y., and C. Porte. 1999. Evidence of endocrine disruption in the imposex-affected gastropod Bolinus brandaris. Environmental Research 81:349-354.

Moroz, L. L., J. R. Edwards, S. V. Puthanveettil, A. B. Kohn, T. Hla, A. Heyland, L. Knudsen, A. Sahni, F. H. Yu, L. Liu, S. Jezzini, P. Lovell, W. Iannucculli, M. C. Chen, T. Nguyen, H. T. Sheng, R. Shaw, S. Kalachikov, Y. V. Panchin, W. Farmerie, J. J. Russo, J. Y. Ju, and E. R. Kandel. 2006. Neuronal transcriptome of Aplysia: neuronal compartments and circuitry. Cell 127:1453-1467.

Murata, S., S. Takahashi, T. Agusa, N. J. Thomas, K. Kannan, and S. Tanabe. 2008. Contamination status and accumulation profiles of organotins in sea otters (Enhydra lutris) found dead along the coasts of California, Washington, Alaska (USA), and Kamchatka (Russia). Marine Pollution Bulletin 56:641-649.

Niu, B. F., L. M. Fu, S. L. Sun, and W. Z. Li. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. Bmc Bioinformatics 11.

Novaes, E., D. R. Drost, W. G. Farmerie, G. J. Pappas, D. Grattapaglia, R. R. Sederoff, and M. Kirst. 2008. High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. Bmc Genomics 9.

Oliveira, I. B., C. A. Richardson, A. C. Sousa, S. Takahashi, S. Tanabe, and C. M. Barroso. 2009. Spatial and temporal evolution of imposex in dogwhelk Nucella lapillus (L.) populations from North Wales, UK. Journal of Environmental Monitoring 11:1462-1468.

Parchman, T. L., K. S. Geist, J. A. Grahnen, C. W. Benkman, and C. A. Buerkle. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. Bmc Genomics 11.

Parkinson, J., A. Anthony, J. Wasmuth, R. Schmid, A. Hedley, and M. Blaxter. 2004. PartiGene--constructing partial genomes. Bioinformatics 20:1398-1404.

Quinn, N. L., N. Levenkova, W. Chow, P. Bouffard, K. A. Boroevich, J. R. Knight, T. P. Jarvie, K. P. Lubieniecki, B. A. Desany, B. F. Koop, T. T. Harkins, and W. S. Davidson. 2008. Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. Bmc Genomics 9.

Saavedra, C., and E. Bachere. 2006. Bivalve genomics. Aquaculture 256:1-14.

Satkoski, J. A., R. S. Malhi, S. Kanthaswamy, R. Y. Tito, V. S. Malladi, and D. G. Smith. 2008. Pyrosequencing as a method for SNP identification in the rhesus macaque (Macaca mulatta). Bmc Genomics 9.

Shagin, D. A., D. V. Rebrikov, V. B. Kozhemyako, I. M. Altshuler, A. S. Shcheglov, P. A. Zhulidov, E. A. Bogdanova, D. B. Staroverov, V. A. Rasskazov, and S. Lukyanov. 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Research 12:1935-1942.

Spade, D. J., R. J. Griffitt, L. Liu, N. J. Brown-Peterson, K. J. Kroll, A. Feswick, R. A. Glazer, D. S. Barber, and N. D. Denslow. 2010. Queen Conch (Strombus gigas) Testis Regresses during the Reproductive Season at Nearshore Sites in the Florida Keys. Plos One 5.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25:705-712.

Sugarbaker, D. J., W. G. Richards, G. J. Gordon, L. Dong, A. De Rienzo, G. Maulik, J. N. Glickman, L. R. Chirieac, M. L. Hartman, B. E. Taillon, L. Du, P. Bouffard, S. F. Kingsmore, N. A. Miller, A. D. Farmer, R. V. Jensen, S. R. Gullans, and R. Bueno. 2008. Transcriptome sequencing of malignant pleural mesothelioma tumors. Proceedings of the National Academy of Sciences of the United States of America 105:3521-3526.

Tangphatsornruang, S., P. Somta, P. Uthaipaisanwong, J. Chanprasert, D. Sangsrakru, W. Seehalak, W. Sommanas, S. Tragoonrung, and P. Srinives. 2009. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (Vigna radiata (L.) Wilczek). Bmc Plant Biology 9.

Tanguy, A., N. Bierne, C. Saavedra, B. Pina, E. Bachere, M. Kube, E. Bazin, F. Bonhomme, P. Boudry, V. Boulo, I. Boutet, L. Cancela, C. Dossat, P. Favrel, A. Huvet, S. Jarque, D. Jollivet, S. Klages, S. Lapegue, R. Leite, J. Moal, D. Moraga, R. Reinhardt, J. F. Samain, E. Zouros, and A. Canario. 2008. Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. Gene 408:27-36.

Toth, A. L., K. Varala, T. C. Newman, F. E. Miguez, S. K. Hutchison, D. A. Willoughby, J. F. Simons, M. Egholm, J. H. Hunt, M. E. Hudson, and G. E. Robinson. 2007. Wasp gene expression

supports an evolutionary link between maternal behavior and eusociality. Science 318:441-444.

Trussell, G. C., and R. J. Etter. 2001. Integrating genetic and environmental forces that shape the evolution of geographic variation in a marine snail. Genetica 112:321-337.

Vera, J. C., C. W. Wheat, H. W. Fescemyer, M. J. Frilander, D. L. Crawford, I. Hanski, and J. H. Marden. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636-1647.

Wheat, C. W. 2008. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138:433-451.

Wheeler, B. M., A. M. Heimberg, V. N. Moy, E. A. Sperling, T. W. Holstein, S. Heber, and K. J. Peterson. 2009. The deep evolution of metazoan microRNAs. Evolution & Development 11:50-68.

Wicker, T., E. Schlagenhauf, A. Graner, T. J. Close, B. Keller, and N. Stein. 2006. 454 sequencing put to the test using the complex genome of barley. Bmc Genomics 7.

Wicker, T., S. Taudien, A. Houben, B. Keller, A. Graner, M. Platzer, and N. Stein. 2009. A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant Journal 59:712-722.

Woyke, T., G. Xie, A. Copeland, J. M. Gonzalez, C. Han, H. Kiss, J. H. Saw, P. Senin, C. Yang, S. Chatterji, J. F. Cheng, J. A. Eisen, M. E. Sieracki, and R. Stepanauskas. 2009. Assembling the Marine Metagenome, One Cell at a Time. Plos One 4.

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert. 2001. Reverse transcriptase template switching: A SMART (TM) approach for full-length cDNA library construction. Biotechniques 30:892-897.

Zhulidov, P. A., E. A. Bogdanova, A. S. Shcheglov, L. L. Vagner, G. L. Khaspekov, V. B. Kozhemyako, M. V. Matz, E. Meleshkevitch, L. L. Moroz, S. A. Lukyanov, and D. A. Shagin. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Research 32.

CHAPTER 3

Endocrine disruption in the marine gastropod Nucella lapillus: combination of pyrosequencing and microarray technologies as a diagnostic instrument

Chapter submitted as part of an original article:

Pascoal S, Carvalho G, Vasieva O, Hughes R, Cossins A, Fang Y, Ashelford K, Olohan L, Barroso C, Mendo S, Creer S. A Transcriptomic Perspective of Imposex Endocrine Disruption in an Ecological Sentinel (Nucella lapillus) Suggests Involvement of RXR:PPARγ Signalling. PloS Biology

3.1. ABSTRACT

Gene-environment interactions using ecological model species are now facilitated by the use of contemporary functional genomic tools. Imposex, the superimposition of male sexual characteristics in females, is mainly caused by tributyltin (TBT) contamination and provides one of the best examples of anthropogenically-induced endocrine disruption in aquatic ecosystems. The dogwhelk, Nucella lapillus, is a recognized sentinel for TBT-induced imposex and although the phenomenon is well documented ecologically, less is known at the genomic and biochemical level. Herein, we applied a combination of 454 Roche sequencing and microarray (Agilent 4*180K) technologies to elucidate the mechanisms underpinning this poorly understood, but widespread gene-environment interaction. Transcript data were assembled and used to reconstruct and annotate a partial Nucella transcriptome (chapter 2), from which a 180,000-feature oligonucleotide array was designed. Microarray analyses for environmental/functional genomic interpretation of the imposex mechanism support and complement (e.g. CypA71, CypB71 and CypA391) the 3 previously suggested hypotheses and show evidence for their interaction. Some new targets for TBT: nuclear receptors and transcription factors (e.g. RAR, ROR, Rev-Erba, ARNT, SRY and PPARγ) and their related pathways were identified. Impacts on the immune system, cell proliferation and apoptosis, DNA repair and tumour suppressors were evident and the possibility for a TBT-inhibited-transporter-based hypothesis is also suggested. TBT seems to be a multi-site binding compound and the contribution of several causal effects and their interaction is evident. Moreover, the generated data support a common mechanism in signalling of endocrine disruption along taxa. The produced genomic resources largely contribute to the molecular understanding of imposex and provide valuable insights for further examination of responses to TBT contamination exposure. We anticipate that the produced tools will represent a platform for exploring additional responses of dogwhelks to environmental variation.

3.2. INTRODUCTION

Mainly due to industrialization, man-made endocrine disruptors have been deliberatively released into the environment with particularly negative impacts in aquatic environments. Endocrine disruption is of worldwide concern affecting a large taxonomic range of organisms that may present various deleterious effects, including malformations in frogs and decreased immunity in cetaceans (Ratcliff 1970; Guillette et al. 1994; Sumpter and Johnson 2005; Murata et al. 2008) and may, ultimately, lead to local population extinction (Bryan et al. 1986; Gibbs et al. 1991; Huet et al. 1996). Consequently it has been a major topic of research in recent years (Ketata et al. 2008).

Most information on the biological effects and mechanisms of action of endocrine disruptors (EDs) has been focused on vertebrates (Porte et al. 2006). To date, most of these studies have been directed to the reproductive processes regulated by hormonal signalling mediated by members of the family of nuclear receptors and thyroid hormone receptors. More recently, retinoid X receptor (RXR) and peroxisome-activated receptors have also been investigated (PPARs), (Swedenborg et al. 2009). EDs can affect these targets in different ways, mainly, by directly interfering with receptors signalling or by activating other signalling pathways, in particular the aryl hydrocarbon receptor (AhR) that is envolved in the metabolism of many xenobiotic substances (Swedenborg et al. 2009). Gene expression is regulated at several levels; transcription level, post-transcription level with mRNA formation, mRNA translation with post-translational processing (Migeon and Wisniewski 2000). Transcription factors (TFs) are trans-acting elements that bind selectively to the appropriate cis-acting DNA sequences of a gene promoter that can act as activators or repressors of gene transcription. TFs are vital for many important cellular processes (e.g. basal transcription regulation, development, response to stressors) and particularly, it has been shown that TFs have a relevant role in human sex differentiation (Migeon and Wisniewski 2000).

The imposex phenomenon (superimposition of non-functional male sexual characteristics in females (Morcillo and Porte 1999; Ketata et al. 2008)) provides a key ecological example of anthropogenically-induced endocrine disruption in aquatic ecosystems (Matthiessen and Gibbs 1998). The marine gastropod, Nucella lapillus, is widely used for biomonitoring, via the imposex response, since it is very sensitive to TBT pollution (Gibbs et al. 1987; Oliveira et al. 2009). Despite intensive research, mainly related with the ecological and population impacts of TBT (Horiguchi 2006; Castro et al. 2007a), the underlying molecular and biochemical mechanisms associated with this phenomenon remain poorly understood.

Nevertheless, three main hypotheses for the TBT-induced imposex - steroid, neuroendocrine and retinoid - have been suggested so far, and have contributed to our understanding of the endocrine system and reproductive physiology of gastropods (Sternberg et al. 2010). Briefly, the steroid hypothesis proposes that imposex is caused by an increase of free testosterone levels in exposed females that may involve the inhibition of enzymes that metabolize steroids. The main targets that have been proposed are the Cytochrome P450 aromatase (Spooner et al. 1991; Bettin et al. 1996; Santos et al. 2002), sulfotransferase (Ronis and Mason 1996) and acyl coenzyme A: steroid acyltransferase (Gooding et al. 2003; Janer et al. 2005). The neuroendocrine hypothesis proposes that TBT acts as a neurotoxicant in gastropods causing the aberrant secretion of neurohormones that contribute to male sexual differentiation (Féral et al. 1983; Oberdorster and McClellan-Green 2000a, 2002). Finally, the retinoid hypothesis is the most recent, suggesting that the retinoid X receptor (RXR), a nuclear receptor, plays an important role in inducing the development of imposex (Nishikawa et al. 2004; Nishikawa 2006; Castro et al. 2007a; Horiguchi et al. 2007). Furthermore, it suggests that organotins may mimic the endogenous ligand of RXR, disrupting the signalling pathway that are retinoic acid dependent (Castro et al. 2007a). In support, more recently, it was suggested that RXR-mediated signalling may have an important role in sex differentiation and that TBT would induce imposex by initiating RXR signalling prematurely in females (Sternberg et al. 2008). The interplay between the suggested pathways is presumably complex, and remains unclear, and not all the claimed molecular targets have been identified in

gastropods (Castro et al. 2007a). Therefore, further investigation on the molecular mechanisms of the imposex is warranted as it has been largely limited by the lack of adequate genomic resources in target species.

Recent technological advances in high throughput sequencing (e.g. 454 Roche sequencing) have greatly enhanced the genomic toolkits available to gain functional genomics insights, even with non-model species (Margulies et al. 2005; Ellegren 2008). Array-based technologies have been the main platforms for undertaking large-scale gene expression screens and are invaluable genomic resources for a global overview of how the expression of each gene in the organism is responding to a particular stressor (Gracey 2007). For non-model species, however, the generation of the underpinning resources have been a problem (Gracey and Cossins 2003). The emergence of second-generation sequencing is now, however, filling this gap providing the necessary datasets for array design for any species of interest. Consequently, combinations of such contemporary technologies - pyrosequencing and microarray - are recently being applied, and are predicted to enable the study of key gene-environment interactions in non-model organisms, that were difficult to address so far (e.g. Spade et al. 2010).

Here, combinations of ultrasequencing and microarray technology were applied to disentangle the functional genomic mechanism of imposex using N. lapillus as a model organism. Accordingly, a N. lapillus oligonucleotide array was designed from the partial Nucella transcriptome obtained in chapter 2. Microarray analyses for differential gene expression in response to TBT contamination were performed revealing that it is not possible to discriminate equivocally between the 3 suggested hypotheses. However, supporting evidence that all pathways are involved in the imposex phenomenon and all interact in gene differential expression is presented. Additionally, some new TBT targets and related pathways involved in the toxic chain were identified in the Nucella response to TBT contamination. We anticipate that the produced transcriptomic data and subsequent genomic tools will be valuable resources for further exploration of the genomic basis of the imposex mechanism and potentially other environmental and evolutionary processes in N. lapillus.

3.3. MATERIAL AND METHODS

3.3.1. RNA samples for array hybridizations

Test dogwhelk females (N=16) were obtained from the laboratory exposure to TBTCl that is described in chapter 2. A total of 48 individual RNA extractions (by tissue: digestive gland (H), ovary (O) and penis (I) from 4 biological replicates per treatment) were performed with the Trizol (Invitrogen) method from control and TBT exposed females (200 ng Sn/L) after 1 week and 3 months of exposure. Total RNA was DNase treated (Qiagen RNase-free DNase set) and cleaned (Qiagen RNeasy MinElute cleanup Kit) before final quality control (QC) using an Agilent 2100 bioanalyser. A RNA integrity number (RIN) was not available due to the absence of the 28S rRNA band from undegraded total RNA. This is possibly related to "hidden break" in invertebrates (Ishikawa 1977; Spade et al. 2010), also observed here in N. lapillus.

3.3.2. Microarray probe design

A custom gene expression microarray composed of approximately 180,000 60-mer oligonucleotide probes was designed in Agilent's web-based application, eArray. Of the 88,016 sequences selected for representation on the array (obtained from chapter 2), 84,392 sense and 84,390 anti-sense probes were successfully designed. A set of 15 "non-control" sense probes corresponding to housekeeping genes, were also designed; these were used for subsequent use in the analysis of spatial variance during data acquisition from the scan images. The non-control probes were replicated 10 times across the array and Agilent's QC grid, which includes negative control probes for use in background subtraction and positive control probes complementary to spike-in transcripts, was also added to the array. The remainder of the array was filled with randomly selected replicate sense probes. A random probe layout was selected with a format of four 180K probe arrays per slide. The final array design was submitted to Agilent for array fabrication.

3.3.3. Preparation of labelled target

Fluorescently labelled amplified complementary RNA (cRNA) was generated using a Quick Amp Labelling kit (Agilent), according to the manufacturer's instructions. The method employs an oligo (dT) primer bearing a T7 promoter and MMLV-RT to produce double stranded cDNA from mRNA; the cDNA then serves as template for in vitro transcription with T7 RNA Polymerase, which linearly amplifies target material whilst simultaneously incorporating cyanine 3- or cyanine 5-labelled CTP. For each labelling reaction, 200 ng of quality assured total RNA was used as input, along with appropriately diluted spike mixtures from the Two-colour RNA Spike-In kit (Agilent). cRNA was purified using an RNeasy Mini Kit (Qiagen) and quantified on a NanoDrop ND-1000 Spectrophotometer version 3.3.0.

3.3.4. Array hybridization

Samples to be co-hybridized to the same array were combined (825 ng each of cyanine 3- and cyanine 5-labelled cRNA), together with 11 µl of 10x Blocking Agent and 2.2 µl of 25x Fragmentation Buffer (both from the Agilent Gene Expression Hybridization kit) to a total volume of 55 µl. Target mixtures were then incubated at 60°C for 30 min to fragment the RNA to approximately 150 nucleotides. Fragmentation was terminated by the addition of 55 µl of 2x GEx Hybridization Buffer HI-RPM (Agilent). Microarrays were loaded and hybridized using Agilent hardware, namely, gasket slides, SureHyb chambers and hybridization oven. Hybridization was carried out at 65°C with rotation at 10 rpm for 17 h. After this time, microarrays were washed using an Agilent Gene Expression Wash Buffer provided in the kit, according to the manufacturer's instructions. A final wash with Stabilization and Drying Solution to prevent cyanine 5 degradation by ozone was carried out. Arrays were then scanned at a resolution of 3 µm using an Agilent DNA Microarray Scanner to generate 20 bit tiff images. Data were extracted and QC reports generated using Agilent Feature Extraction version 10.5.1.1. Array quality was assessed by visual inspection of each tiff image and analysis of the associated QC report, which indicates the dynamic range of the experiment, hybridization and background uniformity, as well as an

evaluation of 11 metrics associated with the RNA spike-ins added to the labelling reactions.

3.3.5. Data analyses

The array contrasts were made for each of three tissues (H, I, O) at the beginning and the end of the TBT exposure experiment (week 1, week 13), generating six array groups, each one of them containing four biological replications. The analysis of the dogwhelk microarray data included data normalization, model based analysis, multivariate analysis and gene set based analysis. Firstly, the data was normalized using limma (http://www.bioconductor.org/help/bioc-views/release/bioc/), in which the within-array bias and inter-array bias were corrected through loess and quartile normalization respectively. A linear model which contains six parameters for six array groups was then generated for the normalized data, and the contrasts of two time points for each of the three tissues were also assessed by the modeling tool in limma package. The adjusted p-value for each probe by F-test which was available from the modeling tool of limma was used to extract differential expressed (DE) probes. A first group of DE probes was produced using a cut-off at 5% (long list) and a second group of DE probes was obtained (shortlist) reinforcing the criterion by a fold change threshold (fold change greater $\sqrt{2}$).

In order to attribute a biological meaning to the produced data, gene annotation was obtained by BLAST searches against the NCBI non-redundant protein (NR) and nucleotide (NT) databases using different e-values thresholds, BLAST2GO and AMIGO were used to access the associated Gene ontology (GO) terms.

Two multivariate analysis tools were conducted on the DE probes. One, principal component analysis (PCA) was applied to the whole dataset and to the partial datasets which contained only the DE probes. The second approach, based on the cluster method was applied only to the shortlisted DE probes.

Finally, to test for significant GO terms and to evaluate the biological meaning or randomness of the DE probes GOs, a GO based gene set analysis was applied to the DE-replicated results of the data model, in which annotated probes were DE-replicated to unique genes and probes without any annotation were excluded. A difficulty arose here as only a small number of probes have been annotated by GO and the annotated DE genes were too few to perform hypergeometric test method analyses. Consequently, a rank test method was employed to overcome this problem that ranks the significance of all the annotated genes and uses them, rather than merely the annotated DE genes.

3.3.6. Pathway analysis

All the 454 contig sequences from each experiment (control and treated libraries assembled individually or from co-assembled libraries) were submitted to KAAS (KEGG - Kyoto encyclopedia of genes and genomes - automated annotation server http://www.genome.jp/tools/kaas/). The KAAS server provides the capability to annotate genomes and collections of ESTs using KEGG's families of orthologous sequences and to project annotated functions onto KEGG's collection of metabolic and regulatory pathways. KAAS's reference eukaryotic genome set has been used for SDH (single-directional hit) orthology searches (similarity index threshold: 60). We also exploited a minimal threshold (30) in a search for particularly interesting missing functions. KEGG's Brite hierarchical classification of functions was used to retrieve the functional groups of interest from the annotated data. IPA software (Ingenuity http://www.ingenuity.com) was applied to generate a protein regulatory network from a manually composed list of functions related to gender regulation and development, known lipophilic nuclear receptors and functions that were shown to be involved in binding of TBT. The Washington University Biomedical Informatics Core (http://bioinformatics.wustl.edu/webTools/PromoterAnalysis.do) was employed to analyse Transcription factor (TF) binding sites in promoters of eukaryotic groups of isofunctional homologous genes.

3.3.7. qPCR validation

For microarray validation, quantitative real-time PCR (qPCR) analyses were performed using the same 48 purified total RNA samples as for the microarray hybridizations. Five transcripts of interest (selected from the transcriptome sequences obtained in chapter 2) were used to develop primers for SYBR-green detection (www.designmyprobe.com) and a Nucella reference gene (Actin, that was not differentially expressed in our microarray study) was selected from the literature (Castro et al. 2007b), (Table 1). A two step reaction qPCR was performed. In the first step, 0.5 µg of purified total RNA was used to synthesise cDNA using the SuperScript III first-strand synthesis super mix for qRT-PCR (Invitrogen). In the second step, cDNA samples were analysed using the MESA qPCR MasterMix Plus for SYBR Assay sample (Eurogentec) following the manufacturer´s protocol. For real time quantification, reactions were run in the 7900HT Fast Real-Time PCR System (Applied Biosystems) using the standard protocol, that included an initial denaturation at 95 ^{o}C for 10 min, followed by 40 cycles of denaturation at 95 ^{o}C for 15 sec and annealing at 60 ^{o}C for 60 sec. Preliminary amplifications were performed in order to optimize primer concentrations and cDNA volume, and to verify primer specificity to each assay by adding a dissociation curve after amplification. Standard curves to assess PCR efficiency for both the gene of interest and the reference gene were also performed. Standards and experimental samples were run in duplicate along with a no-template control (NTC) by gene. For each gene, its expression in each tissue was normalized to Actin and the fold change due to treatment was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

Table 1. Primer sequences for qPCR.

Primer name	Primer sequence 5'- 3'	Amplicon size (bp)
DetoxF	CGTCTGACATCCGATACA	122
DetoxR	GTCCATTTCTTGTTCCATCA	
ProgF	TCCAAGACCTACTACACT	89
ProgR	AAGATGAGCCAGGAATAG	
EstF	TTGTGACTATTGCCAGTT	189
EstR	TTGCCTTTCTTATACCCTAA	
TBTbpF	TCTGTCACAACCTGGAAT	114
TBTbpR	TGGATGTTAGACTAAGGGAAA	
PPARγF	ATAATACTCAGAGCGGCTAG	117
PPARγR	AACAATCAATCACAGACACAT	
ActinF*	GCCGTGACCTGACAGACTACC	
ActinR*	CCTTGATGCCACTGACGATTTCC	

Transcripts: Detox: carbamoyl-phosphatase synthase; Prog: G-protein coupled progestin receptor; Est: Estrogen receptor; TBTbp: TBT binding protein; PPARγ: Peroxisome proliferator-activated receptor gamma; Actin*: Actin gene as described by (Castro et al. 2007b).

3.4. RESULTS

3.4.1. Microarray results

The oligoarray hybridization analyses resulted in concerted patterns of differential gene expression in response to TBT contamination with statistical precision and a high number of gene models (Fig. 1). A list of 13,075 differentially expressed (DE) probes were extracted using the 5% cut-off, and of these, 617 strongly DE genes were revealed if the selection criterion was reinforced by the fold change threshold. Accordingly, the 617 shortlisted genes were identified as target sequences of interest (candidate genes) for the gastropod response to TBT contamination and were used as the basis for the subsequent analyses.

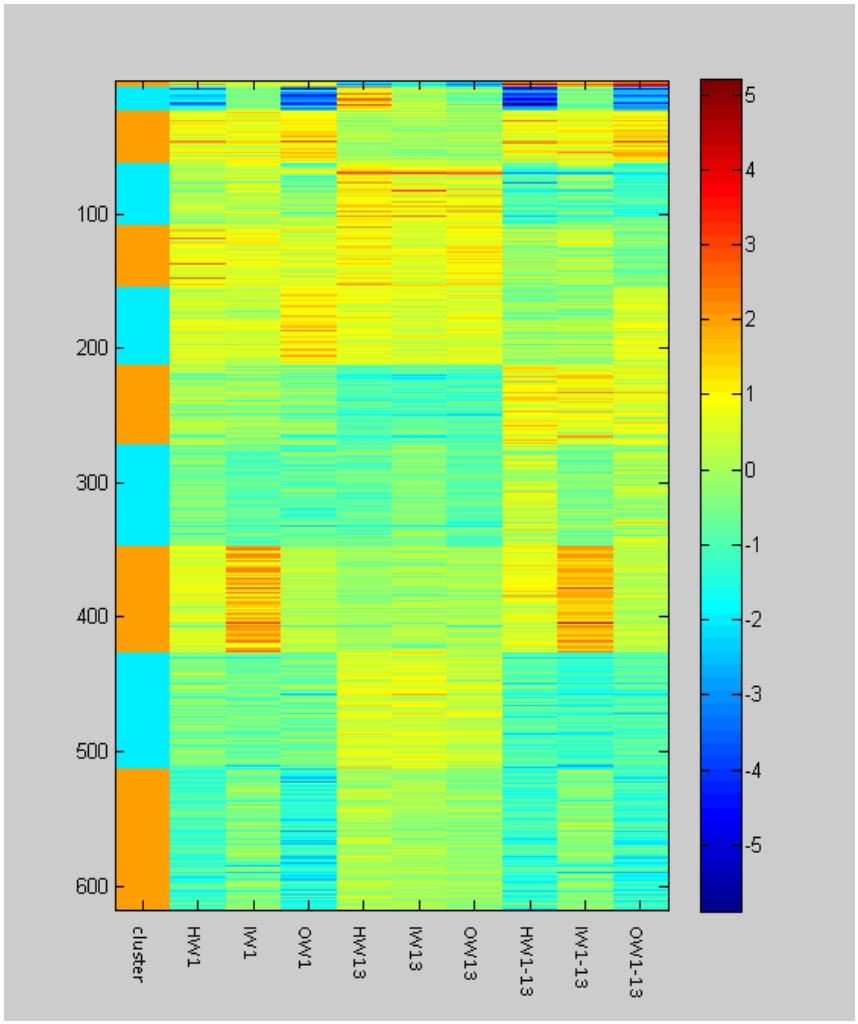


Figure 1. Heat-map resulting from the microarray analysis after clustering. HW1: Digestive gland week 1; IW1: Penis week 1; OW1: Ovary week 1; HW13: Digestive gland week 13; IW13: Penis week 13; OW13: Ovary week 13. Heat-map scale is $\log_2$ based. Blue represents down-regulation and red up-regulation.

The principal component analysis (PCA) results revealed a much clearer separation of array groups using the DE probes (Fig. 2) than the whole dataset. This implies that DE probes were well identified and that useful information may be derived from them. From the cluster method, data were grouped into 11 clusters (Fig. 1), displaying useful information about the patterns of gene response across the tissues and at different time points. These methods showed a clear time and tissue-specific response to contamination.

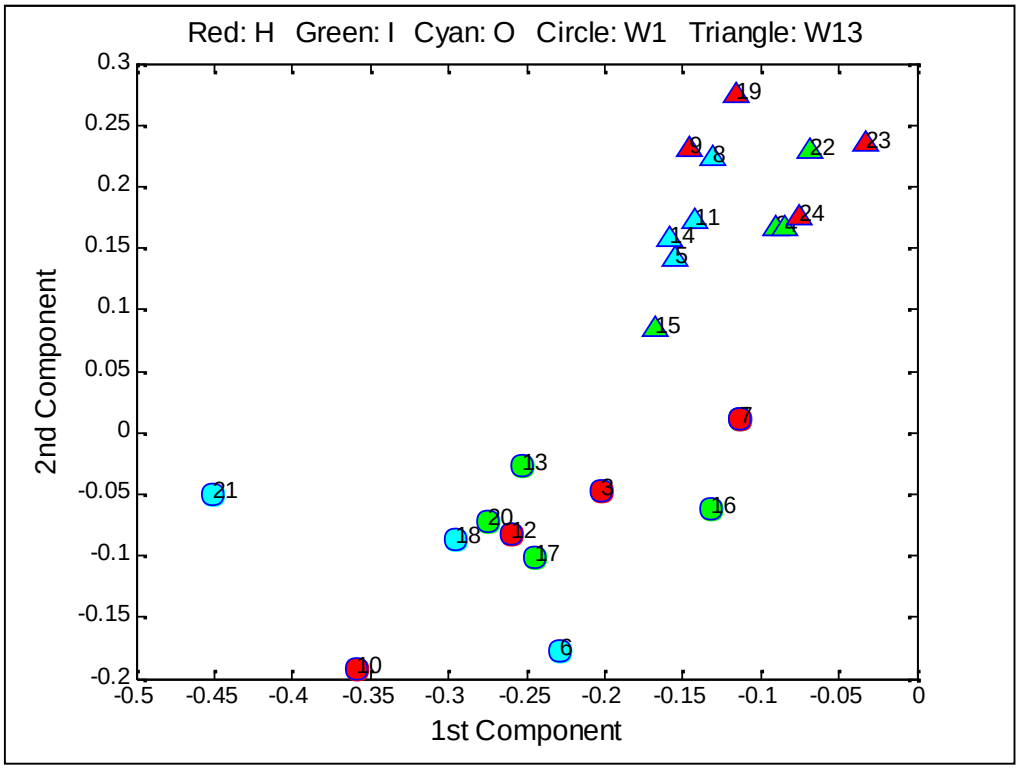


Figure 2. PCA based data of differentially expressed gene probes for the most differentially expressed genes. H: Hepatopancreas; I: Penis; O: Ovary; W1: Week 1; W13: Week 13.

3.4.2. Gene annotation and gene set analysis by GO classification

From the 617 candidate sequences, we were able to annotate 87 sequences performing a BLAST search against the NCBI NR database (e-value threshold≤10^{-6}) and using BLAST2GO, 64 sequences could be associated to a GO term (Fig. 3).

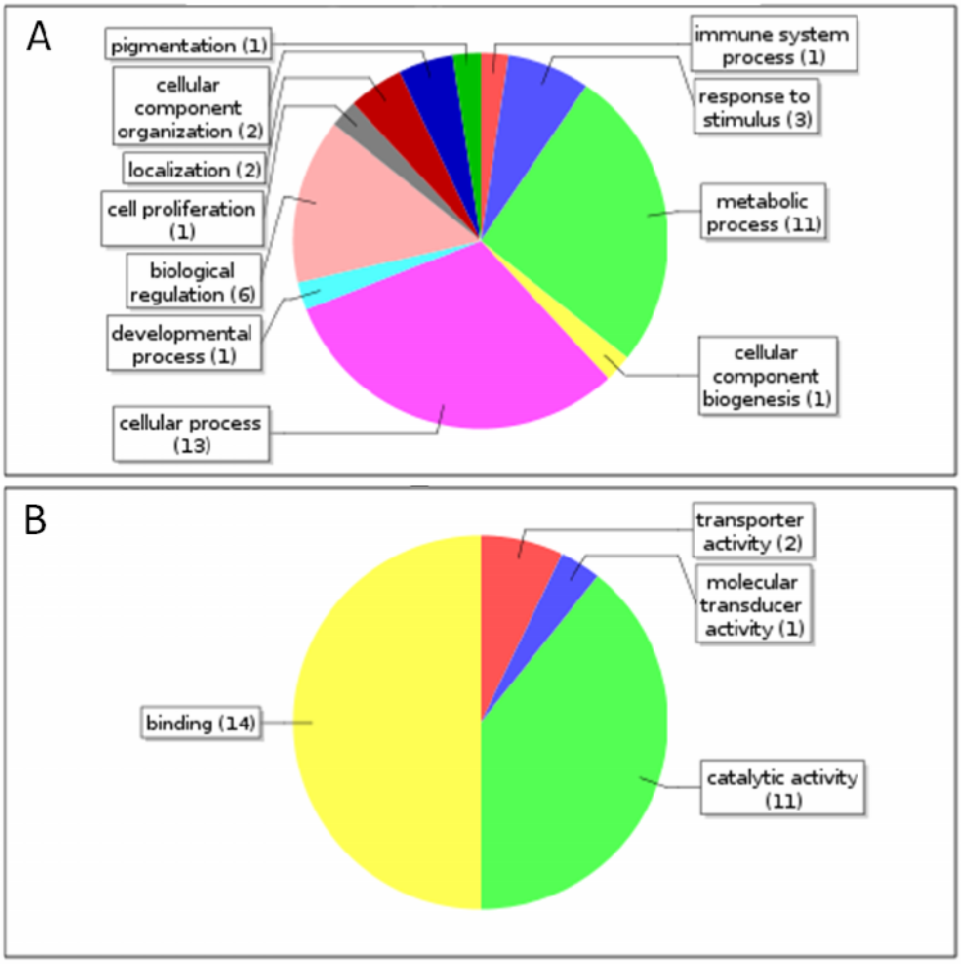


Figure 3. Microarray shortlist (617) candidate genes gene ontology; A) Biological process; B) Molecular function.

Due to the lack of gene annotation in gastropods we conducted another "gene-finding" blast search, using a very relaxed e-value threshold (10^2). The latter strategy was adopted to serve as a "capture all" search, facilitating the downstream scrutiny of putative annotations via reference to biology, physiology and homology. Using this approach, 520 sequences with a gene annotation (please see appendix 3.1) were identified; from these, 218 had an associated GO term. From the GO based analysis using the rank test method, no significant GO terms were identified for all the contrasts except "PenisW13treated" vs "PenisW13control". Here, 20 GO terms in "Biological process" were identified as significant GO terms (Table 2) with the highest level of significant GO

terms being regulation of translation and rRNA process. All the other significant GO terms were the ancestors of the two terms indicating biological significance in the observed annotations. It seems that this experimental condition differs greatly in gene functional annotation and that translation is being highly affected. Since there is a biological meaning in GO terms, the annotations obtained by gene finding were also used in further analysis.

Table 2. Significant GO term represented in the Penis w13 response, resultant from the IW13T vs IW13C contrast. DE: differentially expressed (please see appendix 3.2 for diagram view).

	DE_GOid	DE_GOterm	P. value	q. value
1	GO:0043170	macromolecule metabolic process	0.001357	0.054148
2	GO:0009058	biosynthetic process	0.001832	0.069444
3	GO:0044249	cellular biosynthetic process	0.000184	0.012689
4	GO:0009059	macromolecule biosynthetic process	2.04E-05	0.003487
5	GO:0043283	biopolymer metabolic process	0.000461	0.025861
6	GO:0010467	gene expression	3.90E-07	0.000295
7	GO:0044260	cellular macromolecule metabolic process	0.00053	0.026766
8	GO:0034960	cellular biopolymer metabolic process	0.000478	0.025861
9	GO:0043284	biopolymer biosynthetic process	6.41E-05	0.006077
10	GO:0019538	protein metabolic process	0.00014	0.011505
11	GO:0042254	ribosome biogenesis	0.000955	0.045245
12	GO:0034645	cellular macromolecule biosynthetic process	1.56E-05	0.003487
13	GO:0044267	cellular protein metabolic process	0.000152	0.011505
14	GO:0034961	cellular biopolymer biosynthetic process	6.41E-05	0.006077
15	GO:0006412	translation	3.14E-06	0.001189
16	GO:0034660	ncRNA metabolic process	0.001033	0.046055
17	GO:0016072	rRNA metabolic process	2.76E-05	0.003487
18	GO:0034470	ncRNA processing	0.000467	0.025861
19	GO:0006446	regulation of translational initiation	0.001167	0.049162
20	GO:0006364	rRNA processing	2.76E-05	0.003487

3.4.3. Network and pathway analysis involved in the response to TBT contamination

3.4.3.1. Previous hypotheses: involvement and gene expression interaction

According to the DE gene analyses there is evidence that all previously identified mechanisms (steroid, neuroendocrine and retinoid) are responsible for, and interact, during the imposex phenomenon in N. lapillus.

1. Steroid: Biosynthesis of steroid hormones seems to be affected in the tissues analysed since differential regulation of 3 enzymes - CypA71, CypB71 and CypA391 - that can change the steroid profile was observed. Searches for the transcription factor (TF) binding sites in the promoters of mammalian CypA71, CypB71 and CypA391 suggest that they can be regulated by gender-determination genes (Sex determining region Y (SRY), SRY-box9 (SOX9), GATA-4), or by lipophilic nuclear receptors (e.g. Retinoic acid-related orphan receptor (e.g. RORα)) (Fig. 4). Indeed, all these genes were significantly differentially expressed, predominantly by up-regulation in the ovary and hepatopancreas. Additionally, the sterol regulatory element binding protein (SREBP1), the PPARγ and the progestin receptor (nuclear receptor subfamily 3 (NR3C3)), that are involved in regulation of a number of cytochromes in vertebrates (Rogue et al. 2010; Inoue et al. 2011) and might be related to the cytochromes up-regulation in Nucella were also present in the dataset.

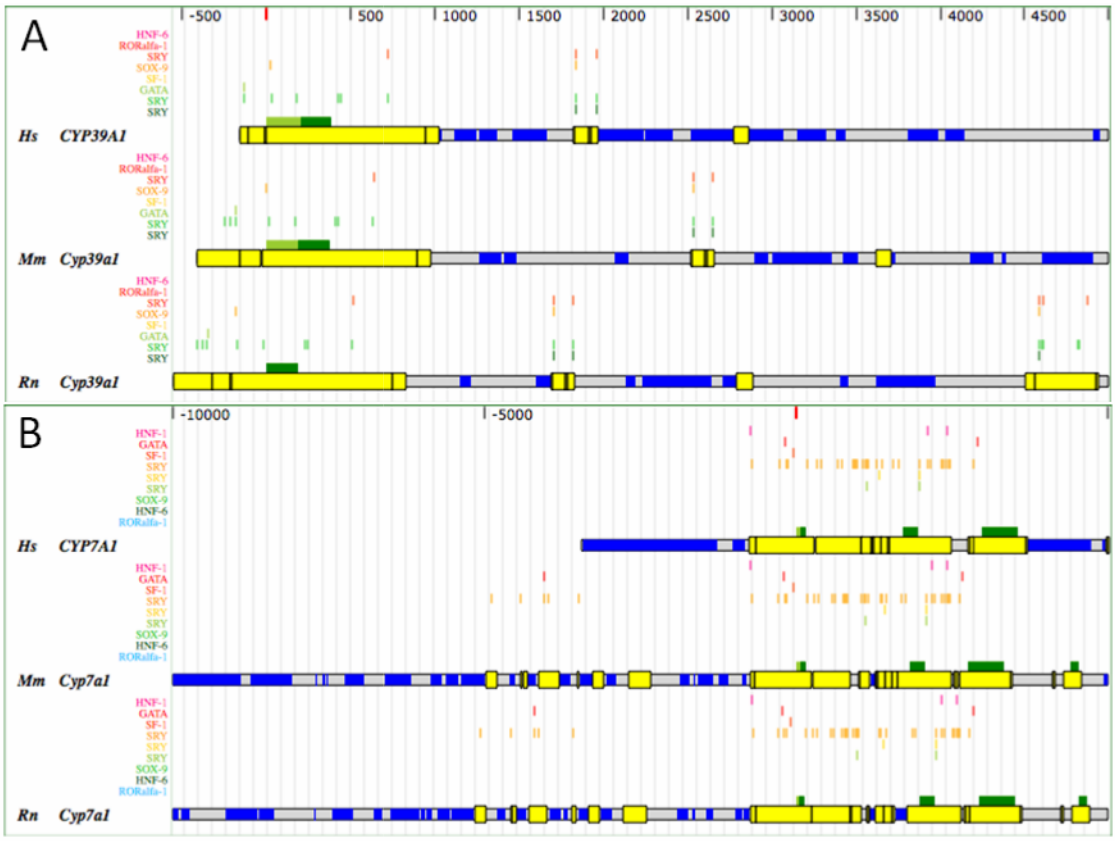


Figure 4. Illustration for Transcription factors binding sites distribution in promoter of Cyp39A1 (A) and Cyp7A1 (B).

2. Neuroendocrine: A number of neuro-regulatory receptors (e.g. opioid-like, odour and taste) that can be regulated by gender-related TFs and by nuclear receptors, were differentially expressed during the imposex response.

3. Retinoid: There was no direct evidence for the differential expression of RXR within the candidate sequences. However, a gene that shares the same histidine scaffold as the vertebrate PPARγ (PPARγ homolog), exhibited general up-regulation in all tissues and times throughout the observed endocrine disruption (Fig. 5).

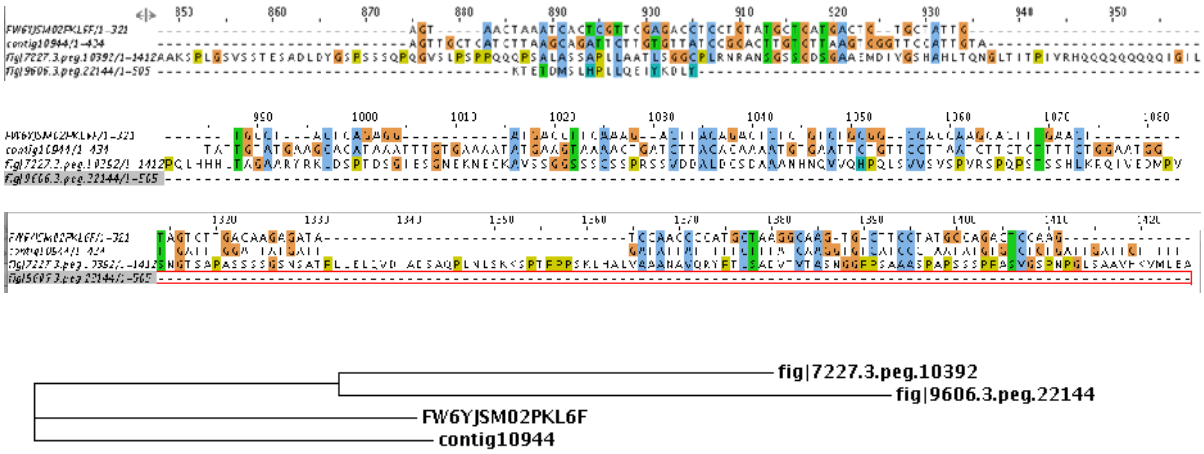


Figure 5. Peroxisome proliferator activated receptor gamma (PPARγ) homologue sequences alignments and phylogram. Human PPARγ: fig|7227.3.peg.10392; Drosophila PPARγ: fig|9606.3.peg.22144 and two Nucella sequences.

Further exploration of how gender determining factors crosstalk with nuclear receptors in vertebrates illustrate that hormonal effects are central to the network and link lipophilic TFs and gender determining factors sub-networks (Fig. 6).

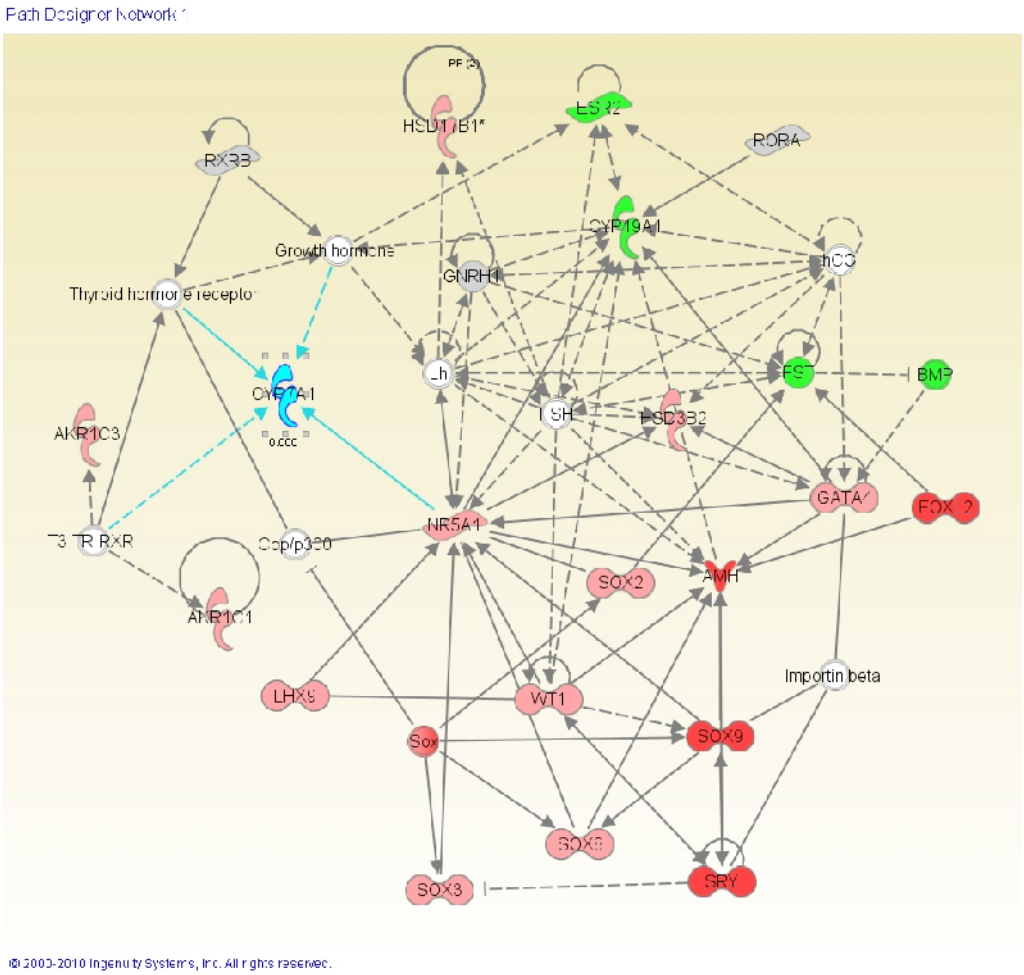


Figure 6. Network showing how gender-determining factors crosstalk with the nuclear receptors. Cyp7A1 is found in this network as well. Red: male-related functions; Green: female-related functions; Grey: nuclear receptors not yet related to gender determination; Blue: Cyp7A1.

Up-regulation of a potential homolog to TBT binding protein - a known TBT-binder (Oba et al. 2007) was observed in all the tissues and more differentially in the hepatopancreas, which may lead to a higher accumulation of TBT in this organ.

3.4.3.2. Identification of new potential gender-related TBT targets

Gene annotation of the transcriptome sequences via KAAS services allowed the identification of potential N. lapillus transcription factors, nuclear receptors (NR), G-protein coupled receptors and opioid receptors (as they are likely to bind the neural gender differentiation hormone) - that are gender-related and may be involved in the toxicity response. Considering the first order TBT binders such as PPARγ and TBTbp, we were able to identify TFs that may not be differentially expressed, but are putative first order binder candidates involved in the organismal response to TBT exposure. Besides nuclear receptors that require RXR for their transcription activation function such as retinoic acid receptor (RAR), liver X receptor (LXR), PPARs, Pregnane X receptor (PXR); those that are lipophilic orphan receptors: NR6A1, NR2C1, RORα, RORβ and AhR/AhR nuclear translocator-like (ARNTL), that may be involved in TBT binding (Table 3) were also found. Additional candidates included G-protein coupled receptors: GPR40, prostaglandin receptor (PTGIR) and receptor tyrosine kinase (RTK) receptors. The latter TFs, like ROR1 and ROR2 (also retinoic receptor), could be involved in initial TBT binding leading to inductions of gender-related TFs and a receptor for ecdysterol was also identified. Finally, an array of opioid receptors was delimited, since they are likely to bind the neural gender differentiation hormone (Table 3).

Table 3. List of potential transcription factors and receptors expressed in Nucella lapillus that may be involved in TBT binding and toxicity.

G-PROTEIN COUPLED RECEPTORS AND OPIOID RECEPTORS
1. Prostacyclin contig14149; K04263 PTGIR; prostacyclin receptor FW6YJSM01CSOS3; K04263 PTGIR; prostacyclin receptor
6. Free fatty acid FW6YJSM01DICE2; K04325 FFAR1, GPR40; free fatty acid receptor 1 contig13261; K04328 FFAR2, GPR43; free fatty acid receptor 2 contig05745; K08425 GPR120; G protein-coupled receptor 120
8. RTK class XII (ROR receptor family)-retinoic orphan receptors contig10606; K05122 ROR1, NTRKR1; receptor tyrosine kinase-like orphan receptor 1 FW6YJSM01B7YIZ; K05122 ROR1, NTRKR1; receptor tyrosine kinase-like orphan receptor 1 FW6YJSM01C72FZ; K05122 ROR1, NTRKR1; receptor tyrosine kinase-like orphan receptor 1 contig19298; K05123 ROR2, NTRKR2; receptor tyrosine kinase-like orphan receptor 2 FW6YJSM02SHB7A; K05123 ROR2, NTRKR2; receptor tyrosine kinase-like orphan receptor 2
9. Opioid

FW6YJSM02QDR8F; K04213 OPRD1; opioid receptor delta 1 FW6YJSM02PL85L; K04213 OPRD1; opioid receptor delta 1 contig12543; K04214 OPRK1; opioid receptor kappa 1 FW6YJSM02QFWZ8; K04215 OPRM1; opioid receptor mu 1 FW6YJSM01CCR0K; K04215 OPRM1; opioid receptor mu 1 FW6YJSM01CMSKM; K04215 OPRM1; opioid receptor mu 1
NUCLEAR RECEPTORS AND LIPOPHILIC ORPHAN RECEPTORS
0. Cys4 hepatocyte nuclear factor 4-like FW6YJSM02QVNL3; K08525 NR2B2, RXRB; retinoid X receptor beta FW6YJSM02R4H82; K08526 NR2B3, RXRG; retinoid X receptor gamma FW6YJSM01C43Z4; K08543 NR2C1, TR2; testicular receptor 2 -regulated by PPARg FW6YJSM01AO3TL; K14031 NR2CN; nuclear receptor subfamily 2 group C
1. Cys4 thyroid hormone-like FW6YJSM02PKL6F; K08530 NR1C3, PPARG; peroxisome proliferator-activated receptor gamma contig05927; K08532 NR1F1, RORA; RAR-related orphan receptor alpha FW6YJSM02RGHOV; K08532 NR1F1, RORA; RAR-related orphan receptor alpha contig27298; K08533 NR1F2, RORB; RAR-related orphan receptor beta contig01857; K08534 NR1F3, RORC; RAR-related orphan receptor gamma contig24966; K08534 NR1F3, RORC; RAR-related orphan receptor gamma FW6YJSM01B7891; K08535 NR1H2, LXRB; liver X receptor beta FW6YJSM01DTGHL; K08535 NR1H2, LXRB; liver X receptor beta FW6YJSM02SH67O; K14034 NR1H1, EcR; ecdysone receptor FW6YJSM02P374B; K08540 NR1I2, PXR; pregnane X receptor
2. Cys4 estrogen-like contig03196; K08550 NR3A1, ESR1; estrogen receptor alpha contig04559; K08551 NR3A2, ESR2; estrogen receptor beta contig08456; K08552 NR3B1, ESRRA; estrogen-related receptor alpha contig24650; K05771 NR3C1, GR; glucocorticoid receptor FW6YJSM02P983J; K08556 NR3C3, PGR; progesterone receptor
4. Cys4 Fushi tarazu-F1-like FW6YJSM02TDZIZ; K08560 NR5A1, SF1; steroidogenic factor 1 contig01962; K08561 NR6A1, GCNF; germ cell nuclear factor FW6YJSM01CR4QK; K08561 NR6A1, GCNF; germ cell nuclear factor-retinoic acid receptor-related testis-associated receptor
7. Factors with PAS domain contig26744; K09093 AHR; aryl hydrocarbon receptor FW6YJSM01C166A; K09093 AHR; aryl hydrocarbon receptor FW6YJSM01DDGO5; K09093 AHR; aryl hydrocarbon receptor contig11046; K02296 ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1 contig12355; K09099 ARNTL2, BMAL2; aryl hydrocarbon receptor nuclear translocator-like protein 2 contig17225; K09099 ARNTL2, BMAL2; aryl hydrocarbon receptor nuclear translocator-like protein 2

Additional TF searches in the transcriptomic data resulting from individual cDNA library assemblies - from non-treated animals and from animals exposed to TBT contamination - allowed the identification of some more TFs in the response to contamination (Table 4).

Table 4. Nucella lapillus transcription factors and receptors identified for individually assembled control and TBT treated libraries.

		Treated library	Control library
E1: Thyroid hormone like			
E1A	Thyroid hormone receptor (THR)	Contig07735; contig20577: thyroid hormone receptor beta	----
E1B	Retinoic acid receptor (RAR)	----	----
E1C	Peroxisome proliferator-activated receptor (PPAR)	----	----
E1D	Rev-Erb	contig06997; contig10060; contig24060; contig15922; contig00699; contig15916: nuclear receptor subfamily 1 group D	contig08171; contig06213; contig09699; contig10567: : nuclear receptor subfamily 1 group D
E1F	RAR-related orphan receptor	contig05750; contig19850: RAR-related orphan receptor gamma	----
E1H	Liver X receptor like receptor	----	----
E1I	Vitamin D3 like receptor	contig12186; pregnane x receptor	contig09088: nuclear receptor subfamily 1 group I
E2: Hepatocyte nuclear factor 4 like			
E2A	Hepatocyte nuclear factor 4 receptor	----	----
E2B	Retinoid X receptor (RXR)	contig14380: nuclear receptor subfamily 2 group B member 4	contig07122: nuclear receptor subfamilly 2 group B member 4, ecdysone receptor
E2C	Testicular receptor	----	----
E2E	Tailless like receptor	contig18846: nuclear receptor subfamily 2 group E member 3	contig11297: nuclear receptor subfamily 2 group E member 3 , IPB000003 Retinoic acid signature, IP001723 Steroid hormone receptor signature
E2F	COUP-TF like receptor	contig20508; contif04321: COUP transcription factor; contig09822:nuclear receptor subfamily 2 group F member 6	contig12879: COUP transcription factor 1
E3: Estrogen like			
E3A	Estrogen receptor	contig21499: estrogen receptor alpha; contig06713: estrogen receptor beta	contig02346: estrogen receptor alpha
E3B	Estrogen-related receptor	----	contig00355: estrogen-relatedreceptor beta
E3C	3-Ketosteroid receptor	----	----
E6: Germ cell nuclear factor			
E6A	Germ cell nuclear factor receptor, EGF receptor?	----	contig06630; contig06823; contig12548: nuclear receptor subfamilly 6 group A

Expression profiles of 94 gene probes corresponding to functions regulated by PPARs (Table 5, Fig. 7) were analysed. These probes have been defined by projection of the transcriptome sequences to KEGG via KAAS and extraction of all probes hitting the

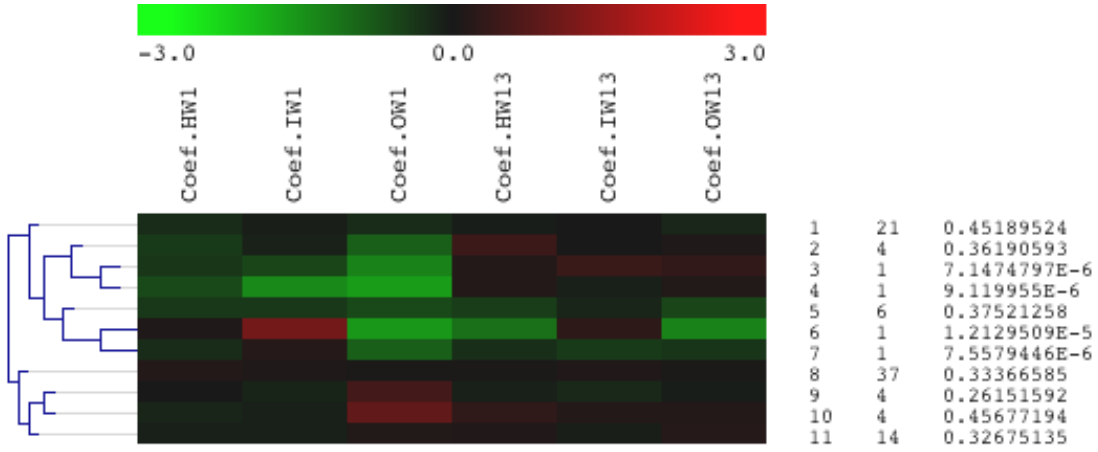
'PPAR pathway' map. Groups of clones annotated as stearoyl-CoA desaturase and long-chain acul-CoA synthase and also single clones with homology to apolipoprotein ApoA5, carnitine palmitoyltransferase and acyl-CoA dehydrogenase were differentially activated in the ovary (cluster11, Fig. 7). Two clones with homology to fatty acid binding protein (cluster5, Fig. 7) were down-regulated in ovarian samples.

Table 5. PPAR signalling pathway (PATH: ko03320) mapped by the Nucella lapillus dataset. **: present in the long list of differentially expressed genes

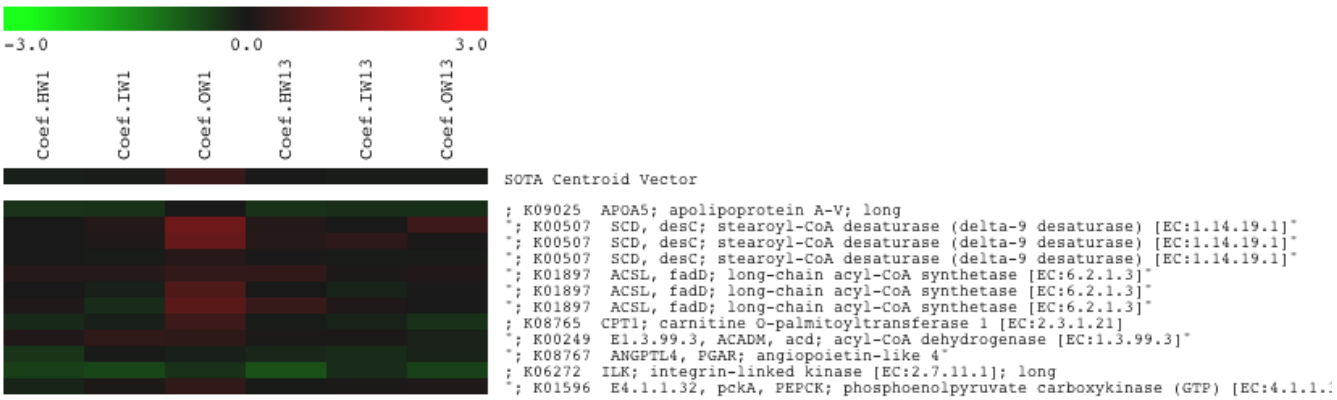
** contig00212	K06259 CD36; CD36 antigen
FW6YJSM01C4KW8	K06259 CD36 CD36 antigen
contig20040	K08745 SLC27A1_4, FATP1, FATP4 solute carrier family 27 (fatty acid transporter), member 1/4
FW6YJSM01CL0K6	K08745 SLC27A1_4, FATP1, FATP4 solute carrier family 27 (fatty acid transporter), member 1/4
FW6YJSM01CJCJW	K08745 SLC27A1_4, FATP1, FATP4 solute carrier family 27 (fatty acid transporter), member 1/4
FW6YJSM02SKL6R	K08745 SLC27A1_4, FATP1, FATP4 solute carrier family 27 (fatty acid transporter), member 1/4
FW6YJSM01BPS8Y	K08748 SLC27A5, FATP5 solute carrier family 27 (fatty acid transporter), member 5
FW6YJSM02T1E69	K08748 SLC27A5, FATP5 solute carrier family 27 (fatty acid transporter), member 5
FW6YJSM01B9O5B	K08749 SLC27A6, FATP6 solute carrier family 27 (fatty acid transporter), member 6
contig14705	K08752 FABP3 fatty acid-binding protein 3, muscle and heart
contig03127	K08755 FABP6 fatty acid-binding protein 6, ileal (gastrotropin)
FW6YJSM02QVNL3	K08525 NR2B2, RXRB retinoid X receptor beta
FW6YJSM02R4H82	K08526 NR2B3, RXRG retinoid X receptor gamma
FW6YJSM02PKL6F	K08530 NR1C3, PPARG peroxisome proliferator-activated receptor gamma
contig09156	K08757 APOA1 apolipoprotein A-I
** FW6YJSM02Q6EKJ	K09025 APOA5 apolipoprotein A-V
contig10097	K08761 PLTP phospholipid transfer protein
contig00976	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
contig01147	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
contig03911	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
contig05657	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
contig15884	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM02Q0IY7	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01ERHJN	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01D8GFQ	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01DR7IO	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01AINPJ	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
** FW6YJSM01BBQ8G	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01AMO4J	K00507 SCD, desC stearoyl-CoA desaturase (delta-9 desaturase)
FW6YJSM01EMLKR	K00489 CYP7A1 cytochrome P450, family 7, subfamily A (cholesterol 7alpha-monooxygenase)
FW6YJSM01BM187	K00489 CYP7A1 cytochrome P450, family 7, subfamily A (cholesterol 7alpha-monooxygenase)
contig12297	K08762 DBI, ACBP diazepam-binding inhibitor (GABA receptor modulator, acyl-CoA-bp)
contig05784	K01897 ACSL, fadD long-chain acyl-CoA synthetase

contig13882	K01897	ACSL, fadD long-chain acyl-CoA synthetase
contig14168	K01897	ACSL, fadD long-chain acyl-CoA synthetase
contig15311	K01897	ACSL, fadD long-chain acyl-CoA synthetase
contig19836	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM02P37QD	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM01EX8U6	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM01EYFAD	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM01A7FA1	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM01A3J21	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM02QM542	K01897	ACSL, fadD long-chain acyl-CoA synthetase
** FW6YJSM02TGVQX	K01897	ACSL, fadD long-chain acyl-CoA synthetase
** FW6YJSM01BDCEL	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM01BVXO1	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM02Q156M	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM02RS7IG	K01897	ACSL, fadD long-chain acyl-CoA synthetase
FW6YJSM02S6BQA	K08763	OLR1 oxidised low-density lipoprotein receptor 1
FW6YJSM01BLA0S	K08763	OLR1 oxidised low-density lipoprotein receptor 1
contig03040	K07425	CYP4A cytochrome P450, family 4, subfamily A
FW6YJSM02SC2WR	K07425	CYP4A cytochrome P450, family 4, subfamily A
FW6YJSM01C9497	K07425	CYP4A cytochrome P450, family 4, subfamily A
FW6YJSM02R45JY	K07513	ACAA1 acetyl-CoA acyltransferase 1
contig08549	K08764	SCP2, SCPX sterol carrier protein 2
** contig19554	K08764	SCP2, SCPX sterol carrier protein 2
FW6YJSM01EODDJ	K08764	SCP2, SCPX sterol carrier protein 2
** contig19232	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM01BXZA6	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM02PY07Y	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM01EKRD0	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM01B2WJH	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM01DV3E9	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM02QWUH1	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
FW6YJSM01C50GL	K00232	E1.3.3.6, ACOX1, ACOX3 acyl-CoA oxidase
contig21676	K08765	CPT1 carnitine O-palmitoyltransferase 1
FW6YJSM02Q6W23	K08765	CPT1 carnitine O-palmitoyltransferase 1
FW6YJSM01COYDQ	K08765	CPT1 carnitine O-palmitoyltransferase 1
FW6YJSM02QBTDM	K08765	CPT1 carnitine O-palmitoyltransferase 1
FW6YJSM01CII96	K08765	CPT1 carnitine O-palmitoyltransferase 1
FW6YJSM01CRMC1	K00249	E1.3.99.3, ACADM, acd acyl-CoA dehydrogenase
FW6YJSM01CKLR9	K00249	E1.3.99.3, ACADM, acd acyl-CoA dehydrogenase
FW6YJSM01EPRAV	K00249	E1.3.99.3, ACADM, acd acyl-CoA dehydrogenase
FW6YJSM02RZC3J	K08767	ANGPTL4, PGAR angiopoietin-like 4
FW6YJSM02SD1CF	K08767	ANGPTL4, PGAR angiopoietin-like 4
contig12307	K06086	SORBS1, SH3D5, PONSIN, CAP sorbin and SH3 domain containing 1
FW6YJSM01CXOGL	K07296	ACDC adiponectin

FW6YJSM02R28ZD	K01388	MMP1 matrix metalloproteinase-1 (interstitial collagenase)
FW6YJSM02QPU9G	K06272	ILK integrin-linked kinase
FW6YJSM01A16DD	K06272	ILK integrin-linked kinase
FW6YJSM01DH9FV	K06272	ILK integrin-linked kinase
FW6YJSM02R06LW	K06272	ILK integrin-linked kinase
** FW6YJSM02PK4WN	K06272	ILK integrin-linked kinase
FW6YJSM01BLGAO	K06276	PDPK1 3-phosphoinositide dependent protein kinase-1
FW6YJSM01BTK5L	K08770	UBC ubiquitin C
contig26458	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
FW6YJSM02P0IKX	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
FW6YJSM01DH0G3	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
FW6YJSM02TPRBI	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
** FW6YJSM02PMPKN	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
FW6YJSM02Q6CX5	K01596	E4.1.1.32, pckA, PEPCK phosphoenolpyruvate carboxykinase (GTP)
FW6YJSM02QTFKY	K00864	E2.7.1.30, glpK glycerol kinase
FW6YJSM01AKOGS	K00864	E2.7.1.30, glpK glycerol kinase
** contig10481	K08771	AQP7 aquaporin-7



Cluster11 illustrates a band of PPAR-regulated functions differentially activated in OW1:

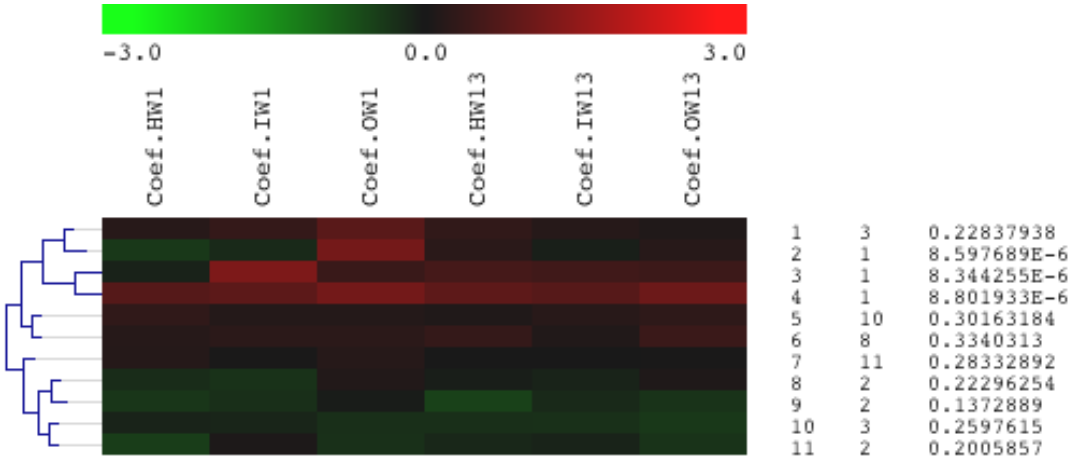


Cluster5 illustrates a band of PPAR-regulated functions differentially suppressed in OW1:

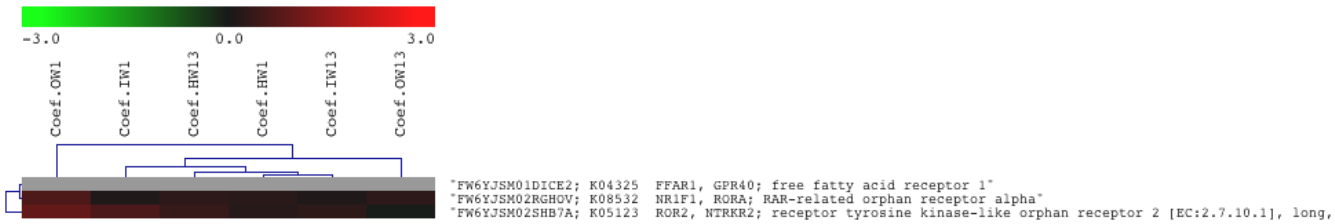
Figure 7. Expression profile analysis for the PPAR network mapped by the Nucella lapillus sequences. The first right column is an N of the cluster followed by the number of functions in each cluster. Green: down-regulated, red: up-regulated. Euclidean distance (cell diversity p<0.05, min distanceMax cell div 0.01. min b error e-4). (Please see appendix 3.3 for more individual cluster detail).

Additionally, expression profiles of the detected lipophilic receptors (Fig. 8) were also analysed. It is clear that the expression of ROR, that is linked to activation of the SRY pathway, increases with time (cluster1, Fig. 8). Additionally, the most consistently expressed TF was ARNTL2 (cluster4, Fig. 8) that has PPARγ binding sites and SOX9 binding sites in the promoter region (at least in vertebrates). Two other highly expressed TFs were ROR2 that have also PPARγ binding sites and opioid receptor kappa 1 (OPRK1) that has the AhR/ARNT binding site in vertebrate gene promoter regions. A vertebrate homolog of estrogen-related receptor alpha (ESRRA), that was induced in both the ovaries and hepatopancreas in week 13 of the experiment, also has SOX9 and AhR binding sites in its promoter region.

The ovary after 1 week of exposure (OW1) responded differently from the other samples and was the main organ where PPAR-related functions were affected. We predict that PPARγ and AhR are likely to be the first targets of TBT and ROR2, SOX9, OPRK1 are induced during downstream processes. Expression of OPRK1 may lead to further inductions of differentiation by one of the brain hormones.



Cluster1: RORA, ROR2, GPR40 up-regulated in OW1,IW1:



Cluster4: ARNTL2: up-regulated in all experimental conditions:

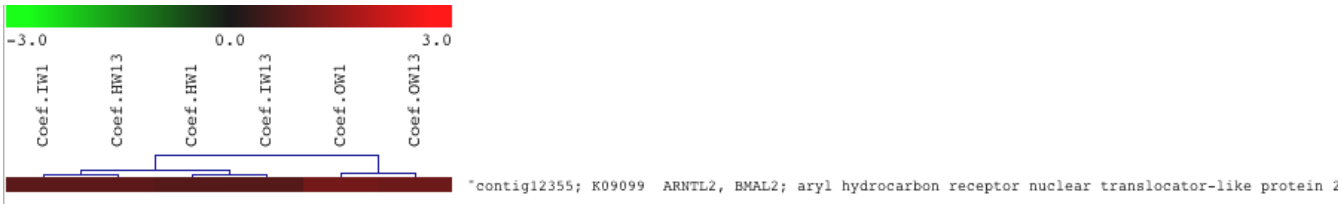


Figure 8. Expression profile analysis for the Nucella lapillus receptors; General for clustering of functions; Euclidean SOMs, (For hierarchies: single-linkage clustering). Green: down-regulated; red: up-regulated. (Please see appendix 3.3 for more individual cluster detail).

3.4.3.3. Phylogenetic conservation of gender-related TFs

The sequence conservation (sequence alignment was performed by T-Coffee [version 5.3]) of gender-related TFs throughout a range of vertebrates and invertebrates (please see appendix 3.4) was investigated and showed that putative SRY (sex-determining region Y) group homologs are highly conserved. Within the SRY group, SOX9 exhibited the highest levels of sequence conservation throughout Rat, Ciona, Tetraodon, Xenopus and Drosophila, indicative of a common function. The sequences with the

strongest homology to the human SOX9 gene (NCBI: NM_000346) were derived from the SEED database and are listed under SEED identities. Within SOX9, there is a conservative core, which is likely to be the DNA binding region (Fig. 9).

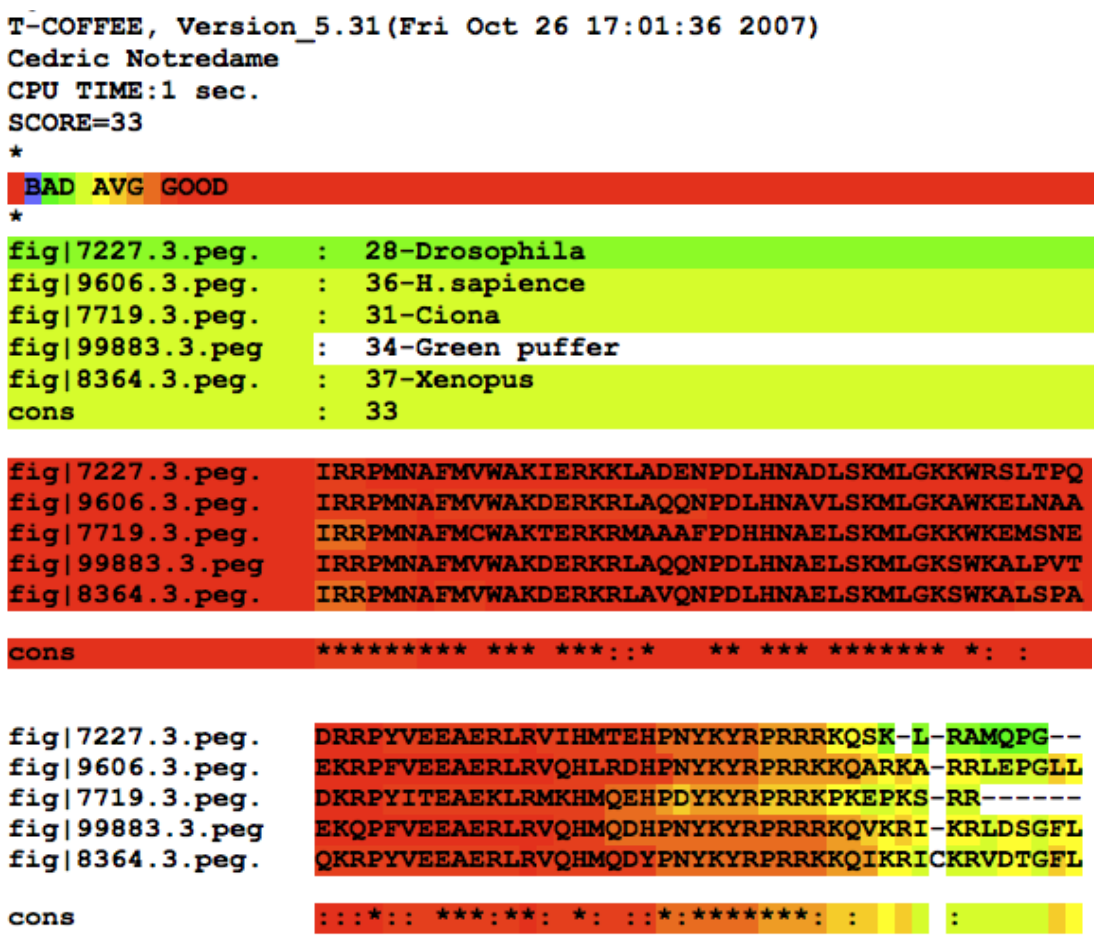


Figure 9. Sequence alignment of SOX9 throughout Rat, Ciona, Tetraodon, Xenopus and Drosophila with the evidence of a conservative core highlighted in red.

Moreover, there is pronounced conservation throughout promoter region motifs (Fig. 10) and in particular, the SRY-specific binding motif (SOX9 motif 1), highlighted conservation of structure and regulatory feature for a gender-regulation-related gene in both vertebrates and invertebrates.

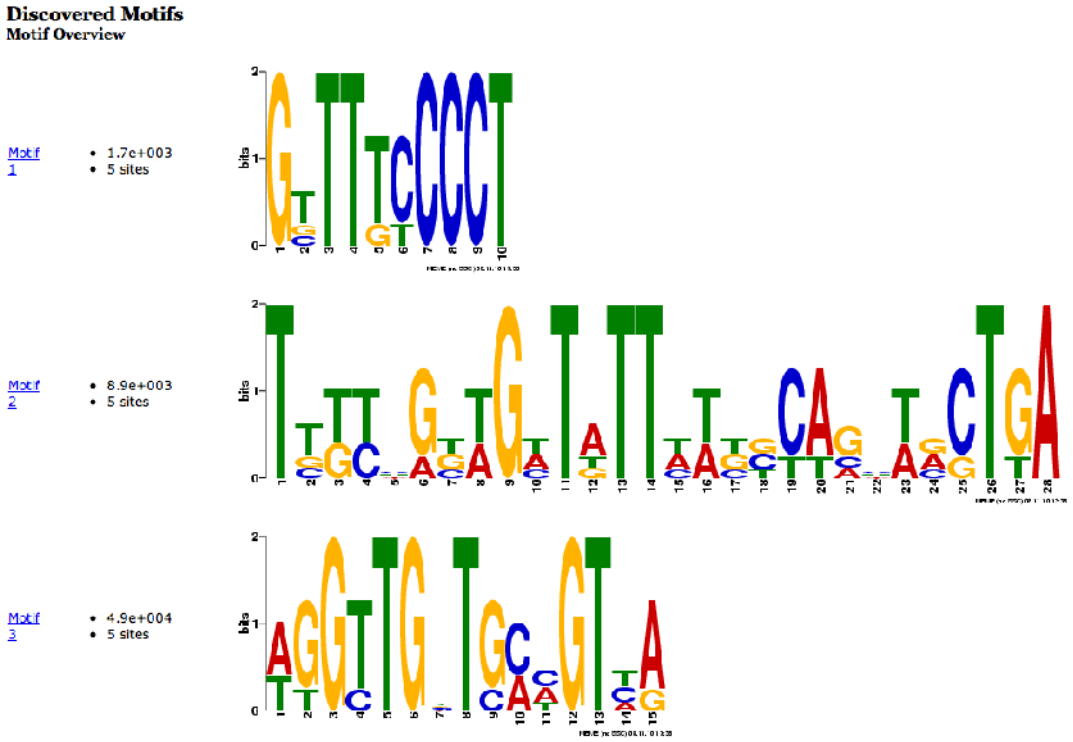


Figure 10. Example of conservation in the promoter region sequences of gender-related transcription factors binding sites (SOX9); motif 1, 2 and 3 were the most conserved.

This example demonstrates that there is conservation of a structure and regulatory feature for a gender-regulation-related gene (please see appendix 3.5 for details and more examples).

3.4.3.4. Detoxification, oxidative phosphorylation and transport

Several differentially expressed genes (please see appendix 3.1) were related to transporter activity, and it seems that it is highly affected in the response to TBT. Particularly, carbamoyl-phosphate synthase enzyme, associated with the ammonia/NO metabolism, was strongly inhibited in all the tissues, especially after 1 week of exposure. There were also pronounced up-regulation of genes linked to oxidative phosphorylation (adenosine triphosphate [ATP] production, metabolic and transporting ATPases) and sulfur-iron/thiol metabolism and oxidoreductase activity. One of the top up-regulated functions in ovary and hepatopancreas was also mitochondrial activity and it can be

functionally linked to a number of functions for fatty acid metabolism differentially up-regulated in ovaries and may be controlled by PPARs.

3.4.3.5. Other potential TBT targets

Impacts in the immune system, cell proliferation and apoptosis, DNA repair, tumour suppressors among others were also evident from the DE gene list (please see appendix 3.1) and these may also be potential targets of endocrine disruption; however, those were not explored in depth since our main objective was focused on the impacts of TBT in the reproductive/gender determination process.

3.4.4. Microarray validation by qPCR

For each experimental condition, the dissociation curves revealed single amplicon amplification. The qPCR efficiencies ranged from 100% to 109% and their correlation coefficients ranged from 0.988 to 0.998. For all the tested genes in the different tissues, the direction of regulation obtained by qPCR was the same as revealed by the microarray data (Fig. 11). Overall, qPCR and microarray results were similar but the transcript fold change in response to treatment was slightly reduced, though it was not significantly different (paired samples t-test p=0.085).

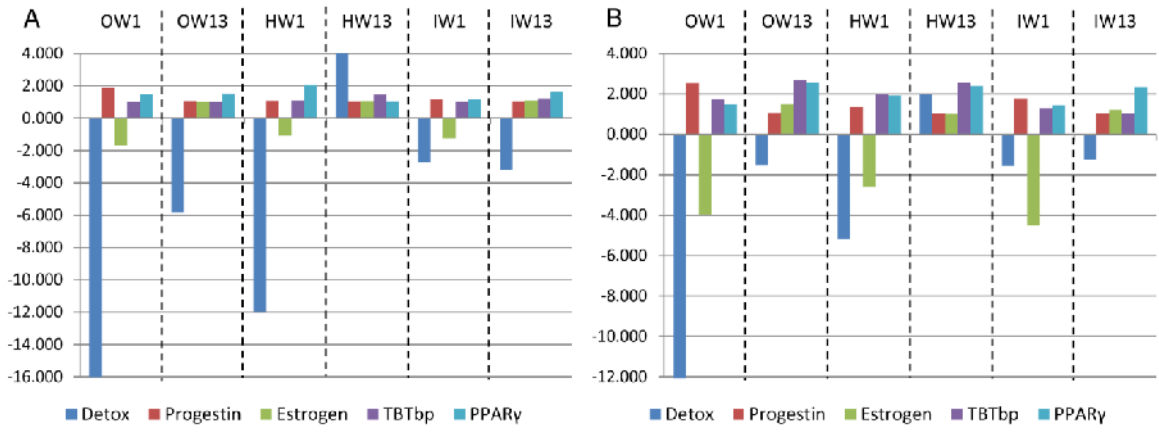


Figure 11. Microarray validation by qPCR; representation of fold change per gene: A) qPCR results; B) microarray results. OW1: Ovary week1; OW13: Ovary week 13; HWI: Hepatopancreas w1, HW13: Hepatopancreas w13; IW1: Penis w1; IW13: Penis W13.

3.5. DISCUSSION

The imposex mechanism is one of the key examples of anthropogenically-induced endocrine disruption in aquatic ecosystems and not surprisingly it has received special attention in recent years. Important insights have been achieved and various theories on causal mechanisms have been proposed: steroid (Spooner et al. 1991; Bettin et al. 1996; Gooding et al. 2003), neuroendocrine (Féral et al. 1983; Oberdorster and McClellan-Green 2000b, 2002) and retinoid (Nishikawa et al. 2004; Castro et al. 2007a; Horiguchi et al. 2007). However, the underlying molecular mechanism of imposex is still poorly understood and further investigation is required. Taking advantage of the new advanced technologies, here we applied combinations of pyrosequencing and microarrays in order to explore the genomic profile of response to TBT exposure in the dogwhelk, N.lapillus. Indeed, the applied technology and the produced genomic tools – a specific Nucella lapillus oligoarray – successfully revealed several new insights to the imposex mechanism and response to TBT in general. Moreover, the produced array provides an unparallel genomic resource for N. lapillus and we anticipate that it will be a valuable resource for exploring additional responses of dogwhelks to environmental changes/variation.

Regarding the lack of gene annotation of the Nucella transcriptome and consequent unknown sequences sense, all selected sequences that have been used in the array design were used for a sense and antisense probe production – acknowledging the capacity of probe representation in the selected array. By doing this, it was ensured that all the sequences had equal possibility to hybridize with the study samples and, generally, just one of the probe senses had response in gene expression (we found one sequence in the shortlisted sequences that showed differential gene expression in both senses but the pattern of gene expression was the same for both directions). The generation of such transcriptomic and sequence data in a non-model organism provides an informative framework for undertaking similar investigations of genomic response to environmental stress. Furthermore, we validated gene expression data obtained by microarray using qPCR and the same patterns of gene expression were obtained confirming reproducibility between techniques and the accuracy of our findings. Also noteworthy is the observed time and tissue-specific response to contamination, reflecting the need to analyse

different tissues and time scales in order to capture a representative profile pattern and dynamics of gene expression.

The data show that the response to TBT contamination is a complex phenomenon that involves several molecular pathways and their crosstalk. As expected, large suites of genes involved in endocrine disruption/gender determination, the primary focus of our analyses, were differentially espressed in the present dataset. However, genes involved with the immune system response, cell proliferation and apoptosis, DNA repair and tumour suppressors, among others, were also differentially expressed and are therefore likely to be involved in the response to contamination. Interactions between the immune and endocrine systems are well documented and it has been suggested that the immune system may be susceptible to endocrine disruption (Swedenborg et al. 2009). Associations of endocrine disruptors with functions other than reproduction, such as the immune response, may also help identifying specific targets for endocrine disruption in invertebrates (Porte et al. 2006). From our data, TBT seems to be a very multi-site binding compound and so, it is likely that several different causal pathways are involved in the toxic response.

This study is one of the most extensive and pioneering using a non-model organism with almost no previous molecular data in order to elucidate a well-known gene-environment question. Moreover, we predict that many further questions and studies may arise from our genomic resources, tools and findings.

3.5.1. Lipophilic Nuclear receptors as transducers of TBT toxicity

The oligoarray derived gene expression analyses in response to TBT exposure, support the involvement of all three previously suggested causal mechanisms (steroid (Spooner et al. 1991; Bettin et al. 1996; Gooding et al. 2003), neuroendocrine (Féral et al. 1983; Oberdorster and McClellan-Green 2000b, 2002) and retinoid (Nishikawa et al. 2004; Castro et al. 2007a; Horiguchi et al. 2007)) that may crosstalk at the transcription level. Biosynthesis of steroid hormones seems to be affected in all the tissues since CypA71,

CypB71 and CypA391 that were found to be differentially expressed in the samples are all related to the pathway. These cytochromes can be regulated by gender-determination genes, by lipophilic nuclear receptors and also by SREBP1, the PPARγ and the progestin receptor which were all found to be expressed in the candidate sequences obtained.

From the produced transcription profile and the following bioinformatic analysis it was possible to identify nuclear receptors as new candidate TBT binders in N. lapillus that may be involved in the toxic response chain. We anticipate that they may be strong candidate targets for the endocrine disruption mechanism as several nuclear receptors have recently been identified as mediators of endocrine disruption in vertebrates (Iguchi and Katsu 2008).

Retinoid signalling that has been shown as an important contributor of TBT toxicity in Nucella is transduced by several families of nuclear receptors (RXR, RARs, PPARs), receptor tyrosine kinases (RORs) and G-protein coupled receptors. The most general, RXR, affects multiple regulatory pathways because it dimerizes with all two-subunit nuclear receptors. As shown above, expression of several potential receptors or retinoids in Nucella were affected by TBT exposure. RARα and ROR1 were up-regulated in all tissues and identified in both the transcriptome and in the treated assembled library transcription factor searches. Additionally, a PPARγ homolog gene was also identified within the candidate sequences and was differentially expressed in all tissues showing general up-regulation. To our knowledge PPARγ has not been described in gastropods, but a PPARγ like protein (NCBI: NW_001955054) has been identified in the sea squirt, Ciona intestinalis, genome and in other invertebrates such as marine molluscs (Perrigault et al. 2009). PPARγ has been shown to bind TBT in other organisms and can be regulated by gender-determining TFs and other lipophilic nuclear receptors (Hiromori et al. 2009). Moreover, ovarian PPAR-related functions are regulated by PPARγ and SREBP in vertebrates (Memon et al. 2000). Ovary tissue may contain fat-storing cells, homologous to fat tissue of higher animals and that may be regulated by similar mechanims, including

highly fat-related PPARγ. Consequently, we predict that the RARα, ROR1 and PPARγ nuclear receptors and TFs are likely to be involved in TBT binding in N. lapillus.

Down-regulation and up-regulation of some genes located downstream of the PPAR pathway were observed, potentially related to the presence of additional factors in regulation. Lipophilic receptors crosstalk by competition for RXR, coactivators and the DNA binding sites (Swedenborg et al. 2009). SREBP and ARNT genes that have binding sites in promoters of PPAR-regulated vertebrate genes were among genes activated by TBT exposure and may modulate a potential effect of PPAR-like TBT-binders. If the same is true for Nucella's genes, we would suggest that ARNT may play a role in down-regulation of these genes. Thus, the resulting mechanism may be driven by a balance of activating role of PPARγ-SREBP complex and inhibiting function of ligated ARNT (Alexander et al. 1998). The Rev-Erba (orphan nuclear receptor) TF also identified in the datasets is known to negatively regulate RARα binding and may affect transcription of genes linked to gender development. The REv-Erba ligand is not known, but our data suggest that it might be inhibited by TBT. Overall, we can suggest that PPARγ and AhR are the first targets of TBT and ROR2, SOX9, OPRK1 are induced later. The expression of OPRK1 may lead to further induction of differentiation by one of the brain hormones.

A receptor for ecdysterol was also found among the differentially expressed functions obtained by KEGG BLAST, and, interestingly, the cytochrome that is involved in ecdysteroid biosynthesis - Cyp306A1 (Niwa et al. 2004) - was also up-regulated in ovary. The ecdysterol hormone is responsible for metamorphosis in insects and it is hypothesised to be functional in molluscs as well. This hormone was found to be associated with ovarian differentiation in hermaphroditic species (Nolte et al. 1986). We hypothesise that the ecdysterol receptor is one of the primary targets of TBT binding in Nucella.

3.5.2. Potential effect of TBT exposure on regulators of gender differentiation

After one week of TBT exposure, expression (up-regulation) of putative TF homologs that may be involved in gonad re-differentiation was detected. Particularly, androgen-receptor-like gene and also putative GATA and SRY-box (SOX) TFs were found up-regulated in ovary and hepatopancreas that may be involved in early stages of male related differentiation at the level of steroid production and gonad differentiation. Analysis of promoter regions of gender-related TFs were explored revealing that the binding sites for SOX9 were quite conserved in the promoter sequences. Previous studies (Ren et al. 2001; Nishida et al. 2003) on phylogenetics of sex determinant genes also confirm a certain degree of conservation of sex determining pathways, thus providing a rationale for the exploration of mechanisms of gender regulation in Nucella, extrapolated from previously studied organisms.

3.5.3. Mitochondrial toxicity of TBT

Mitochondrial functions were among the most highly ranked up-regulated functions in the ovary and hepatopancreas. Activation of transcription of respiratory functions may be due to the activation of mitochondrial sterol biosynthetic steps, but might be related to TBT effect on the plasma and mitochondrial membrane via inhibition of mitochondrial channels or its uncoupling features (Liu et al. 1996). Liu et al. (1996) has shown that TBT can bind to thiol groups, as its toxicity can be reversed by S2- or di-thiols. TBT can also catalyze (Cl-/OH-) exchange via membrane (similar to chloride channel), and it can also inhibit ATPases (Powers and Beavis 1991). A number of carboxylate transporters were in the list of differentially regulated functions indicative of a deficiency in corresponding compounds. The possible inhibition of transport of carboxylates and/or hormones (Powers and Beavis 1991) by TBT may be crucial for gonad differentiation and can cause a chain of toxic events. Some simple compounds like lactate, pyruvate or succinate may be important to keep gonads from re-differentiation. Deregulation of energetics due to a block in anionic exchange, in general, can affect sterol biosynthesis,

PPARγ expression, change in retinoic balance affect cytochrome function, intercellular pH and ammonia metabolism (Suzuki et al. 2003; Yamada et al. 2008).

Overall, gene expression data revealed several new insights to the endocrine disruption mechanism (N. lapillus response to TBT), which were not possible to identify to date, mainly due to technological limitation. Overall, and based on our results we can suggest:

i) Steroid, neuroendocrine and retinoid pathways interact and contribute for the imposex mechanism;

ii) TBT affects byosynthesis of steroid hormones, and detoxification related genes were mainly down-regulated;

iii) Some TFs and receptors e.g. RAR, ROR, Rev-Erba, ecdysterol are good candidate TBT targets involved in the mechanism;

iv) A PPARγ homolog and AhR are likely to be the first targets of TBT and ROR2, SOX9, OPRK1 are later induced leading to further induction of differentiation by one of the brain hormones;

v) Transport function is affected and there is some evidence for deregulation of energetics;

vi) Functions other than reproduction (e.g. immune responses) also constitute potential targets of endocrine disruption;

vii) There are several examples and strong evidence for common signalling mechanisms for endocrine disruption in invertebrates and vertebrate species;

viii) The findings have provided a better understanding of endocrine disruption in invertebrates and revealed that N. lapillus may be a good candidate “transcriptome-enabled” ecological model organism for endocrine disruption studies.

3.6. CONCLUSION

Combinations of pyrosequencing and microarray technologies (design of an oligonucleotide array (Agilent 4*180K) from the partial Nucella transcriptome) were employed in order to help to disentangle the functional genomic mechanism of imposex, using N. lapillus as the model organism. Microarray analyses for differential gene expression in response to TBT contamination were performed and contributed significantly to our understanding of the molecular basis of the imposex mechanism. Overall, the presence and crosstalk in gene differential expression of the 3 previously suggested hypotheses were identified and complemented. Moreover, new candidate TBT targets and related pathways involved in the toxic chain were discovered in the Nucella response to TBT exposure. Regarding some conservation in the transcription factors binding sites and pathways, we anticipate that the produced results may have a broader application among taxa and related pollutants. Moreover, a common mechanism of signalling of endocrine disruption was evident between gastropods and vertebrates suggesting that the underlying molecular mechanism between taxa is probably more conserved than expected.

Finally, we anticipate that the transcriptomic data obtained herein, together with the functional genomic tools will provide a valuable resource to further explore endocrine disruption and additional environmental and evolutionary responses of this sentinel organism to challenges of the intertidal environment that hitherto has not been possible.

3.7. REFERENCES

Alexander, D., L. Ganem, P. Fernandez-Salguero, F. Gonzalez, and C. Jefcoate. 1998. Aryl-hydrocarbon receptor is an inhibitory regulator of lipid synthesis and of commitment to adipogenesis Journal of Cell Science 111:3311-3322.

Bettin, C., J. Oehlmann, and E. Stroben. 1996. TBT-induced imposex in marine neogastropods is mediated by an increasing androgen level. Helgolander Meeresuntersuchungen 50:299-317.

Bryan, G. W., P. E. Gibbs, L. G. Hummerstone, and G. R. Burt. 1986. The decline of the gastropod Nucella lapillus around Southwest England - evidence for the effect of tributyltin from antifouling paints. Journal of the Marine Biological Association of the United Kingdom 66:611-640.

Castro, L. F. C., D. Lima, A. Machado, C. Melo, Y. Hiromori, J. Nishikawa, T. Nakanishi, M. A. Reis-Henriques, and M. M. Santos. 2007a. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod Nucella lapillus. Aquatic Toxicology 85:57-66.

Castro, L. F. C., C. Melo, R. Guillot, I. Mendes, S. Queiros, D. Lima, M. A. Reis-Henriques, and M. M. Santos. 2007b. The estrogen receptor of the gastropod Nucella lapillus: Modulation following exposure to an estrogenic effluent? Aquatic Toxicology 84:465-468.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17:1629-1631.

Féral, C., C. le Gall, J. Lever, and H. H. Boer. 1983. The influence of a pollutant factor (tributylin) on the neuroendocrine mechanism responsible for the occurrence of a penis in the females of Ocenebra erinacea. Molluscan neuro-endocrinology. Proceedings of the International Minisymposium on Molluscan Endocrinology, held in the Department of Biology, Free University, Amsterdam, the Netherlands, August 16-20, 1982.:173-175.

Gibbs, P. E., G. W. Bryan, and P. L. Pascoe. 1991. Tbt-Induced Imposex in the Dogwhelk, Nucella-Lapillus - Geographical Uniformity of the Response and Effects. Marine Environmental Research 32:79-87.

Gibbs, P. E., G. W. Bryan, P. L. Pascoe, and G. R. Burt. 1987. The use of the dog-whelk, Nucella lapillus, as an indicator of tributyltin (TBT) contamination. Journal of the Marine Biological Association of the United Kingdom 67:507-523.

Gooding, M. P., V. S. Wilson, L. C. Folmar, D. T. Marcovich, and G. A. LeBlanc. 2003. The biocide tributyltin reduces the accumulation of testosterone as fatty acid esters in the mud snail (Ilyanassa obsoleta). Environmental Health Perspectives 111:426-430.

Gracey, A. Y. 2007. Interpreting physiological responses to environmental change through gene expression profiling. Journal of Experimental Biology 210:1584-1592.

Gracey, A. Y., and A. R. Cossins. 2003. Application of microarray technology in environmental and comparative physiology. Annual Review of Physiology 65:231-259.

Guillette, L. J., T. S. Gross, G. R. Masson, J. M. Matter, H. F. Percival, and A. R. Woodward. 1994. Developmental abnormalities of the gonad and abnormal sex-hormone concentrations in juvenile alligators from contaminated and control lakes in Florida. Environmental Health Perspectives 102:680-688.

Hiromori, Y., J.-I. Nishikawa, I. Yoshida, H. Nagase, and T. Nakanishi. 2009. Structure-dependent activation of peroxisome proliferator-activated receptor (PPAR) γ by organotin compounds. Chemico-Biological Interactions 180:238-244.

Horiguchi, T. 2006. Masculinization of female gastropod mollusks induced by organotin compounds, focusing on mechanism of actions of tributyltin and triphenyltin for development of imposex. Environmental Sciences 13:77-87.

Horiguchi, T., T. Nishikawa, Y. Ohta, H. Shiraishi, and M. Morita. 2007. Retinoid X receptor gene expression and protein content in tissues of the rock shell Thais clavigera. Aquatic Toxicology 84:379-388.

Huet, M., Y. M. Paulet, and M. Glemarec. 1996. Tributyltin (TBT) pollution in the coastal waters of west Brittany as indicated by imposex in Nucella lapillus. Marine Environmental Research 41:157-167.

Iguchi, T., and Y. Katsu. 2008. Commonality in Signaling of Endocrine Disruption from Snail to Human. Bioscience 58:1061-1067.

Inoue, S., K. Yoshinari, M. Sugawara, and Y. Yamazoe. 2011. Activated Sterol Regulatory Element-Binding Protein-2 Suppresses Hepatocyte Nuclear Factor-4-Mediated Cyp3a11 Expression in Mouse Liver. Molecular Pharmacology 79:148-156.

Ishikawa, H. 1977. Evolution of Ribosomal RNA. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 58:1-7.

Janer, G., R. M. Sternberg, G. A. LeBlanc, and C. Porte. 2005. Testosterone conjugating activities in invertebrates: are they targets for endocrine disruptors? Aquatic Toxicology 71:273-282.

Ketata, I., X. Denier, A. Hamza-Chaffai, and C. Minier. 2008. Endocrine-related reproductive effects in molluscs. Comparative Biochemistry and Physiology C-Toxicology and Pharmacology 147:261-270.

Liu, G. Y., B. Hinch, and A. D. Beavis. 1996. Mechanisms for the transport of alpha, omega-dicarboxylates through the mitochondrial inner membrane. Journal of Biological Chemistry 271:25338-25344.

Livak, K., and T. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Matthiessen, P., and P. E. Gibbs. 1998. Critical appraisal of the evidence for tributyltin-mediated endocrine disruption in mollusks. Environmental Toxicology and Chemistry 17:37-43.

Memon, R., L. Tecott, K. Nonogaki, A. Beigneux, A. Moser, C. Grunfeld, and K. Feingold. 2000. Up-Regulation of Peroxisome Proliferator-Activated Receptors (PPAR-a) and PPAR-g Messenger Ribonucleic Acid Expression in the Liver in Murine Obesity: Troglitazone Induces Expression of PPAR-g-Responsive Adipose Tissue-Specific Genes in the Liver of Obese Diabetic Mice. Endocrinology 141.

Migeon, C., and A. Wisniewski. 2000. Human sex differentiation: from transcription factors to gender. Hormone Research 53:111-119.

Morcillo, Y., and C. Porte. 1999. Evidence of endocrine disruption in the imposer-affected gastropod Bolinus brandaris. Environmental Research 81:349-354.

Murata, S., S. Takahashi, T. Agusa, N. J. Thomas, K. Kannan, and S. Tanabe. 2008. Contamination status and accumulation profiles of organotins in sea otters (Enhydra lutris) found dead along the coasts of California, Washington, Alaska (USA), and Kamchatka (Russia). Marine Pollution Bulletin 56:641-649.

Nishida, S., L. A. Pastene, M. Goto, and H. Koike. 2003. SRY gene structure and phylogeny in the cetacean species. Mammal Study 28:57-66.

Nishikawa, J. 2006. Imposex in marine gastropods may be caused by binding of organotins to retinoid X receptor. Marine Biology 149:117-124.

Nishikawa, J., S. Mamiya, T. Kanayama, T. Nishikawa, F. Shiraishi, and T. Horiguchi. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Environonmental Science and Technology 38:6271-6276.

Niwa, R., T. Matsuda, T. Yoshiyama, T. Namiki, K. Mita, Y. Fujimoto, and H. Kataoka. 2004. CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and Drosophila. Journal of Biological Chemistry 279:35942-35949.

Nolte, A., J. Koolman, M. Dorlochter, and H. Straub. 1986. Ecdysteroids in the dorsal bodies of pulmonates (gastropoda) - synthesis and release of ecdysone. Comparative Biochemistry and Physiology a-Physiology 84:777-782.

Oba, Y., Y. Shimasaki, Y. Shima, H. Satone, T. Kitano, M. Nakao, S. Kawabata, and T. Honjo. 2007. Purification and characterization of tributyltin-binding protein type 2 from plasma of Japanese flounder, Paralichthys olivaceus. Journal of Biochemistry 142:229-238.

Oberdorster, E., and P. McClellan-Green. 2000a. The neuropeptide APGWamide induces imposer in the mud snail, Ilyanassa obsoleta. Peptides 21:1323-1330.

Oberdorster, E., and P. McClellan-Green. 2000b. The neuropeptide APGWamide induces imposex in the mud snail, Ilyanassa obsoleta. Peptides 21:1323-1330.

Oberdorster, E., and P. McClellan-Green. 2002. Mechanisms of imposex induction in the mud snail, Ilyanassa obsoleta: TBT as a neurotoxin and aromatase inhibitor. Marine Environmental Research 54:715-718.

Oliveira, I. B., C. A. Richardson, A. C. Sousa, S. Takahashi, S. Tanabe, and C. M. Barroso. 2009. Spatial and temporal evolution of imposex in dogwhelk Nucella lapillus (L.) populations from North Wales, UK. Journal of Environmental Monitoring 11:1462-1468.

Perrigault, M., A. Tanguy, and B. Allam. 2009. Identification and expression of differentially expressed genes in the hard clam, Mercenaria mercenaria, in response to qualog parasite unknown (QPX). Bmc Genomics 10:1471-2164.

Porte, C., G. Janer, L. C. Lorusso, M. Ortiz-Zarragoitia, M. P. Cajaraville, M. C. Fossi, and L. Canesi. 2006. Endocrine disruptors in marine organisms: approaches and perspectives. Comparative Biochemistry and Physiology C-Toxicology and Pharmacology 143:303-315.

Powers, M. F., and A. D. Beavis. 1991. Triorganotins inhibit the mitochondrial inner membrane anion channel. Journal of Biological Chemistry 266:17250-17256.

Ratcliff, D. 1970. Changes attributable to pesticides in egg breakage frequency and eggshell thickness in some British birds. Journal of Applied Ecology 7:67-115.

Ren, L. I., H. H. Cheng, Y. Q. Guo, X. Huang, L. Liu, and R. J. Zhou. 2001. Evolutionary conservation of Dmrt gene family in amphibians, reptiles and birds. Chinese Science Bulletin 46:1992-1995.

Rogue, A., C. Spire, M. Brun, N. Claude, and A. Guillouzo. 2010. Gene expression changes induced by PPAR gamma agonists in animal and human liver. PPAR Research 2010.

Ronis, M. J. J., and A. Z. Mason. 1996. The metabolism of testosterone by the periwinkle (Littorina littorea) in vitro and in vivo: Effects of tributyl tin. Marine Environmental Research 42:161-166.

Santos, M. M., C. C. ten Hallers-Tjabbes, N. Vieira, J. P. Boon, and C. Porte. 2002. Cytochrome P450 differences in normal and imposex-affected female whelk Buccinum undatum from the open North Sea. Marine Environmental Research 54:661-665.

Spade, D. J., R. J. Griffitt, L. Liu, N. J. Brown-Peterson, K. J. Kroll, A. Feswick, R. A. Glazer, D. S. Barber, and N. D. Denslow. 2010. Queen Conch (Strombus gigas) Testis Regresses during the Reproductive Season at Nearshore Sites in the Florida Keys. Plos One 5.

Spooner, N., P. E. Gibbs, G. W. Bryan, and L. J. Goad. 1991. The Effect of Tributyltin Upon Steroid Titers in the Female Dogwhelk, Nucella-Lapillus, and the Development of Imposex. Marine Environmental Research 32:37-49.

Sternberg, R. M., M. P. Gooding, A. K. Hotchkiss, and G. A. LeBlanc. 2010. Environmental-endocrine control of reproductive maturation in gastropods: implications for the mechanism of tributyltin-induced imposex in prosobranchs. Ecotoxicology 19:4-23.

Sternberg, R. M., A. K. Hotchkiss, and G. A. Leblanc. 2008. Synchronized expression of retinoid X receptor mRNA with reproductive tract recrudescence in an imposex-susceptible mollusc. Environmental Science & Technology 42:1345-1351.

Sumpter, J. P., and A. C. Johnson. 2005. Lessons from endocrine disruption and their application to other issues concerning trace organics in the aquatic environment. Environmental Science & Technology 39:4321-4332.

Suzuki, T., T. Onogawa, N. Asano, H. Mizutamari, T. Mikkaichi, M. Tanemoto, M. Abe, F. Satoh, M. Unno, K. Nunoki, M. Suzuki, T. Hishinuma, J. Goto, T. Shimosegawa, S. Matsuno, S. Ito, and T. Abe. 2003. Identification and characterization of novel rat and human gonad-specific organic anion transporters. Molecular Endocrinology 17:1203-1215.

Swedenborg, E., J. Ruegg, S. Makela, and I. Pongratz. 2009. Endocrine disruptive chemicals: mechanisms of action and involvement in metabolic disorders. Journal of Molecular Endocrinology 43:1-10.

Yamada, Y., K. Davis, and C. Coffman. 2008. Programmed cell death of primordial germ cells in Drosophila is regulated by p53 and the Outsiders monocarboxylate transporter. Development 135:207-216.

CHAPTER 4

Induction of imposex in Nucella lapillus by Rosiglitazone: evidence for a potential involvement of the RXR:PPARγ signalling pathway

Chapter submitted as part of an original article:

Pascoal S, Carvalho G, Vasieva O, Hughes R, Cossins A, Fang Y, Ashelford K, Olohan L, Barroso C, Mendo S, Creer S. A Transcriptomic Perspective of Imposex Endocrine Disruption in an Ecological Sentinel (Nucella lapillus) Suggests Involvement of RXR:PPARγ Signalling. PloS Biology

4.1. ABSTRACT

The heterodimer Retinoid X receptor (RXR): Peroxisome proliferator-activated receptor gamma (PPARγ) is well known in vertebrates and is recognized as being involved in endocrine disruption mechanisms. However, there is so far no evidence for its active role in the imposex mechanism but there is evidence for the involvement of the RXR pathway. Based on indications from the transcriptomic study described in this thesis, here we investigated the possible involvement of PPARγ signalling pathway in imposex induction. To test the potential contribution of the PPARγ pathway, previously unexposed Nucella lapillus were injected with Rosiglitazone (Rosi), a known PPARγ ligand, and maintained in the laboratory for 2 months. Imposex was significantly induced in snails injected with Rosi at a degree comparable to that of TBT. We thereby propose that one possible mechanism by which TBT triggers the imposex development in N. lapillus is via the activation of the RXR:PPARγ heterodimer signalling pathway, hitherto not described in invertebrates, leading to the transcription of PPARγ target genes and generating a cascade of events that will ultimately cause the masculinisation of females.

4.2. INTRODUCTION

Substantial progress has already been made regarding our understanding of the individual morphological processes involved in the imposex (superimposition of male characters (penis and vas deferens) onto prosobranch females (Smith 1971)) development and the impacts of this phenomenon at the population and ecosystem levels (Gibbs et al. 1987; Sternberg et al. 2010). However, disentangling the underlying molecular and biochemical mechanisms has been a challenge due in part to the phylogenetic disparity of prosobranchs from other genomic model species from other taxonomic groups. To date, three main mechanisms (steroid, neuroendocrine and retinoid) have been proposed to explain how TBT induces imposex in gastropods but the exclusivity and/or level of interplay between the suggested pathways remain unclear, and many molecular targets have yet to be characterised.

Throughout this thesis, new insights have enabled enhanced understanding of the imposex phenomenon. Application of a flexible suite of contemporary transcriptomic and post-genomic tools (chapter 2 and chapter 3) indicated that all three putative imposex mechanisms might interact and underpin the imposex response. However, there was also evidence that indicated potential novel candidate targets and related pathways involved in the TBT toxicity response, such as transcription factors and nuclear receptors. Of these, we have identified a PPARγ homolog gene, not previously reported for gastropods (Stewart et al. 1994; Cajaraville et al. 2003), which is up-regulated when females are exposed to TBT. In vertebrates the PPARs are members of the superfamily of nuclear hormone receptors and initially identified as mediators of peroxisome proliferation (Issemann and Green 1990). The PPARs have been identified in a wide range of vertebrate species as ligand-activated transcription factors playing vital roles in a variety of cell functions (Ibabe et al. 2005). There are three described PPAR subtypes - PPARα, PPARβ and PPARγ. The latter is mainly expressed in adipose tissue and has important roles in lipid metabolism, cell proliferation, and the inflammation processes (Kersten et al. 2000).

Since the produced molecular data (chapter 3) suggested a putative link between imposex development and PPARγ signalling, here we investigated the role of this novel

candidate gene and accompanying pathway by inducing the focal mechanism in vivo. Therefore, Nucella lapillus females were injected with Rosiglitazone (Rosi), a well-known PPARγ ligand (Nakanishi 2007), to test if this compound can induce imposex in these animals.

4.3. MATERIAL AND METHODS

4.3.1. Sampling and laboratory exposure to Rosiglitazone

Adult specimens of N. lapillus were collected in September 2010 from Cable Bay (Anglesey, UK), one of the reference "pristine" sites (see chapter 2) where imposex levels are very low (Vas Deferens Sequence Index (VDSI)=0.36)). After 2 weeks of acclimatization, females were narcotized in an aqueous solution of 7% $MgCl_2$ for approximately 60 minutes and then injected into the foot, with the ethanol solutions of the compounds to be tested: Rosi and TBT using a microsyringe (Hamilton). A group of females were also injected with ethanol, in order to provide a negative control (solvent control) as this was used as a carrier of the tested compounds. Another group of non-narcotized and non-injected females was also included in this experiment to additionally control the effects of injection and narcotization (seawater control). The animals that were injected received a volume of solution proportional to their soft body wet weight (SBW) (i.e., weight without shell). In order to estimate the SBW, a linear regression was obtained between the weight of animals with shell (X), and after removing the shell (SBW) (±0.0001 g) for a total of 30 females, with the formula: Y=0.082X+0.270. Hence, each female used for the experiment (N=400) was individually standardised according weight and the volume of solution injected was 1 µl/µg SBW. The following experimental conditions were tested: "Rosi1": 1 µg of Rosi/g SBW; "Rosi2": 2 µg Rosi/g SBW; "TBT": TBTCl 1µg/g SBW (positive control); "ethanol" (solvent control); "seawater" (seawater control). Four replicate flasks per treatment, each with 10 animals, were used with 1L filtered seawater from the Menai Strait, UK, with constant aeration. Animals were

maintained with no food supply at 15 ^{o}C ± 1, and exposed to the same laboratory conditions as described in chapter 2 for the initial TBT exposures for up to 2 months. Physical evidence of imposex: penis length (measured to the nearest 0.01 mm with a stereo microscope and eyepiece graticule), VDS and percentage of affected females (%I) was determined after 1 and 2 months of the beginning of the experiment following the procedures described by (Gibbs et al. 1987).

4.3.2. Statistical data analysis

The non-parametric Kruskal-Wallis test (H) was used to assess significant differences in penis length and vas deferens sequence (VDS) between treatment groups at the end of each experiment. This test was followed by pairwise comparisons to identify the groups that were significantly different from each other. Analyses were performed using PASW Statistics 18 software. Differences in the percentage of females with imposex between treatments were analysed by chi-square test at a significance level of 0.05.

4.4. RESULTS

After one month of experiment, female penis length (H=12.17, P=0.016) and VDS levels (H=31.14, P<0.001) were significantly different between treatments (Fig. 1A). In the case of penis length the results account for significant differences between the seawater control and Rosi2 and also between females injected with Rosi1 and Rosi2. No significant differences were observed between females injected with ethanol versus any other treatment (Table 1). Regarding VDS, all pairwise comparisons between females from the seawater control and the other treatments were significantly different but no other treatment comparisons were significant (Table 1).

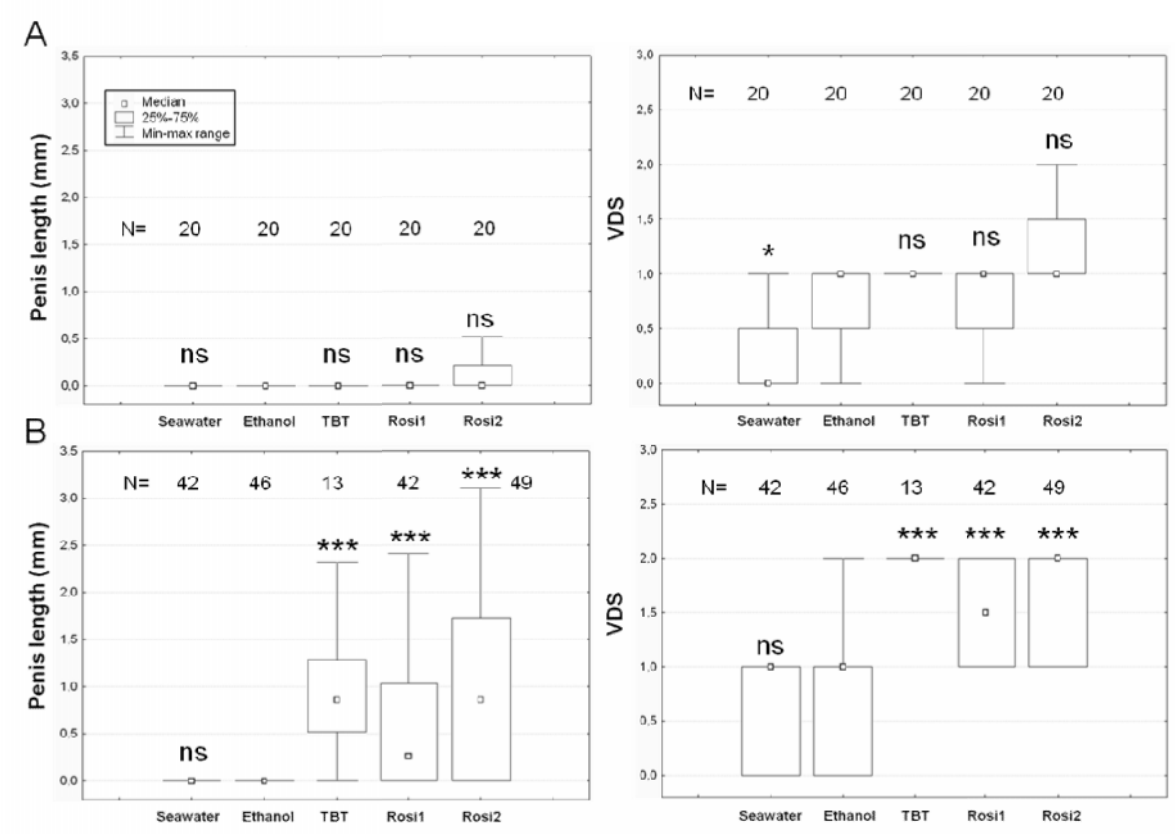


Figure 1. Nucella lapillus imposex levels (A) after 1 month (N=100) and (B) after 2 months (N=192) of injection with Ethanol (solvent control), TBTCl 1µg/g SBW (TBT), Rosiglitazone 1 µg/g SBW (Rosi1), Rosiglitazone 2 µg/g SBW (Rosi2). A further group of females were maintained without being injected (Seawater). VDS (Vas Deferens Sequence). Boxplots represent medians (point), the 25%-75% quartiles (box) and minimum and maximum range of values (±bars). The significance of the statistical comparisons between ethanol and the other experimental conditions are represented by: ns=not significant; *=P<0.05; **=P<0.01; ***=P<0.001.

After 2 months the significant differences between treatments were much more pronounced (penis length: H=84.31, P<0.001; VDS levels: H=103.73, P<0.001) (Fig. 1B). This is expected because time is needed for imposex development and for the possible imposex induction by TBT or Rosi to become evident. Pairwise comparisons revealed that there were no significant differences concerning penis length between seawater control and solvent control females. We can thereby assume that solvent had no influence in the observed results. However, females injected with TBT, Rosi1 and Rosi2 developed penises that were significantly larger than those from the solvent control, clearly indicating that both former compounds can induce penis growth. The extent of penis development with

TBT and Rosi is comparable for equivalent doses after 2 months of experiment (Fig. 1; Table 1). Similarly, there were no significant differences in VDS levels between seawater and solvent controls and, besides, VDS levels of females injected with either TBT or both concentrations of Rosi were significantly different from solvent control. In this case, TBT and Rosi also caused a significant growth of the vas deferens at comparable degrees. Moreover, the percentage of females affected by imposex (seawater=54%; ethanol=60%, TBT, Rosi1 and Rosi2=100%) was not significantly different between seawater and solvent controls (chi-square P=0.562) but it was highly significant between controls and all the other treatments (chi-square P<0.001). Overall, the results noticeably indicate that Rosi can induce imposex development in N. lapillus as efficiently as TBT two months after injection.

Table 1. Pairwise comparisons between treatments, for penis length and VDS, after 1 month (1M) and 2 months (2M) after injection. ts: Test statistic; significant differences indicated in bold.

		Seawater	Ethanol	TBT	Rosi1
Penis 1M	**Ethanol**	P=1.000			
	ts	-2.350			
	TBT	P=1.000	P=1.000		
	ts	-4.825	-2.475		
	Rosi1	P=1.000	P=1.000	P=1.000	
	ts	0.000	2.350	4.825	
	Rosi2	**P=0.030**	P=0.152	P=0.638	**P=0.030**
	ts	-12.825	-10.475	-8.000	-12.825

		Seawater	Ethanol	TBT	Rosi1
VDS 1M	**Ethanol**	**P=0.014**			
	ts	-24.800			
	TBT	**P<0.001**	P=1.000		
	ts	-35.800	-11.000		
	Rosi1	**P=0.030**	P=1.000	P=0.991	
	ts	-23.000	1.800	12.800	
	Rosi2	**P<0.001**	P=0.692	P=1.000	P=0.405
	ts	-38.900	-14.100	-3.100	-15.900

		Control	Ethanol	TBT	Rosi1
Penis 2M	**Ethanol**	P=1.000			
	ts	-1.380			
	TBT	**P<0.001**	**P<0.001**		
	ts	-82.231	-80.850		
	Rosi1	**P<0.001**	**P<0.001**	P=0.145	
	ts	-45.726	-44.346	36.505	
	Rosi2	**P<0.001**	**P<0.001**	P=1.000	P=0.316
	ts	-67.000	-65.620	15.231	-21.274

		Control	Ethanol	TBT	Rosi1
VDS 2M	**Ethanol**	P=1.000			
	ts	-5.533			
	TBT	**P<0.001**	**P<0.001**		
	ts	-100.038	-94.506		
	Rosi1	**P<0.001**	**P<0.001**	P=0.441	
	ts	-67.250	-61.717	32.788	
	Rosi2	**P<0.001**	**P<0.001**	P=1.000	P=1.000
	ts	-79.112	-73.580	20.926	-11.862

4.5. DISCUSSION

Several hypotheses have been proposed over the recent years to explain how the environmental contaminants TBT and triphenyltin (TPT) induce the development of imposex in prosobranch gastropods. One important hypothesis suggests that imposex is triggered through binding of these triorganotins to the retinoid X receptors (RXRs). Nishikawa et al. (2004) showed that TBT and TPT are high affinity ligands for the human RXR (hRXR) and that the injection of 9-cis retinoic acid (RA), a known ligand of hRXRs in vitro, into females of the rock shell (Thais clavigera) induced the development of imposex in these animals. Subsequently, it was shown that 9-cis RA is able to induce imposex in other prosobranch species such as N. lapillus and Nassarius reticulatus (Castro et al. 2007; Sousa et al. 2010), providing further evidence that the RXR signalling pathway may have a key involvement in this phenomenon. The pathway may also be implicated in accessory sex organ development in male gastropods, leading to an increase in penis length when specimens are treated with TBT or 9-cis RA (Castro et al. 2007; Lima et al. 2011).

Nishikawa et al. (2004) cloned the RXR homologue from T. clavigera and found that the ligand-binding domain (LBD) of rock shell RXR is very similar to vertebrate RXR and can bind to 9-cis RA and to organotins. Similarly, Castro et al. (2007) cloned the orthologue of N. lapillus RXR and also proved that it binds 9-cis RA in vitro. According to Le Maire et al. (2009) the high affinity of TBT for hRXRα derives from the covalent interaction linking the tin atom to residue Cys432 in the RXRα LBD and also from the direct van der Waals contacts between all the TBT atoms and the RXRα residues; likewise, triphenyltin show similar features to TBT in this respect.

The above experiments developed with the RXR ligand 9-cis RA yield important insights regarding the involvement of nuclear receptors in the imposex phenomenon, but it does not mean necessarily that 9-cis RA itself is involved in this process. In fact, RXR:RAR dimers and RAR activation by the morphogen alltrans retinoic acid (ATRA) is well known but the 9-cis RA isomer, recognized as a high-affinity ligand to RXR (all subtypes -α, -β and -γ) in vitro, was still not clearly detected in vivo (Germain et al. 2006; Lefebvre et al. 2010). Besides, it is noteworthy that while 9-cis RA is known to be a ligand for both RARs and RXRs in vertebrates (Germain et al. 2006), its action on molluscs is perhaps

made exclusively through RXR binding because is remains questionable whether RARs occur in molluscs; in fact, so far there are evidences that RARs are recent chordate-specific novelties that have no protostome (where molluscs are included) orthologs (Thornton 2003). However, RAR-related homolog genes were identified in N. lapillus response to TBT (chapter 3) suggesting that they may be present in molluscs. On the contrary, RXRs are supposed to be ancient and distributed throughout the Eumetazoa and, accordingly, their presence/activity in prosobranchs is well established (Nishikawa et al. 2004; Bouton et al. 2005; Castro et al. 2007; Sternberg et al. 2008; Sousa et al. 2010). RXR is also ubiquitously present in the whole gastropod body as demonstrated by Castro et al. (2007) who detected, by real-time PCR, ubiquitous basal expression of RXR in a variety of tissues of N. lapillus, though the highest levels were recorded in ovary and testis. Lima et al. (2011) observed RXR transcription levels in several tissues of N. lapillus after TBT exposure and found that females in advanced stages of imposex displayed elevated RXR transcription in penis, identical to those of males, suggesting a functional role of RXR in the penis growth; in other tissues the response was very different: in gonads and digestive gland the transcription was not affected by TBT, whereas in the central nervous system a down-regulation was observed in females both before and after imposex initiation. Horiguchi et al. (2010) obtained somewhat different results, although working with a different gastropod species and a different organotin, as they observed a significant increase in RXR expression levels in the central nervous system, penis forming area and ovary of T. clavigera females exposed to TPT.

Currently therefore, there is strong evidence implicating the RXR signalling pathway in the development of imposex in prosobranchs. However, the results obtained in the current work disclose innovative and important complementary information. In fact, following the results from pyrosequencing and microarray analysis (chapter 3) indicating that a PPARγ homolog gene is up-regulated in N. lapillus females exposed to TBT, we tested if Rosi, a potent and selective PPARγ ligand, could by itself trigger the development of imposex. Females were injected with Rosi and the results clearly demonstrated that this compound can effectively induce the development of imposex at a degree comparable to that of TBT. Consequently, a more elaborated hypothesis is

proposed where TBT could act in the RXR:PPARγ signalling pathway and activate the transcription of PPARγ target genes, which would then trigger imposex development.

It is important to introduce here a brief description of the possible interplay between RXR and other nuclear receptors. In fact, RXR has a promiscuity and enigmatic role in the sense that it can bind to DNA as homodimers and homotetramers to regulate their own specific signalling pathways but it can also dimerize with diverse nuclear receptors and exert transcriptional control on other cell biology functions (Lefebvre et al. 2010). The heterodimers containing RXR can be functionally classified as 'permissive' or 'non-permissive' depending on whether they can be activated, or not, through the RXR moiety. Nuclear receptors like RARs, for instance, have a high affinity for their cognate ligands and belong to the non-permissive category and exert repressive activity in the unliganded state. On the contrary, lipid-activated nuclear receptors with general low affinity for their ligands like, for example, the PPARs, are considered to be permissive and can be under the functional control of RXR binding partner (Germain et al. 2006; Lefebvre et al. 2010). The DNA motifs to which RXR heterodimers bind are generally direct repeats (DR) that contain the sequence AGGTCA and follow the 1-5 rule, i.e., RXR homodimers and RXR:PPAR heterodimers bind to DR1 motifs, RXR:RAR heterodimers preferentially bind to DR2 and DR5, and DR3 and DR4 favour DNA binding of RXR heterodimers with, respectively, nuclear receptors VDR and T3R (Lefebvre et al. 2010).

Therefore RXR heterodimers that contain the permissive partner PPAR can be activated by agonists of both RXR and the PPAR partner receptor independently or together to induce a synergistic activation (Germain et al. 2006). In this context, and according to our hypothesis, Rosi has the ability to bind to PPARγ in N. lapillus tissues and activate the heterodimer RXR:PPARγ, which then binds to the PPARγ response elements in the target gene promoter. In the case of organotins we suggest that this could occur via each or both nuclear receptors, i.e., TBT could bind to RXR and/or to PPARγ and activate the heterodimer RXR:PPARγ to produce the same result.

TBT and TPT are potent agonists to RXR, as described above, but there are contradictory reports regarding their ability to bind PPARγ. According to some authors (Kanayama et al. 2005; Hiromori et al. 2009), these organotins are strong agonists of

PPARγ but, in contrast, Le Maire et al. (2009) showed that TBT is a potent agonist of RXRα and activates efficiently this receptor, but has a weak affinity for PPARγ. The interplay between PPAR and RXR pathways is even more complex and intriguing as it was observed that in vivo activation of PPAR target genes containing a DR-1 may occur in response to RXR homodimers in the presence of RXR agonists (IJpenberg et al. 2004). Regardless of the degree of involvement of TBT on each of these receptors for triggering the RXRα:PPARγ signalling, we hypothesize that TBT may promote transcription of PPARγ target genes that prompt the development of imposex. The ability of 9-cis RA to promote imposex in laboratory experiments could putatively reside in the same pathway. In fact, it is known that 9-cis RA and synthetic RXR agonists can promote the transcription of PPAR target genes (Feige et al. 2006).

The exact mechanism downstream PPARγ target gene expression that leads to imposex development may involve many possible crossroads and still have many mysteries to unveil. It is known that in vertebrates PPARγ controls the expression of multiple genes implicated in a variety of physiological and pathophysiological processes. Its own name is misleading because PPARγ has multiple biological roles beyond those for which they were initially named because, in fact, the mediation of peroxisome proliferation was the first discovered function for one other PPAR subtype (Issemann and Green 1990). It is known that in vertebrates, PPARγ is a key transcription factor in adipocyte differentiation and an important regulator of target genes involved in glucose and lipid metabolism and in macrophage development and function; PPARγ agonists are currently commercialized as antidiabetic agents and, for instance, Rosi has been used as an oral hypoglycaemic agent in the treatment of Type II diabetes in humans for many years by sensitizing target tissues to insulin; PPARγ is also known to be involved in terminal differentiation of malignant breast epithelial cells (Knoblauch et al. 1999; Shearer and Hoekstra 2002; Feige et al. 2006; Capobianco et al. 2008; Hiromori et al. 2009). Besides all these roles, we anticipate the PPARγ may be involved in many other further important physiological functions to be discovered either with RXR or integrating additional crosstalks with other nuclear receptors. The complexity of interactions with other receptors is even wider if we consider that all nuclear receptors acting as

heterodimers with RXR can potentially compete with PPAR signalling in tissues where the amount of RXR is limiting (Feige et al. 2006). Thus, we predict that the cascades following the putative involvement of PPARγ leading to imposex are likely to involve many pathways. Some potential targets are highlighted in this thesis (e.g. RAR, ROR, ARNT, SREBP1, progestin receptor; see chapter 3), that warrant further investigation for future research unravelling the functional genomic basis of endocrine disruption.

4.6. CONCLUSION

Although there is no doubt that TBT and TPT are high affinity ligands of RXR and that this receptor occurs ubiquitously in prosobranchs, it is still not clear which are the full sets of signalling pathways that RXR is implicated as a homodimer. Moreover, the interactions between RXR and other nuclear receptors is also complex due to it being a versatile dimerization partner. Hence, the present work constitutes an important additional step to better understand the molecular mechanisms underlying imposex development in N. lapillus since we propose that TBT may act through the RXR:PPARγ signalling. Nevertheless, significant work remains to elucidate the downstream cascade of events that lead to imposex development in this species.

4.7. REFERENCES

Bouton, D., H. Escriva, R. de Mendonça, C. Glineur, B. Bertin, C. Noël, M. Robinson-Rechavi, A. de Groot, J. Cornette, V. Laudet, and R. Pierce. 2005. A conserved retinoid X receptor (RXR) from the mollusk Biomphalaria glabrata transactivates transcription in the presence of retinoids. Journal of Molecular Endocrinology 34:567-582.

Cajaraville, M. P., M. Cancio, A. Ibabe, and A. Orbea. 2003. Peroxisome proliferation as a biomarker in environmental pollution assessment. Microscopy Research and Technique 61:191-202.

Capobianco, E., V. White, R. Higa, N. Martınez, and A. Jawerbaum. 2008. Effects of natural ligands of PPARg on lipid metabolism in placental tissues from healthy and diabetic rats. Molecular human reproduction 14:491-499.

Castro, L. F. C., D. Lima, A. Machado, C. Melo, Y. Hiromori, J. Nishikawa, T. Nakanishi, M. A. Reis-Henriques, and M. M. Santos. 2007. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod Nucella lapillus. Aquatic Toxicology 85:57-66.

Feige, J., L. Gelman, L. Michalik, B. Desvergne, and W. Wahli. 2006. From molecular action to physiological outputs: Peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. Progress in Lipid Research 45:120-159.

Germain, P., P. Chambon, G. Eichele, R. Evans, M. Lazar, M. Leid, A. Lera, R. Lotan, D. Mangelsdorf, and H. Gronemeyer. 2006. International Union of Pharmacology. LXIII. Retinoid X Receptors. Pharmacological Reviews 58:760-772.

Gibbs, P. E., G. W. Bryan, P. L. Pascoe, and G. R. Burt. 1987. The use of the dog-whelk, Nucella lapillus, as an indicator of tributyltin (TBT) contamination. Journal of the Marine Biological Association of the United Kingdom 67:507-523.

Hiromori, Y., J.-I. Nishikawa, I. Yoshida, H. Nagase, and T. Nakanishi. 2009. Structure-dependent activation of peroxisome proliferator-activated receptor (PPAR) γ by organotin compounds. Chemico-Biological Interactions 180:238-244.

Horiguchi, T., T. Nishikawa, Y. Ohta, H. Shiraishi, and M. Morita. 2010. Time course of expression of the retinoid X receptor gene and induction of imposex in the rock shell, Thais clavigera, exposed to triphenyltin chloride. Analytical and Bioanalytical Chemistry 396:597-607.

Ibabe, A., E. Bilbao, and M. P. Cajaraville. 2005. Expression of peroxisome proliferator-activated receptors in zebrafish (Danio rerio) depending on gender and developmental stage. Histochemistry and Cell Biology 123:75-87.

IJpenberg, A., N. Tan, L. Gelman, S. Kersten, J. Seydoux, J. Xu, D. Metzger, L. Canaple, P. Chambon, W. Wahli, and B. Desvergne. 2004. In vivo activation of PPAR target genes by RXR homodimers. EMBO Journal. 23:2083-2091.

Issemann, I., and S. Green. 1990. Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature 347:645-650.

Kanayama, T., N. Kobayashi, S. Mamiya, T. Nakanishi, and J. Nishikawa. 2005. Organotin compounds promote adipocyte differentiation as agonists of the peroxisome proliferator-activated receptor gamma/retinoid x receptor pathway. Molecular Pharmacology 67:766-774.

Kersten, S., B. Desvergne, and W. Wahli. 2000. Roles of PPARs in health and disease. Nature 405:421-424.

Knoblauch, H., A. Busjahn, B. Muller-Myhsok, H. Faulhaber, H. Schuster, R. Uhlmann, and F. Luft. 1999. Peroxisome Proliferator–Activated Receptor γ Gene Locus Is Related to Body Mass Index and Lipid Values in Healthy Nonobese Subjects. Arteriosclerosis, Thrombosis and Vascular Biology 19:2940-2944.

Lefebvre, P., Y. Benomar, and B. Staels. 2010. Retinoid X receptors: common heterodimerization partners with distinct functions. Trends in Endocrinology and Metabolism 21.

Lima, D., M. A. Reis-Henriques, R. Silva, A. Santos, L. F. Castro, and M. M. Santos. 2011. Tributyltin-induced imposex in marine gastropods involves tissue-specific modulation of the retinoid X receptor Aquatic Toxicology 101:221-227.

Nakanishi, T. 2007. Potential toxicity of organotin compounds via nuclear receptor signaling in mammals. Journal of Health Science 53:1-9.

Nishikawa, J., S. Mamiya, T. Kanayama, T. Nishikawa, F. Shiraishi, and T. Horiguchi. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Environmental Science and Technology 38:6271-6276.

Shearer, B., and W. Hoekstra. 2002. Peroxisome Proliferator-Activated Receptors (PPARs): Choreographers of Metabolic Gene Transcription. Sigma- Cell Transmissions 18.

Smith, B. S. 1971. Sexuality in the american mud snail Nassarius obsoletus. Proceedings of the Malacological Society of London 39:377-378.

Sousa, A., C. Barroso, S. Tanabe, and T. Horiguchi. 2010. Involvement of Retinoid X Receptor in Imposex Development in Nucella lapillus and Nassarius reticulatus—Preliminary Results. Interdisciplinary studies on environmental chemistry - Biological responses to contaminants:189-196.

Sternberg, R. M., M. P. Gooding, A. K. Hotchkiss, and G. A. LeBlanc. 2010. Environmental-endocrine control of reproductive maturation in gastropods: implications for the mechanism of tributyltin-induced imposex in prosobranchs. Ecotoxicology 19:4-23.

Sternberg, R. M., A. K. Hotchkiss, and G. A. Leblanc. 2008. Synchronized expression of retinoid X receptor mRNA with reproductive tract recrudescence in an imposex-susceptible mollusc. Environmental Science & Technology 42:1345-1351.

Stewart, J. M., R. C. Carlin, J. A. Macdonald, and S. Vaniderstine. 1994. Fatty-acid-binding proteins and fatty-acid catabolism in marine invertebrates, peroxisomal beta-oxidation. Invertebrate Reproduction & Development 25:73-82.

Thornton, J. W. 2003. Nonmammalian nuclear receptors: Evolution and endocrine disruption. Pure and Applied Chemistry 75:1827-1839.

CHAPTER 5

Genetic differentiation and plasticity underlying adaptive shell variation in Nucella lapillus: an assessment using reciprocal transplant and common garden experiments

Chapter submitted as original article:

Pascoal S, Carvalho G, Creer S, Rock J, Kawaii K, Mendo S and Hughes R. 2011. Genetic differentiation and plasticity underlying adaptive shell variation in Nucella lapillus: an assessment using reciprocal transplant and common garden experiments. BMC Evolutionary Biology

Note: This chapter was co-written by Prof. Roger Hughes

5.1. ABSTRACT

Knowledge of the contributions of genotype, plasticity and their interaction to phenotypic expression is crucial for understanding the evolution of adaptive character traits in heterogeneous environments. The above was assessed in relation to adaptive shell morphology of the rocky intertidal snail Nucella lapillus by reciprocal transplantation of snails between two shores differing in exposure to wave action and rearing snails of the same provenance in a laboratory common garden experiment with crab-predation odour as treatment. Microsatellites showed population genetic differentiation indicative of semi-isolated populations. Morphometric analyses revealed plasticity of shell shape in reciprocal transplants, but also the partial retention of parental shape by F_2s in common garden controls, indicating co-gradient variation. Crab-predation odour stimulated the production of thicker shell lips, with greater response in exposed-site snails indicative of counter-gradient variation, and influenced shell shape in exposed-site but not sheltered-site snails. The combination of plasticity and local genetic adaptation may be functionally linked to spatial environmental variation experienced during an individual's lifetime and the occurrence of genetic sweeps during extreme wave action on the exposed shore or sustained selection by crab predation and desiccation on the sheltered shore.

5.2. INTRODUCTION

It has long been known that an understanding of the relative contributions of genotype and plasticity to phenotypic expression is crucial for evaluating the evolution of adaptive character traits in spatially and temporally variable habitats (Bradshaw 1965; Gould 1966; Levins 1968; Endler 1986; Stearns 1989; Travis 1994; Via et al. 1995; Schlichting and Pigliucci 1998). Yet it remains pertinent to ask what conditions favour plasticity, local genetic adaptation, or both (Pigliucci 2001). It has also been suggested that genetic variation is favoured in stable environments (Hori 1993; Smith 1993), whereas phenotypic plasticity is favoured in unstable and fluctuating environments (Stearns 1989; Scheiner 1993; Svanback et al. 2009). Other things being equal, phenotypic plasticity should be selectively advantageous over local genetic adaptation if progeny are randomly distributed among habitats presenting different fitness requirements. The advantage should be reduced, however, if plasticity only achieves an approximate match to the locally optimal phenotype (Moran 1992) and/or incurs a significant fitness cost (DeWitt 1998). However, co-occurrence of local genetic adaptation and phenotypic plasticity has been widely demonstrated in both plants and animals, including intertidal gastropods (Boulding and Hay 1993; Johannesson and Johannesson 1996; Johnson and Black 1998). A limited, but increasing body of data suggests that morphological differentiation among populations tends to be controlled by reinforcing effects of genetic differentiation and plasticity (co-gradient variation), whereas physiological or behavioural differentiation tend to be controlled by correspondingly opposing effects (contra- or counter-gradient variation (Levins 1968; Crispo 2008; Conover et al. 2009).

Here plastic and heritable components of variation in shell morphology and growth rate of the dogwhelk Nucella lapillus (L.) were assessed in relation to contrasting selection regimes associated with high and low wave exposure (Etter 1988a, 1988b; Kirby 2000a; Guerra-Varela et al. 2009). N. lapillus is a predatory snail with limited dispersal ability owing to non-planktonic larvae and a restricted crawling range (Hughes 1972). The species is commonly found on rocky shores of the North Atlantic, ranging from the most wave-exposed to the most sheltered (Crothers 1985). Spatial variation in wave-exposure embodies a complex environmental gradient, including amplitude of mechanical forces,

temperature variation and risk of desiccation that in turn influence community structure and hence the biological environment experienced by N. lapillus. The shell of N. lapillus is more globular at sites exposed to wave action and more elongated at sheltered sites (Kitching et al. 1966). The exposed-site shape possibly offers less drag (Hughes and Taylor 1997) and is characterized by a relatively larger, more rounded aperture (Kirby et al. 1994) that accommodates a larger foot. The latter enables stronger attachment to the rock and therefore greater resistance to dislodgement by waves (Kitching et al. 1966; Etter 1988a). The elongated sheltered-site shape is associated with slower evaporation by having a relatively smaller aperture (Coombs 1973) and with greater capacity for evaporative cooling through holding a relatively greater volume of extra-corporeal water within the basal whorl (Kirby et al. 1994). Moreover, the greater internal volume allows snails to withdraw further into the shell (Palmer 1990) and this, together with thickened shell walls and the relatively narrow aperture of the elongated shell (Currey and Hughes 1982), hinders attacks by crabs which tend to be abundant at sheltered sites but rare at exposed (Hughes and Elner 1979). Shell-lip thickness increases in response to perceived risk of crab predation possibly directly or indirectly through suppressed foraging behaviour and ensuing starvation (Palmer 1990; Trussell and Nicklin 2002; Brookes and Rochette 2007; Edgell and Rochette 2008; Bourdeau 2010).

Using reciprocal-transplant and common-garden experiments complemented by genetic analysis, the present aims were to (1) assess population genetic differentiation of N. lapillus taken from a sheltered and an exposed site where shell morphology and growth rate are known to differ; (2) evaluate plastic and heritable components of variation in the above traits; (3) assess evidence for co-gradient and counter-gradient variation.

5.3. MATERIAL AND METHODS

5.3.1. Reciprocal transplant experiment

Two sites in North Wales, UK, were chosen for reciprocal transplantation of N. lapillus (Fig. 1). One site, Cable Bay, is exposed to strong wave action generated by prevailing south-westerly winds from the English Channel and across the Irish Sea, while the other, Llanfairfechan, is sheltered in the lee of the prevailing winds. The experimental arena at the sheltered site was a glacial boulder of approximately 1.8m height and 7.2m circumference. The boulder was covered by a patchwork of barnacles, Semibalanus balanoides, and mussels, Mytilus edulis, which were densely colonized by barnacles. Near the substratum, the peripheral under-surface of the boulder was bare and was used by adult N. lapillus as a refuge. The experimental arena at the exposed site consisted of bed-rock densely populated by the barnacles Chthamalus montagui and Semibalanus balanoides. The bedrock was devoid of mussels and was dissected by two major crevices used by N. lapillus as refuges. One crevice extended 4.5m along-shore to intersect another crevice running 6m down-shore. N. lapillus foraged within a band some 1.8m wide along the horizontal crevice and on vertical walls about 1.4m high either side the down-shore crevice.

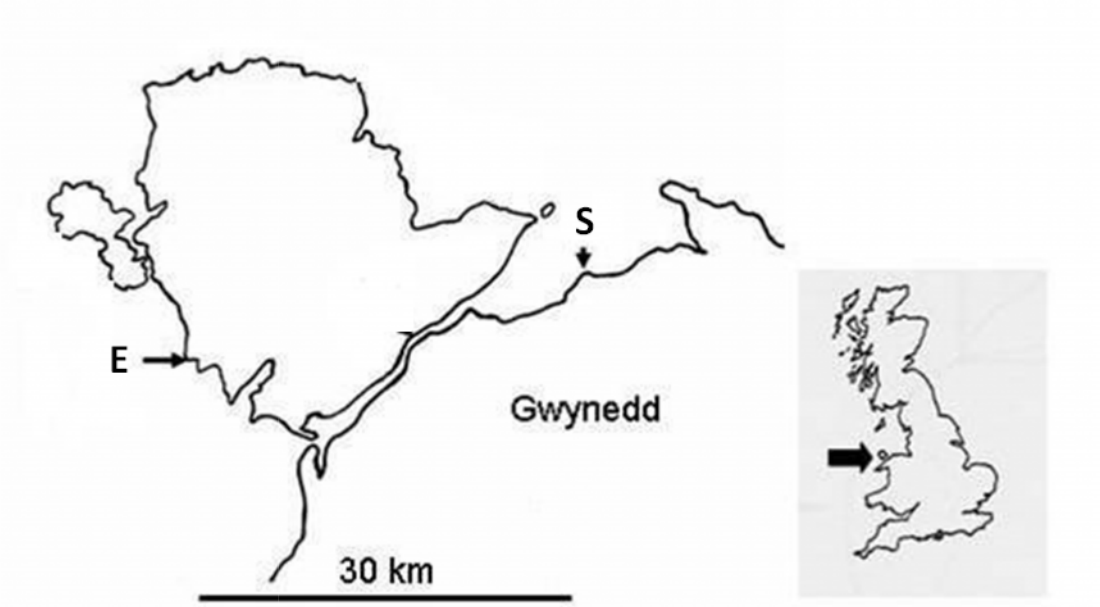


Figure 1. Location of study sites along the North Wales coastline. S: Llanfairfechan (53° 15.456´ N, 03° 58.085´ W, exposure index = 1); E: Cable Bay (53° 12.410´ N, 04° 30.290´ W, exposure index = 13). The wave exposure index is based on mean annual wind energy and fetch together with environmental modifiers (Thomas 1986).

5.3.1.1. Laboratory hatchlings

Adult dogwhelks were collected in February 2008 from the exposed and the sheltered site and maintained in aquaria supplied with running seawater closely tracking ambient outdoor temperature. Barnacles were supplied as prey and replenished as needed. The dogwhelks formed spawning aggregations and deposited egg masses on the walls of the aquaria. Once hatched juveniles had grown large enough (8-12mm, August 2008), they were labelled with a waterproof pen. Labels were covered with superglue (Loctite^{TM}) to protect against abrasion and the marked juveniles were released as summarised in Table 1.

Each of the above treatments was given a unique colour code. On release, individuals were gently irrigated with seawater to encourage them to emerge from their shells and grip the substratum. In early November 2008, marked animals were recaptured (two visits to each site), photographed, re-marked and returned to the field. In

September 2009 the experiment was completed by returning marked individuals to the laboratory where they were photographed and shell-lip thickness measured to 0.01 mm using digital callipers at three points along the lip margin while avoiding any aperture teeth (Edgell and Rochette 2008).

Table 1. Reciprocal transplant experiment: number of juveniles captured and released per treatment. S: sheltered; E: exposed; I: laboratory-hatched juveniles. Symbols as in Material and Methods. *The low number was caused by mortality during marking.

Captured	Sheltered		Exposed		Lab sheltered		Lab exposed		Total
N	1385		1220		272		400		3,277
Released	S-S	S-E	E-E	E-S	IS-S	IS-E	IE-E	IE-S	
N	514	871	610	610	72*	200	200	200	

5.3.1.2. Juveniles collected from the field

Initially it was planned to use only laboratory-hatched young produced by adults collected from the two sites, but owing to limited yield as well as expected high losses in the reciprocal-transplant experiment (Etter 1988a), the laboratory-reared juveniles were supplemented by juveniles collected directly from the field sites. Approximately 1300 N. lapillus juveniles (≤12 mm shell length) were collected from each shore (Table 1) in early July 2008 and subjected to the same mark-recapture protocol as the laboratory-reared juveniles (above). Time between collection and deployment ranged from 24 to 36 h. Unequal ratios were chosen to compensate for greater losses expected among transplants from shelter to exposure (Etter 1996). Population density within each experimental arena was conserved by relocating appropriate numbers of resident snails at a distance of about 10m.

After 3 months from initial release, shell growth was measured as the increase in shell length beyond the growth check caused by marking disturbance (Fig. 2A).

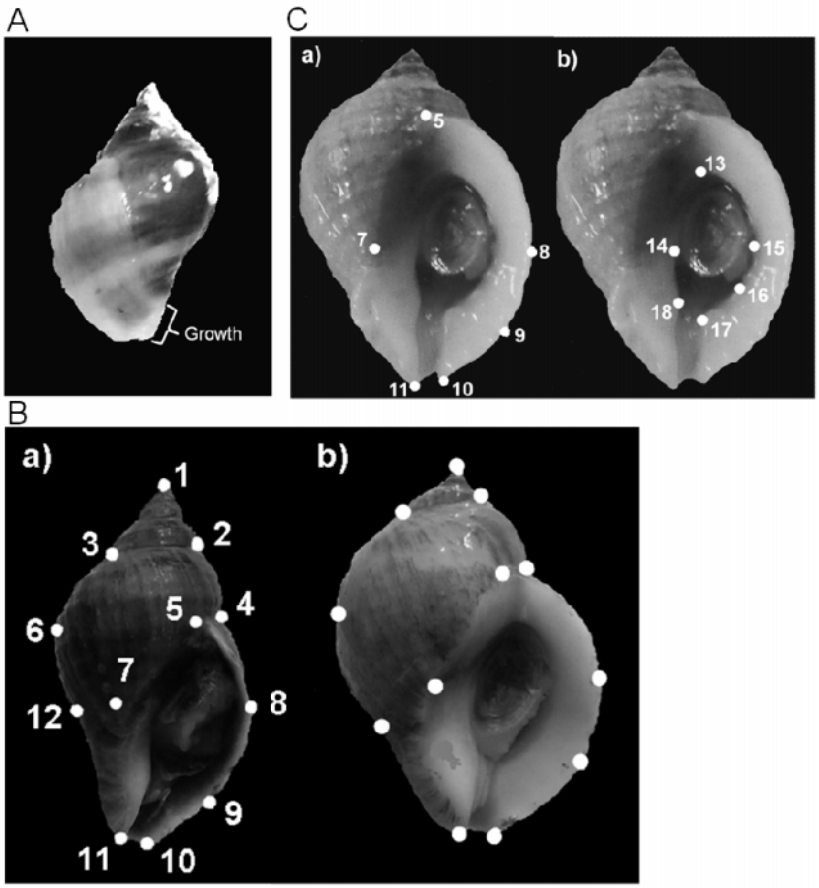


Figure 2. Nucella lapillus shell morphology: (A) Reciprocal transplant experiment: shell growth 3 months after initial release, illustrating the measured increment in shell length; (B) Morphometric analysis of Nucella lapillus: position of landmarks. a) shell collected from a site relatively sheltered from wave action (S Fig.1), b) shell collected from a site exposed to strong wave action (E Fig. 1); shell width was represented by truss 4-12, shell length by truss 1-11, aperture external width by truss 7-8 and aperture external length by truss 5-11. (C) Common garden experiment: landmark positions for (a) external aperture, (b) internal aperture; aperture internal width was represented by truss 14-15 and aperture internal length by truss 13-17.

5.3.2. Common garden experiments

Experiments were run for F_1 and F_2 generations (see below). Each experiment incorporated two duplicated common gardens. One garden, 'control', presented an environment lacking the effects of wave exposure typical of the exposed field site and of crab predation typical of the sheltered field site. Quantitative comparison of traits shown by successive generations would potentially distinguish plastic, maternal and genetic

components of variation. The other garden, 'treatment', presented an olfactory cue signalling risk of crab predation and was included to assess potential counter-gradient variation in the induced response.

Adults from the exposed and the sheltered site were collected in February 2008 and allowed to spawn separately in aquaria as for the reciprocal transplant experiment (above). Having grown large enough for marking (≤12 mm shell length), hatched juveniles (F_1 generation) were apportioned among four tanks: 25 of sheltered site ancestry and 90 of exposed site ancestry per tank (fewer juveniles of sheltered site ancestry were obtained from the brood stock, causing imbalance in numbers per treatment). Seawater was supplied via two constant-head cisterns at a rate of 3 ml s^{-1} with permanent aeration supplied from air-diffusion stones. Ambient temperature fluctuated seasonally between 8-16^{o}C. After allowing snails to acclimatize for 1 wk, two tanks were left unchanged as controls and two were supplied with water-borne olfactory cues assumed to be perceived by dogwhelks as risk of crab predation (Vadas et al. 1994). To generate the olfactory cues, four Carcinus maenas, carapace width 8–12 cm, collected from the Menai Strait (Fig. 1), were placed in the cistern supplying the treatment-tanks and fed on pre-cracked adult N. lapillus. Crabs that died were replaced within 48h. Tanks were spatially transposed at monthly intervals to avoid incidental position effects. Barnacles attached to stones were renewed as needed to maintain an unlimited supply of food for the N. lapillus in each tank. Subjects were photographed and their shell-lip thickness measured as above, first after 6 months and again after 12 months from the beginning of the experiment. Finally, shell and dry tissue weights (±0.0001g) were measured after breaking the shell, extracting the body and drying for 24h at 80^{o}C. Since eggs were laid in all treatments, opportunity was taken to continue the experiment through the F_2 generation, retrospectively assigning parentage by genetic analysis (below).

5.3.3. Morphological analyses

Photographic images were obtained in standard orientation (Fig. 2B), using graph paper as background for accurate scaling. Images were analysed using geometric

morphometrics (GM) (Cavalcanti et al. 1999; Adams et al. 2004; Carvajal-Rodriguez et al. 2005). Twelve landmarks (Fig. 2B), nine of which had previously been employed by Guerra-Varela et al. (2009), were digitized and analysed using the software TPSutil, TPSdig, and TPSrelw (Rohlf 1998; Rohlf and Bookstein 2003). Two analytical approaches were used. First, distances (truss lengths) between selected pairs of landmarks were compared among treatments (Fig. 2B, C). Data were $\log_{10}$-transformed to account for allometry and subjected to ANCOVA using appropriate covariates (see Results) to remove the effect of size. Post-hoc paired comparisons of adjusted means used Bonferroni correction and $\alpha = 0.05$. Second, GM was used to generate relative warps (RWs), from which were derived graphical representations of shell shape and allometric deformation among treatments. MODICOS software (Carvajal-Rodriguez and Rodriguez 2005) was used to obtain centroid size (measure of geometric scale, calculated as the square root of the summed squared distances of each landmark from the centroid of the landmark configuration) and RWs. The RWs, which are free of colinearity (Zelditch et al. 2004), were subjected to forward stepwise discriminant function analysis (DFA) with leave-one-out cross validation, using the software SPSS 12.0. Grouping variables for the reciprocal transplant experiment were: exposed→exposed (E-E), sheltered→sheltered (S-S), exposed→sheltered (E-S), sheltered→exposed (S-E) and for the common garden experiment: odour present, odour absent. For shell shape, GM analysis was based on the 12 landmarks shown in Fig. 2B. To assess the influence of shell thickening on aperture shape, independent GM analyses of the external and internal rims were made based on the landmarks shown in Fig. 2C. Discriminant-function centroids were subjected to ANOVA with Bonferroni correction for post-hoc paired comparisons.

5.3.4. Population genetics

5.3.4.1. Karyotype

To control for potential confounding phylogenetic effects, chromosome numbers were assessed and key mitochondrial and nuclear markers were analysed in each population. The genetic assessments were performed to confirm the absence of

karyotype polymorphism (Kirby et al. 1997), possibly associated with phylogenetic differences that might otherwise underlie adaptive or phenotypic variation (Kirby 2000a, 2000b). Five juveniles from the exposed site and five from the sheltered site were karyotyped using standard protocols (Rock et al. 1996; Pascoe 2006). Briefly, tissues were chopped and treated with two combined colchicine and 0.075 M KCl hypotonic treatments: 0.08% colchicine in 50% sea water for 45 min plus KCl for 30 min followed by colchicine 0.04% in 25% sea water for 45 min plus 60 min in KCl and finally fixed in Carnoy's solution (ethanol:acetic acid, 3:1) at 4^{o}C. The fixed tissues were transferred to a drop of 60% acetic acid on a slide at 40^{o}C, where the cells were dispersed and allowed to dry before staining for 15 min in fresh, 10% Giemsa (VWR) and finally rinsing in tap water. Five to ten slides were prepared from each juvenile. Slides were examined using a Nikon microscope eclipse 50i at 1000x magnification and the clearest chromosome sets photographed for karyotyping. Chromosome counting was complemented by comparison between populations of mitochondrial (16S) and nuclear (mMDH) genes, which vary in association with karyotypic and phenotypic polymorphism, in turn correlated with environmental variables such as wave exposure (Kirby et al. 1997; Kirby 2000a, 2004).

5.3.4.2. Mitochondrial and nuclear gene amplification

Six individual RNA samples from each population were analysed using the mitochondrial gene 16S and the nuclear gene mitochondrial malate dehydrogenase (mMDH). mMDH locus was amplified as described in (Kirby 2000b). Briefly, total RNA was extracted and DNase treated from muscle tissue using an RNeasy kit (Qiagen) followed by cDNA synthesis using the first strand cDNA synthesis kit (Fermentas). cDNA template was amplified following Kirby (2000b) protocol; firstly with mMDHP1 and mMDHP2 primers and then re-amplified with the primer pair mMDHP3 and mMDHP4 in order to get a 91 bp fragment. This gene fragment was amplified once it exhibits similar differentiation levels as the complete gene amplification (Kirby 2004). The mitochondrial 16S gene was amplified using the primers 16SNucFW (5'-TCTGACCTGCCCAGTGAAAT-3') and 16SNucRV (5'-CTCAGTCGGCCCAACTAAAA-3'), (I. Colson, personal communication). PCR

amplifications were carried out in 25 µl reactions containing 1 µl of cDNA, 0.3 pmol of each primer, 1X PCR Buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP (Promega) and 0.5 U Taq DNA polymerase (Promega) on an Biorad DNA engineTetrad2 Thermal cycler. An initial denaturation step of 5 min at 95 ^{o}C was followed by 35 cycles at 95 ^{o}C for 1 min, 53 ^{o}C for 1 min, and 72 ^{o}C for 1min followed by a final extension at 72 ^{o}C for 10 min. PCR results for both genes were sequenced using the Macrogen™ (www.macrogen.com) sequencing facility and subsequently aligned and compared using the software Bioedit (Hall 1999).

5.3.4.3. Microsatellite analysis: sampling, genotyping and statistical analysis

Adult N. lapillus from Cable Bay (N =96) and Llanfairfechan (N =96), (Fig. 1), were collected in September 2008 and fixed in absolute ethanol. DNA was extracted from foot tissue using the CTAB (Hexadecyltrimethylammonium Bromide) DNA Extraction protocol as described in Colson and Hughes (2004). Each individual was genotyped at 9 microsatellite loci (Kawai et al. 2001). Microsatellites were amplified with the Qiagen Multiplex PCR kit following the manufacturer's instructions using two different primer mixes: Nlw2, Nlw3, Nlw8, and Nlw14 in the first mix and Nlw11, Nlw17, Nlw21, Nlw25 and Nlw27 in the second mix. With slightly differences to the PCR reaction and program, the fluorescent M13 tail single-reaction nested PCR method (Schuelke 2000) was used to amplify the loci. An initial denaturation step of 15 minutes at 95 ^{o}C was followed by 13 cycles at 94 ^{o}C for 30 s, 55 ^{o}C for 90 s, and 72 ^{o}C for 60 s. In order to attach the dye tails to the PCR product, an extra 31 cycles at 94 ^{o}C of 30 s, 50 ^{o}C for 90 s and 72 ^{o}C for 60 s were performed and followed by a final extension at 60 ^{o}C for 30 minutes). Extension products were resolved on an ABI 3130xl (Applied Biosystems) and alleles were sized to an internal size standard (GeneScan-500 LIZ; Applied Biosystems) using the GeneMapper software version 4.0 (Applied Biosystems).

Raw data were screened using GenAlEx version 6.2 (Peakall and Smouse 2006) and Micro-checker (Van Oosterhout et al. 2004) to avoid scoring errors. Tests for deviations from Hardy-Weinberg proportions, heterozygote deficiencies, genotypic linkage

equilibrium and genetic heterogeneity among populations were estimated using the exact test of GENEPOP version 3.4 (Raymond and Rousset 1995). Allelic frequencies, mean number of alleles per locus, observed (H_0) and expected heterozygosity (H_E) under Hardy-Weinberg assumptions, estimates of F_{ST}, F_{IS}, and their significance per population over all loci were calculated according to Weir and Cockerham (1984) using FSTAT version 2.9.3.2 (Goudet 1995).

5.3.5. Parental analysis

Potential parents of known gender (F_1, N=112) and offspring (F_2, N=112) were genotyped as described above, using seven microsatellite markers (Nlw2, Nlw3, Nlw8, Nlw11, Nlw21, Nlw25 and Nlw27). Parental analysis was performed using CERVUS (Marshall et al. 1998).

5.4. RESULTS

5.4.1. Reciprocal transplant experiment

5.4.1.1. Recovery rates

Losses of marked snails were heaviest in the first months of the experiment (July-November 2008), reducing to low levels between November 2008 and July 2009 (Fig. 3). Percentage of snails released in July 2008 and recaptured in July 2009 (not shown in Fig. 3) were ranked as follows: snails reared as juveniles in the laboratory E-E > E-S > S-S > S-E, snails collected as juveniles from the field, S-S > E-S > E-E > S-E.

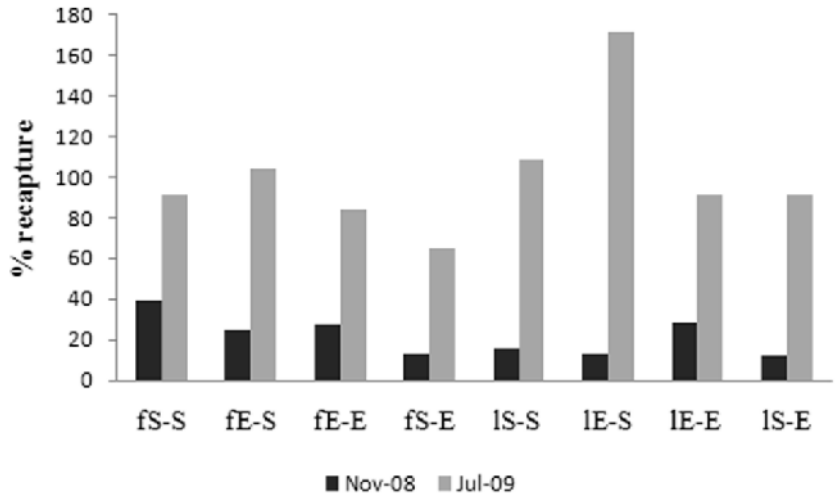


Figure 3. Reciprocal transplant experiment: recapture rates. Nov-08 = percentage of snails released at the beginning of the experiment (August 2008) and recaptured in November 2008; Jul-09 = percentage of snails re-released in November 2008 and recaptured at the end of the experiment in July 2009 (percentages > 100% reflect the recapture of snails released in August 2008 but missed in November 2008). E: exposed; S:sheltered; f: juveniles collected from the field; l: laboratory-hatched juveniles; fS-S (Nnov = 203; Njul = 185); fE-S (Nnov = 153; Njul = 160); fE-E (Nnov = 169; Njul = 142); fS-E (Nnov = 109; Njul =71); lS-S (Nnov = 11; Njul =12); lE-S (Nnov = 25; Njul = 43); lE-E (Nnov = 57; Njul =52); lS-E (Nnov = 24; Njul = 22).

5.4.1.2. Shell morphology

In treatments S-E and S-S, only one and two laboratory-hatched snails respectively were recovered at the end of the experiment. For the remaining treatments, preliminary GM analysis revealed no significant difference in shape characteristics between laboratory-reared and field-collected individuals. Data for laboratory-hatched and field-collected snails therefore were pooled for further analysis.

5.4.1.2.1. Truss lengths

Both shell width and aperture external width adjusted to shell length were greatest in exposed-site control snails, least in sheltered-site controls and intermediate in reciprocal transplants (Table 2). Shell width and aperture external width were greater in

transplants from the exposed to the sheltered site than in sheltered-site controls but less than in exposed-site controls; smaller in transplants from the sheltered to the exposed site than in exposed-site controls but greater than in sheltered-site controls; not significantly different between reciprocal transplants. Aperture external width adjusted to aperture external length did not differ significantly among treatments, but was ranked higher in exposed-site controls and transplants than in sheltered-site controls and transplants.

Table 2. Comparison of standardised truss lengths. ns = non-significant, * = P < 0.05, ** = P < 0.01, *** = P < 0.001. Bold-face type indicates ranked treatment effects.

Reciprocal Transplant Experiment: E-E = (exposed-exposed) (n = 114), E-S = (exposed-sheltered) (n = 127), S-S = (sheltered-sheltered) (n = 127), S-E = (sheltered-exposed) (n = 12).

Response variable	Covariate	Paired comparison of adjusted means
log(shell width)	log(shell length)	(E-E) $>^{***}$ (E-S) $>^{ns}$ (S-E) $>^{***}$ (S-S); E-E $>^{***}$ S-S, E-E $>^{**}$ S-E, E-S $>^{***}$ S-S
Log(aperture external width)	Log(shell length)	(E-E) $>^{***}$ (E-S) $=^{ns}$ (S-E) $>^{*}$ (S-S); E-E $>^{***}$ S-S, E-E $>^{**}$ S-E, E-S $>^{***}$ S-S
log(aperture external width)	log(aperture external length)	(E-E) $>^{ns}$ (E-S) $>^{ns}$ (S-S) $>^{ns}$ (S-E); E-E $>^{ns}$ S-E, E-E $>^{ns}$ S-S, E-S $>^{ns}$ S-E
log(shell-lip thickness)	log(shell length)	(E-E) $>^{ns}$ (E-S) $>^{*}$ (S-E) $>^{*}$ (S-S); E-E $>^{***}$ S-S, E-E $>^{***}$ S-E, E-S $>^{***}$ S-S

Common Garden Experiment: ET = exposed treatment (crab-predation odour) (n = 39), EC = exposed control (n = 51), ST = sheltered treatment (n = 8), SC = sheltered control (n = 16).

Response variable	Covariate	Paired comparison of adjusted means
log(shell width)	log(shell length)	ET $>^{*}$ EC $>^{ns}$ ST $>^{ns}$ SC; ET $>^{**}$ ST, ET $>^{**}$ SC, EC $>^{**}$ ST
Log(aperture external width)	Log(shell length)	ET $>^{ns}$ EC $>^{***}$ SC $>^{ns}$ ST; ET $>^{***}$ ST, ET $>^{***}$ SC, EC $>^{***}$ ST
log(aperture external width)	log(aperture external length)	ET $>^{*}$ EC $>^{ns}$ SC $>^{ns}$ SE; ET $>^{ns}$ ST, ET $>^{***}$ SC, EC $>^{ns}$ ST
log(aperture internal width)	log(aperture internal length)	EC $>^{ns}$ ST $>^{ns}$ SC $>^{ns}$ ET; EC $>^{***}$ ET, EC $>^{*}$ SC, ST $>^{ns}$ ET
log(shell-lip thickness)	log(shell length)	ET $>^{***}$ ST $>^{ns}$ EC $>^{***}$ SC; ET $>^{***}$ SC, ET $>^{***}$ EC, ST $>^{***}$ SC

P v. F1 generations: EP = exposed P generation (n = 31), EF_1 = exposed F_1 generation (n = 29), SP = sheltered P generation (n = 32), SF_1 = sheltered F_1 generation (n = 24). P generation refers to reciprocal-transplant-experiment control snails as proxy for snails that laid eggs in the laboratory.

Response variable	Covariate	Paired comparison of adjusted means
log(shell width)	log(shell length)	$EP >^{***} EF_1 >^{*} SP >^{ns} SF_1$; $EP >^{***} SF_1$, $EP >^{***} SP$, $EF_1 >^{**} SF_1$
Log(aperture external width)	Log(shell length)	$EP >^{ns} EF_1 >^{***} SF_1 >^{ns} SP$; $EP >^{***} SP$, $EP >^{***} SF_1$, $EF_1 >^{***} SP$
log(aperture external width)	log(aperture external length)	$EP = EF_1 >^{ns} SP >^{ns} SF_1$; $EP >^{*} SF_1$, $EP >^{ns} SP$, $EF_1 >^{*} SF_1$

F1 v. F2 generations: EF_1 (n = 29), EF_2 (n = 27), SF_1 (n = 24), SF_2 (n = 5).

Response variable	Covariate	Ranked adjusted means
log(shell width)	log(shell length)	$EF_1 >^{*} SF_1 >^{ns} EF_2 >^{ns} SF_2$; $EF_1 >^{**} EF_2$, $EF_1 >^{ns} SF_2$, $SF_1 >^{ns} SF_2$
Log(aperture external width)	Log(shell length)	$EF_1 >^{***} EF_2 >^{ns} SF_1 >^{ns} SF_2$, $EF_1 >^{***} SF_1$, $EF_1 >^{***} SF_2$, $EF_2 >^{ns} SF_2$
log(aperture external width)	log(aperture external length)	Not tested due to heterogeneity of slopes

F2 generation

Response variable	Covariate	Ranked adjusted means
log(shell width)	log(shell length)	$EF_2 >^{ns} SF_2$
Log(aperture external width)	Log(shell length)	$EF_2 >^{*} SF_2$
log(aperture external width)	log(aperture external length)	$EF_2 >^{ns} SF_2$

5.4.1.2.2. Shell-lip thickness

Shell-lip thickness adjusted to shell length was greatest in exposed-site controls, least in sheltered-site controls and intermediate in transplants (Table 2). Shell-lip thickness in transplants from the exposed to the sheltered site was greater than in sheltered-site controls but not significantly different from exposed-site controls; smaller in transplants from the sheltered to the exposed site than in exposed-site controls but greater than sheltered-site controls; greater in transplants from the exposed to the sheltered site than in transplants from the sheltered to the exposed site.

5.4.1.2.3. Relative warps

Discriminant functions were derived from 8 RWs showing statistically significant treatment effects. All paired comparisons between discriminant-function centroids were statistically significant (ANOVA) except S-E v. E-E and S-E v. E-S. Individuals were correctly classified in 67-82% of cases except for transplants from the sheltered to the exposed site, all of which were misclassified as snails of exposed-site provenance (Table 3).

Table 3. Reciprocal transplant experiment. Cross-validated percentage group membership predicted by discriminant function analysis of RW scores. Symbols as in Table 2.

	Predicted group membership				
Treatment	E-E	E-S	S-S	S-E	Total
E-E	73.7	23.7	2.6	0	100 (114)
E-S	15.8	67.5	16.7	0	100 (120)
S-S	1.6	15.7	81.9	0.8	100 (127)
S-E	50	50	0	0	100 (12)

Juveniles transplanted from the sheltered site to the exposed site developed a more globular shell with a relatively larger aperture than did controls at the sheltered site, whereas juveniles transplanted from the exposed shore to the sheltered site developed a more elongated shell with a relatively smaller aperture than controls at the

exposed site (Fig. 4). Morphological convergence, however, was incomplete, with transplants retaining some resemblance to their controls.

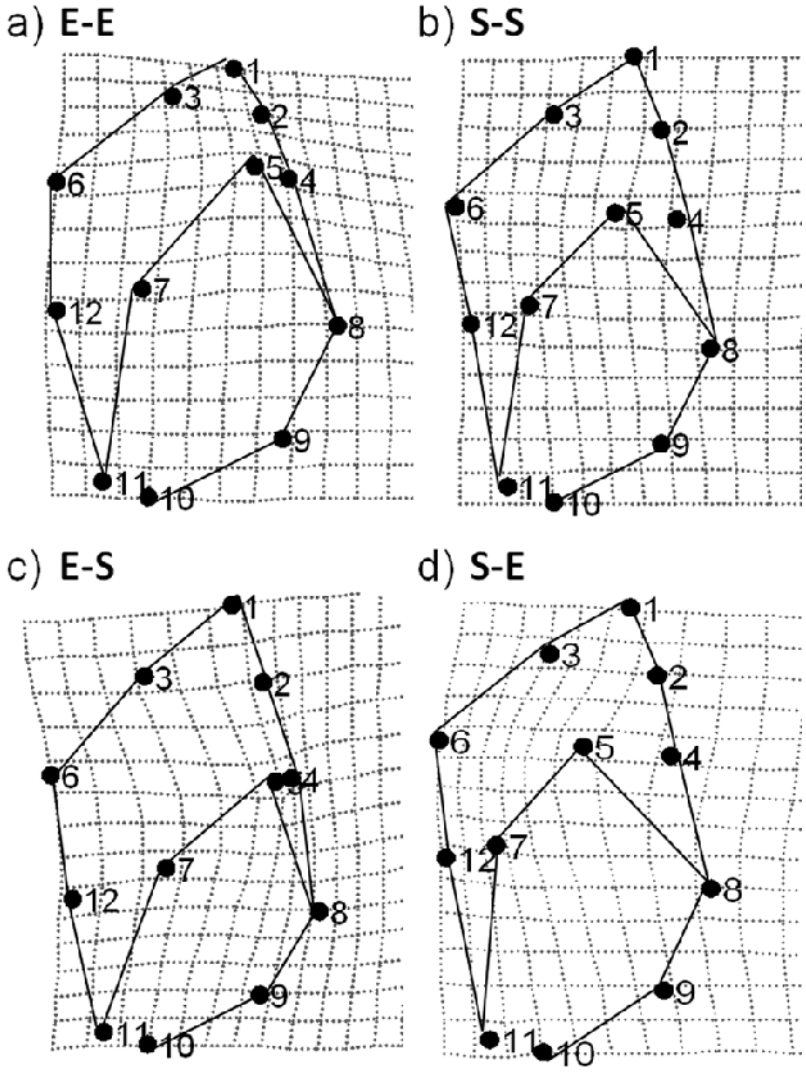


Figure 4. Reciprocal transplant experiment: thin-plate spline deformations from GM analysis of shells after 12 months in the field. E: exposed; S: sheltered.

5.4.1.3. Shell growth

Initial shell length did not differ significantly between snails of exposed-site and sheltered-site ancestry (t test: mean sheltered = 10.0mm, S.E. = 0.9mm; mean exposed = 9.3mm, S.E. = 0.5mm; t = 1.014; P = 0.321). Shell growth over the first 3 months after

initial release was ranked E-S > E-E > S-S > S-E (Fig. 5A). By the end of the experiment, initial growth checks had become obscured, preventing measurement of incremental growth. After 12 months in the field, however, shell length was ranked E-S > S-S > E-E > S-E (Fig. 5B).

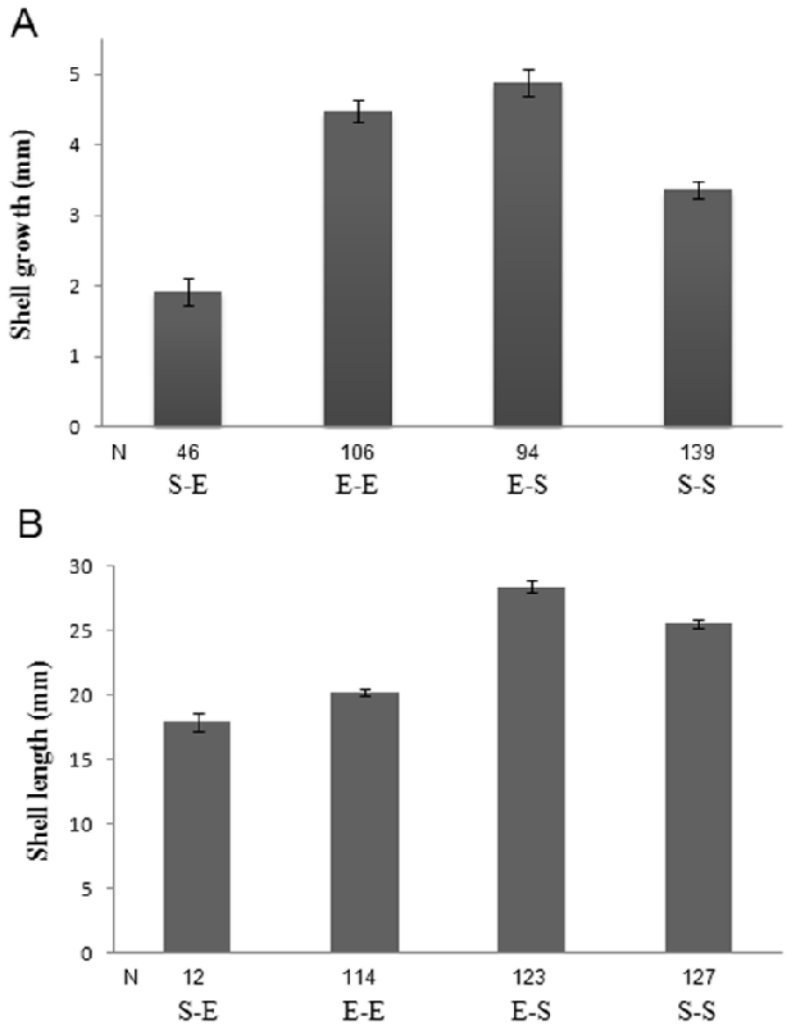


Figure 5. Reciprocal transplant experiment: (A) Shell growth after 3 months in the field; ANOVA, all paired comparisons P<0.001. (B) Shell length after 12 months in the field; ANOVA, all paired comparisons P<0.001. E: exposed; S: sheltered.

5.4.2. Common garden experiment

Within 12h of exposure to crab-predation odour, snails tended to crawl up the sides of the tanks and were less frequently seen than controls among stones bearing the food supply of barnacles. Snails receiving crab-predation odour produced fewer egg capsules than controls (crab-predation odour, N = 47, mean = 54, S.E. = 8.0; controls, N = 57, mean = 247.5, S.E. = 27.5; t = 6.75, P = 0.021), but there was no significant difference in the size of egg capsules produced (crab-predation odour, mean = 6.7, S.E. = 0.185; control, mean = 7.1, S.E. = 0.139, t = 1.56, P = 0.126) or in the number of eggs per capsule (crab-predation odour, mean = 14.4, S.E. = 3.71; control, mean = 12.1, S.E. = 1.87; t = 1.90, P = 0.198).

F_1 parents comprised 36 males and 54 females of exposed-site ancestry and 10 males and 14 females of sheltered-site ancestry. Of the F_2 progeny, 112 snails survived for 12 mo and genotyping unequivocally assigned 85 of these to known parentage, 5 having sheltered-site ancestry, 52 exposed-site ancestry and 28 mixed ancestry.

5.4.2.1. Shell and body growth

Growth measurements were taken only for snails of exposed-site ancestry, since removal of 20 snails per lineage for a parallel study left insufficient numbers of the sheltered-site lineage in the F_1 generation and too few snails of sheltered-site ancestry were produced in the F_2 generation.

At 12 months from the beginning of the experiment, mean shell length of F_1 snails exposed to crab-predation odour did not differ significantly from that of controls (t test: control mean = 21.24mm, S.E. = 0.29, treatment mean = 21.41mm, S.E. = 0.34, t =1.01, P = 0.711). Shell mass was consistently greater in snails exposed to crab-predation odour than in controls (Fig. 6A). Dry tissue mass of snails exposed to crab-predation odour was less than that of controls (Fig. 6B; ANCOVA on log-transformed data, $P < 0.001$). F_2 snails showed similar trends to F_1 snails (ANCOVA, log(shell mass) P <.0.001), log(dry tissue mass) $P < 0.001$).

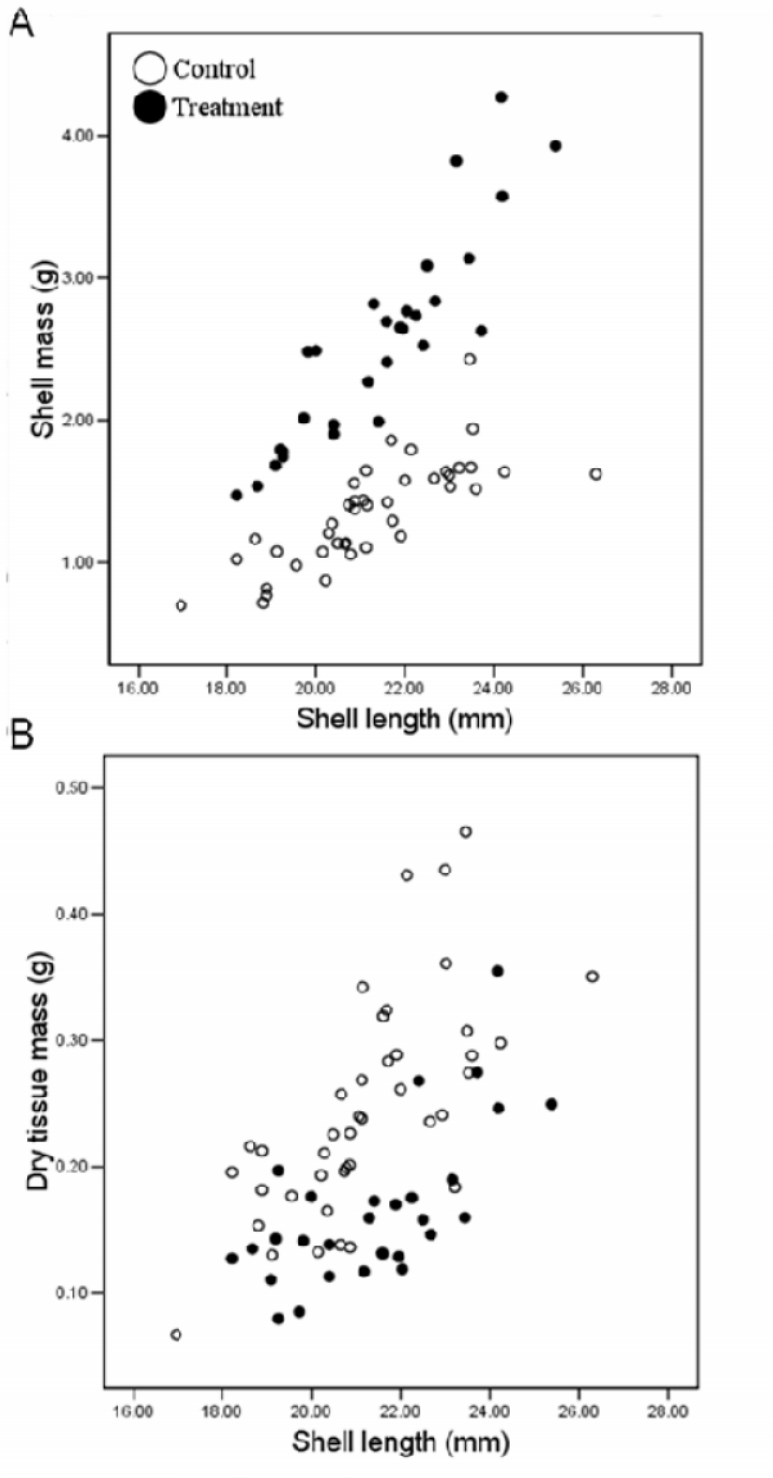


Figure 6. Common garden experiment: size measurements of exposed site snails at 12 months. (A) Shell mass; ANCOVA, treatment*shell length $P < 0.001$. (B) Dry tissue mass; ANCOVA, parallelism confirmed, treatment $P < 0.001$; control mean=0.247, S.E. = 0.009, treatment mean = 0.163, S.E. = 0.011.

5.4.2.2. Effect of crab-predation odour on shell morphology

5.4.2.2.1. Truss lengths

Shell width adjusted to shell length was greater in treatment than in control snails of exposed-site ancestry, greater in both treatment and control snails of exposed-site ancestry than in snails of sheltered-site ancestry, but not significantly different between treatment and control snails of sheltered-site ancestry (Table 2). Aperture external width adjusted to shell length was not significantly different between treatment and control snails of either lineage, but was greater in both treatment and control snails of exposed-site ancestry than in snails of sheltered-site ancestry (Table 2). Aperture external width adjusted to aperture external length was greater in treatment than in control snails of exposed-site ancestry, but not significantly different between treatment and control snails of sheltered-site ancestry, neither between treatment nor control snails of exposed-site ancestry and snails of sheltered-site ancestry (Table 2). Aperture internal width adjusted to aperture internal length was greater in control than in treatment snails of exposed-site ancestry, as well as greater in control snails of exposed-site ancestry than in controls of sheltered-site ancestry, but not significantly different between treatment and control snails of sheltered-site ancestry, nor between treatment snails of exposed-site ancestry and treatment snails of sheltered-site ancestry (Table 2).

Shell-lip thickness adjusted to shell length was greater in treatment than in control snails of exposed-site and sheltered-site ancestry and greater in treatment and control snails of exposed-site ancestry than in snails of sheltered-site ancestry (Table 2).

No F_2 snails of sheltered-site ancestry were available for the crab-predation treatment (above), but among F_2 snails of exposed-site ancestry, shell-lip thickness was greater in those exposed to crab-predation odour than in controls (ANCOVA log-transformed data, $P < 0.001$).

5.4.2.2.2. Relative warps

All paired comparisons were statistically significant (DFA, $P < 0.001$) except sheltered treatment v. sheltered control. Exposed-site treatment and control snails had more globular shells and relatively larger and more rounded apertures than corresponding sheltered-site snails (Fig. 7).

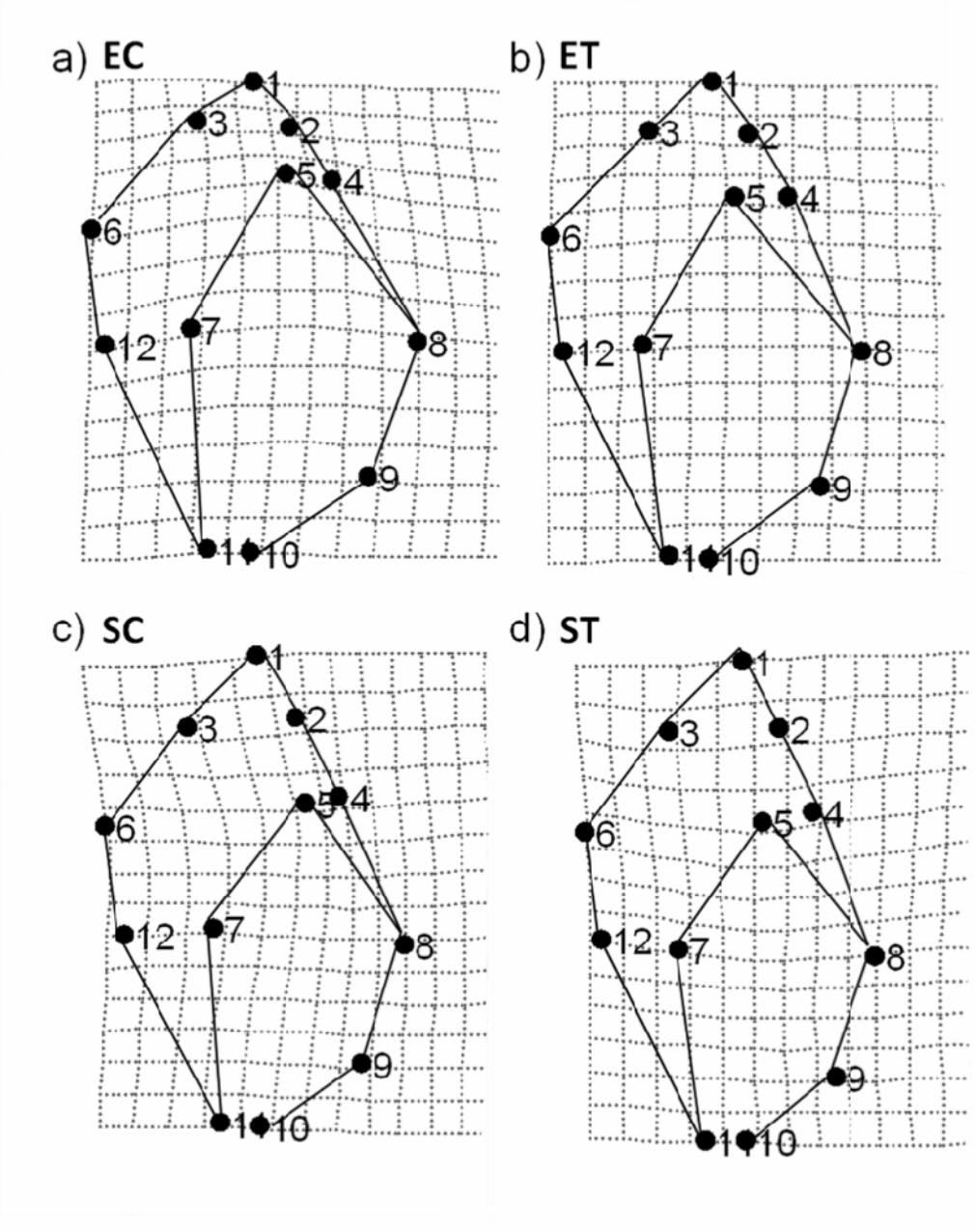


Figure 7. Common garden experiment: thin-plate spline deformations. E: exposed; S: sheltered; C: control; T: treatment (crab-predation odour).

Analysis of aperture measurements alone confirmed the above results on aperture shape both for external landmarks (Fig. 2B; DFA, $P < 0.001$ for all paired comparisons except sheltered treatment v. sheltered control) and internal landmarks (Fig. 2C, DFA, $P < 0.001$ for all paired comparisons).

5.4.2.3. Heritable variation in shell morphology

5.4.2.3.1. Truss lengths

Mean shell width adjusted to shell length was smaller for the F_1 than for the P generation, but within each generation was greater for snails of exposed-site ancestry than for snails of sheltered-site ancestry (Table 2). Mean aperture external width adjusted to shell length did not differ significantly between generations, but within each generation was greater for snails of exposed-site ancestry than for those of sheltered-site ancestry (Table 2).

Mean aperture external width adjusted to aperture external length was not significantly different between generations, neither between snails of exposed-site ancestry and sheltered-site ancestry within the P generation, but was greater for snails of exposed-site ancestry within the F_1 generation (Table 2).

Statistical comparison of F_1 and F_2 snails was avoided due to low numbers of sheltered-site F_2 snails and heterogeneity of slopes within the exposed-site lineage. Within the exposed-site lineage mean shell width and mean aperture external width adjusted to shell length were ranked higher for the F_1 than for the F_2 generation, but within the sheltered-site lineage mean shell width was ranked higher for the F_2 than for the F_1 generation (Table 2). Mean aperture external width adjusted to aperture external length was ranked higher for the F_1 than for the F_2 generation and within generations was ranked higher for the exposed-site than for the sheltered-site lineage (Table 2).

5.4.2.3.2. Relative warps

Discriminant-function centroids for F_1 control snails were different from those for field controls (DFA, $P < 0.001$) for both the sheltered-shore and exposed-shore lineages. F_1 snails of sheltered-site ancestry were more globular with relatively wider apertures than their native counterparts, whereas those of exposed-site ancestry were narrower (Fig. 8). All paired comparisons of discriminant-function centroids between F_1 exposed controls, F_1 sheltered controls, F_2 exposed controls and F_2 sheltered controls were statistically significant (DFA, $P < 0.001$). Snails of exposed-site ancestry developed more globular shells with relatively wider apertures than snails of sheltered-site ancestry, the difference being less pronounced in F_2 compared with F_1 snails.

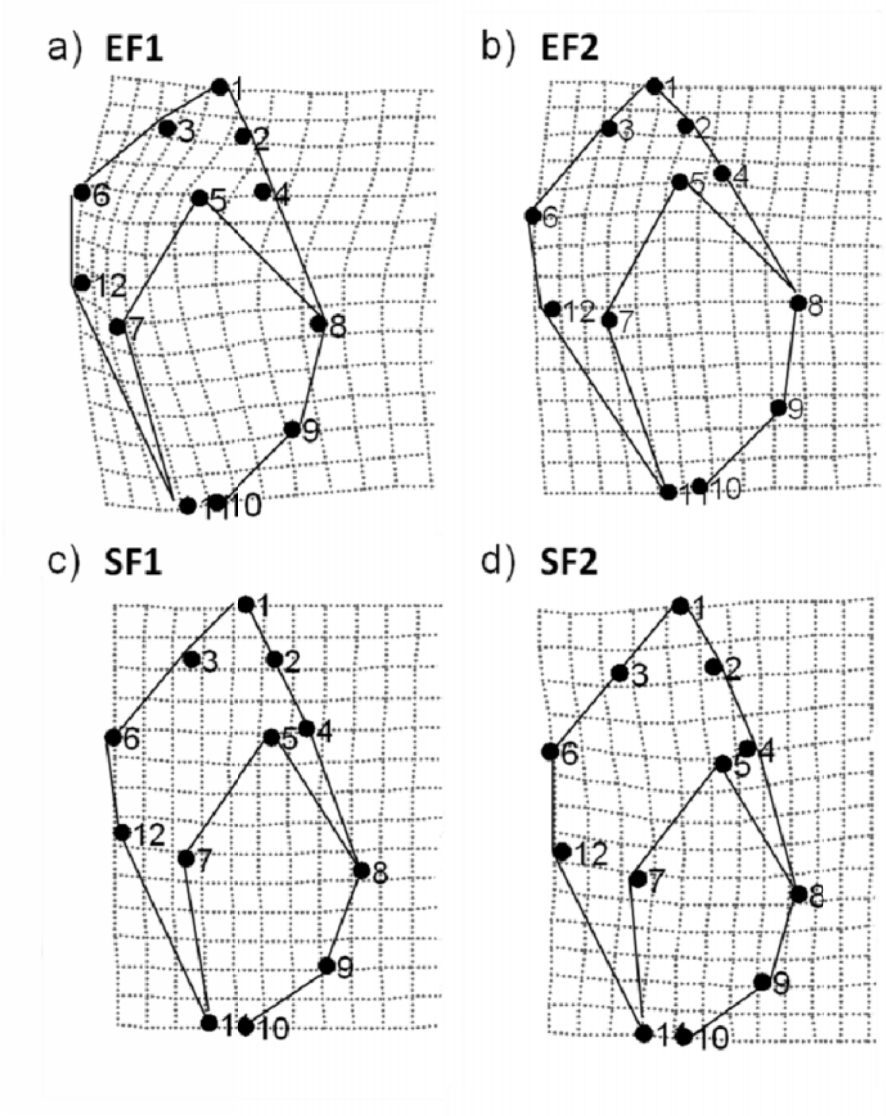


Figure 8. Common garden experiment: heritable variation in shell shape. Thin-plate spline deformations of shell shape of control snails: E: exposed; S: sheltered.

5.4.3. Karyotype and population genetics

Chromosome counts of 2n = 26-28 (please see appendix 5.2) were obtained for both the exposed- and sheltered-site populations, which also had identical 16S sequences and possessed only one mMDH haplotype (mMDH9). There was therefore no evidence of karyotype polymorphism.

All microsatellite loci were polymorphic for both populations. The number of alleles per population per locus ranged from 2 to 19, with a total number of 108 alleles in the global sample. The expected heterozygosity (H_E) per locus ranged from 0.332 to 0.892 and the observed heterozygosity (H_O) from 0.313 to 0.917 (Table 4). A test for concordance with HWE revealed deviations from HWE in locus Nlw2 and Nlw14 (Table 4). No evidence of linkage disequilibrium was observed between loci. Global F_{IS} was -0.0129 suggesting an excess of heterozygotes in the sampling areas. F_{ST} values per locus ranged from -0.0031 and 0.1491 and the global F_{ST} was 0.038 (P=0.001), revealing significant structuring between the two sampling sites (Table 4).

Table 4. Microsatellite analysis: genetic variability measures by locus for each population. Na: number of alleles found per locus; H_E: expected heterozygosity; H_O: observed heterozygosity; F_{IS}: standardised genetic variance within populations at each locus; F_{ST}: standardized genetic variance among populations at each locus; HWE: Hardy-Weinberg P values.

	Exposed				Sheltered				ALL	
	Na	Ho	He	HWE	Na	Ho	He	HWE	F_{IS}	F_{ST}
Nlw2	9	0.906	0.827	0.000	10	0.917	0.821	0.000	-0.101	0.021
Nlw3	9	0.760	0.774	0.595	7	0.583	0.515	0.059	-0.037	0.149
Nlw8	17	0.844	0.841	0.880	19	0.906	0.892	0.225	-0.005	0.007
Nlw11	12	0.792	0.790	0.640	13	0.792	0.830	0.269	0.028	0.077
Nlw14	13	0.719	0.850	0.000	14	0.917	0.848	0.003	0.042	0.024
Nlw17	12	0.802	0.867	0.004	15	0.885	0.878	0.003	0.038	0.008
Nlw21	2	0.344	0.359	0.776	3	0.313	0.332	0.622	0.055	0.008
Nlw25	4	0.448	0.378	0.324	4	0.448	0.476	0.673	-0.044	0.021
Nlw27	7	0.719	0.628	0.306	8	0.594	0.556	0.709	-0.103	0.006
All									-0.0129	0.0376

5.5. DISCUSSION

5.5.1. Phenotype and wave-exposure

5.5.1.1. Phenotypic plasticity

The present results concur with those of previous studies showing that N. lapillus from shores exposed to strong wave action have more globular shells with relatively larger, wider apertures than those from sheltered shores (e.g. Kitching et al. 1966; Kitching 1976; Crothers 1985). It was observed that variation in the above characters is attributable both to plasticity and inheritance. Etter (1988a) reported plastic and heritable variation in foot area of N. lapillus, itself correlated with the above shape characteristics, but that plasticity was confined to transplants from sheltered to exposed sites. Transplantation has also revealed non-reciprocal (asymmetrical) plasticity in littorinids (Etter 1988a; Trussell 1997; Yeap et al. 2001). Asymmetrical plasticity theoretically could be an adaptive response to the risk of error in environmentally cued acclimation (Palumbi 1984). Reduction in relative foot size during protracted calm periods on exposed sites would incur heavy mortality through dislodgement when more typical levels of wave action return (Etter 1988a; Trussell 1997). On the other hand, an increase in relative foot size during prolonged periods of wave action on sheltered sites would be less likely to reduce survivorship when normal conditions return. Similar adaptive interpretation of asymmetrical plasticity was made by Yeap et al. (2001) for the intertidal snail Nodilittorina australis, which readily develops from a striated, wave resistant morph to a nodular, faster cooling morph, but seldom vice versa.

Although the paucity of snails surviving transplantation from the sheltered to the exposed site weakened statistical comparison, the limited data augment those of Etter (1988a) in providing unequivocal evidence of plasticity in transplants from shelter to exposure. On the other hand and in contrast to Etter (1988a), data reveal plasticity in transplants from exposure to shelter. The discrepancy may reflect methodology: smaller initial size (≤12mm v. ≥14mm) and longer experimental duration (12mo v. 5mo) may have allowed greater scope for morphological divergence in the performed reciprocal transplant experiment, although supplementary data show that differentiation of shell

shape between snails of exposed and sheltered ancestry may already be discernible at shell lengths of 5-6mm corresponding to 4-5months of age (Fig. S2, appendix 5.1). It remains clear from the present study, however, that N. lapillus can show reciprocal phenotypic plasticity when transplanted between sites exposed to and sheltered from strong wave action, contrary to the cuing-error hypothesis of Palumbi (1984).

5.5.1.2. Heritable component of phenotypic variation

Morphological differentiation between N. lapillus from native exposed- and sheltered-site populations persisted through F_1 and F_2 generations reared under common garden conditions. Differentiation, however, became progressively weaker in successive generations, possibly due to combined influences of plasticity, reduced maternal effects, selection and genotype. Reduced differentiation between lineages in the F_1 generation compared with the field may have been attributable to plastic convergence, although lineages adapted to different suites of selection forces at exposed and sheltered sites may not necessarily perceive a common laboratory environment in the same way. Reduced differentiation could also result from selection. In all tanks, some F_1 and F_2 individuals showed poor growth and died before reaching maturity. Mortality could be attributable to genetic load, random factors such as inadequate energy supply at a critical stage, or selection imposed by experimental conditions. Although partial phenotypic convergence of lineages might suggest the common influence of selection, there is no apparent reason why the laboratory environment should have favoured intermediate phenotypes. Moreover, both lineages survived well and reproduced freely when brought from the field into the laboratory. Reduced lineage-differentiation between the F_1 and F_2 generations was probably due to diminished maternal effects (Marshall et al. 2008). Residual differentiation of lineages within the F_2 generation, however, strongly supports the conclusion that adaptive phenotypic differentiation between field populations is controlled genetically as well as by plasticity.

We have dealt above only with two sites representing extremes of the local wave-exposure gradient. Supplementary data that include intermediate sites not only support

present results but also show ranked correlation between plastic and heritable components of shell morphology and exposure to wave action (Fig. S3, appendix 5.1). Although selectively important physical variables associated with exposure to wave action may vary continuously among shores, however, other selection forces including risk of crab predation may be typified by more complex non-monotonic variation.

5.5.1.3. Population genetics

Previous genetic studies have shown N. lapillus to be capable of dispersal over several kilometres or more, probably by early juveniles drifting while attached to buoyant mucous threads or debris (Colson and Hughes 2004; Colson et al. 2006). Even at higher levels of gene flow however, local adaptation theoretically may occur if selection is sufficiently strong (Endler 1973; Koehn et al. 1980; Perez-Figueroa et al. 2005). For example, salinity conditions during early ontogeny exert selection strong enough to cause genetic differentiation among mixed migrant populations of herring Clupea harengus (Bekkevold et al. 2005). Depending on local hydrography and long-shore distribution of suitable habitat, N. lapillus may indeed show significant population differentiation on a scale of kilometres (Day and Bayne 1988; Kirby et al. 1997; Kirby 2000a; McInerney et al. 2009). Moreover, the genetic effect of differential selection among enclaves of N. lapillus occupying contrasted microhabitats may be evident even down to a scale of metres (Day 1990; Goudet et al. 1994; Guerra-Varela et al. 2009). N. lapillus aggregate in protective microhabitats in order to mate and spawn and although lacking specific homing behaviour snails tend to use a restricted number of spawning sites distributed within a radius of up to about 10m (RNH personal observation). In principle, therefore, enclaves could form reproductively semi-isolated units, depending on the scale of habitat-heterogeneity. Cliffs or reefs with large crevices and blocks presenting microhabitats exposed to and sheltered from major wave impact might provide a template for fine-scale population structuring as observed by Guerra-Varela et al. (2009). Because sites were selected for relatively uniform surfaces to facilitate recovery of marked snails, however, neither site provided heterogeneity on a scale likely to support reproductive enclaves. The general concordance with Hardy Weinberg equilibria in genotypic frequencies within

populations supports the above premise. Population genetic differentiation between the two sites showed significant divergence at microsatellite loci, suggesting populations were semi-isolated. The level of divergence (global $F_{ST,}$ 0.038 (P=0.001)) coincides with predicted levels for marine taxa lacking pelagic larvae (Palumbi 1994; Hellberg et al. 2002).

Although other factors cannot be ruled out, conditions favouring local adaptation are likely to arise from presumed limited gene flow among study populations (Johnson and Black 1998; Hoskin 2000). A recent synthesis of the scale of adaptive differentiation in marine invertebrates (Sanford and Kelly 2011) indicates that it can occur over a broad range of spatial scales, with marked adaptive variation occurring even at scales of less than 1 km. Such fine-scale heterogeneity has been most commonly observed in populations displaying varying tolerances to stress gradients associated with intertidal zonation (e.g. Janson 1982; Schmidt et al. 2000; Pardo and Johnson 2005). Indeed, the common-garden experiment indicated adaptive population genetic differentiation across such small spatial scales, supporting previous studies that identified genetic sweeps in enclaves of N. lapillus exposed to severe wave action (Carvajal-Rodriguez et al. 2006). It is possible, therefore, that observed genetic differentiation between local populations of N. lapillus is linked to bouts of intense selection under conditions of extreme wave action on exposed shores, selection for resistance to crab predation or desiccation on sheltered shores presumably being more stable across generations. As supported by theoretical studies (Slatkin 1973) and meta-analyses (Hollander 2008), evolution for plasticity rather than local adaptation is more prevalent in species with high gene flow. The presumed low rates of gene flow in direct developers such as N. lapillus, promotes fine-scale adaptive differentiation, especially where environmental gradients are likely to be relatively stable over evolutionary time such those driven by coastal bathymetry and the location of rocky headlands.

5.5.2. Co-gradient variation

5.5.2.1. Shell shape

The performed reciprocal transplant and common garden experiments indicated positive association between genetically based and environmentally induced differences in shell shape of N. lapillus from two sites with contrasted exposure to wave action, which we interpret as co-gradient variation (Levins 1968; Crispo 2008; Conover et al. 2009). Co-gradient variation in shell shape could be explained by at least three mutually non-exclusive hypotheses: (1) Genetically programmed allometry avoids delay in phenotypic expression through plasticity (Padilla and Adolph 1996; Kingsolver and Huey 1998). Plastic modification of shell morphology in N. lapillus is paced by growth of the shell on a time scale of months, which in the case of transplants from shelter to wave exposure bears the cost of reduced survivorship. The response-time model (Padilla and Adolph 1996), however, considers phenotypic variation that is determined either genetically or by plasticity, not in combination as found with N. lapillus. (2) Local genetic adaptation confers greater fitness than achievable by a general purpose genotype through plasticity alone. Incomplete phenotypic convergence of transplanted snails toward residential controls could be attributable to intrinsic limitation of plasticity or to ontogenetic constraint imposed by natal conditions. Ontogenetic constraint is likely to have been relatively unimportant, however, since the laboratory-reared juveniles developed similarly shaped shells to field-collected juveniles after transplantation. Furthermore, snails transplanted from the exposed to the sheltered site had ample scope for expressing plasticity as they grew from small juveniles to a size exceeding that of exposed site and even sheltered site controls. Apparently, therefore, co-gradient variation in shell shape of N. lapillus enables phenotypes to show greater adaptive variation than could be achieved by plasticity alone. (3) Plasticity enhances survivorship following dispersal into new selective environments (Wright 1931; Schlichting and Pigliucci 1998; Ghalambor et al. 2007). Acceleration of morphological differentiation by plasticity could be especially important if dispersal primarily involves younger hatchlings (Colson and Hughes 2004).

5.5.3. Counter-gradient variation

5.5.3.1. Growth rate

Exposed morphs of N. lapillus have previously been reported to have higher size-specific somatic growth rates but similar size-specific shell growth rates compared with sheltered morphs (Burrows and Hughes 1990; Kirby et al. 1994), whereas cumulative growth of body and shell are reported to be greater at sheltered sites than exposed (Menge 1978; Burrows and Hughes 1990). Accordingly, in the present reciprocal transplant experiment, sheltered-site controls reached larger sizes despite having grown more slowly than exposed-site controls, suggesting that they had experienced more time favourable for growth. On the other hand, snails transplanted from the exposed to the sheltered site grew faster and larger than sheltered-site residents, indicating counter-gradient variation with a heritable component promoting faster growth in snails of exposed-site provenance. There is no evidence of heritable differences in prey-handling ability between populations (Sanford and Worth 2010) and indeed this would be unlikely since both selected study sites present an abundance of barnacles that comprise the principal prey of local N. lapillus (Burrows and Hughes 1990). Snails transplanted from the sheltered to the exposed site grew more slowly and to a smaller size than in any other treatment, commensurable with the hypothesized slower potential growth rate of sheltered-site snails and shorter cumulative foraging time favourable for growth at the exposed site. Counter-gradient variation in growth rate has also been reported for Littorina obtusata, in which snails from exposed sites grew faster than those from sheltered under laboratory conditions of low flow velocity (Trussell 2002) and for L. saxatilis in which snails transplanted from high to low shore grew faster than low-shore residents (Pardo and Johnson 2005). The evolution of counter-gradient variation in growth rate of N. lapillus and Littorina spp. may perhaps be explained in terms of physiological compensation for constraint on foraging behaviour imposed by wave action on exposed shores (Trussell 2002) and by desiccation at high shore levels. Pardo and Johnson (2005), however, proposed that slower intrinsic growth rate of low-shore L. saxatilis avoids maladaptive effects of larger size in their native habitat, which offers richer feeding but also receives greater wave action that might select against larger shells

more vulnerable to dislodgement. Greater vulnerability to dislodgement has been invoked to explain the generally smaller maximum size of Nucella spp. at exposed sites (e.g. Kitching 1976, 1977; Burrows and Hughes 1991), but such interpretation is contradicted by the higher potential growth rate of exposed-site snails reported here.

5.5.3.2. Predator-induced shell morphology

The marked suppression of foraging behaviour and changes in shell morphology induced by crab-predation odour in the common garden experiment mirrored results obtained by Palmer (1990). Both studies found qualitatively similar changes in shell morphology of snails receiving crab-predation odour; shells became more globular, with relatively smaller, narrower apertures, the response being more pronounced in snails of exposed-site ancestry. Snails receiving crab-predation odour in this experiment also developed thicker shell lips as found in previous studies (Palmer 1990; Trussell and Nicklin 2002; Brookes and Rochette 2007; Edgell and Rochette 2008; Bourdeau 2010) and again the response was stronger in snails of exposed-site ancestry. Notably, snails transplanted from the exposed, crab-free site to the sheltered, crab-infested site developed relatively thicker shell lips than sheltered-site residents, indicating counter-gradient variation.

Starvation itself causes thickening rather than linear growth of the shell because calcium carbonate deposition continues independently of tissue growth (Appleton and Palmer 1988), leading to debate on whether shell thickening is the passive result of starvation caused by inhibited foraging behaviour or an active physiological response to perceived risk of predation. By controlling food supply as well as predation risk, Bourdeau (2010) showed that starvation resulting from constrained foraging behaviour was sufficient to explain the predator-induced shell thickening he observed. Greater shell-lip thickness of exposed-site snails observed herein therefore might be expected to have resulted from slower growth (above). Design of the performed common garden experiments confounded starvation with predation risk, but contrary to the passive-response hypothesis, linear growth of the shell was undiminished compared with controls, despite constrained foraging behaviour and reduced tissue growth. These

results therefore suggest that shell-lip thickening was a direct response to perceived risk of predation.

As would be expected from the induced response (above), populations of N. lapillus from sheltered sites have been reported with relatively thicker shells than from exposed sites (Kitching et al. 1966; Etter 1988b). The same is true of Littorina obtusata (Trussell and Etter 2001). However, the opposite is true of selected populations and of the neighbouring populations studied by Palmer (1990). On reaching sexual maturity N. lapillus continues thickening the shell lip but virtually ceases linear shell growth. Fig. 3c of Palmer (1990) indicates that although shells grew larger at his sheltered site, they did not achieve greater shell-lip thickness than shells from his exposed site. Similarly in the present experiment, although snails transplanted from the exposed to the sheltered site grew shells to a similar or even larger size than residents, they developed relatively thicker shell-lips. Moreover, in common garden, snails of exposed-site ancestry developed thicker shell-lips than those of sheltered-site ancestry, indicating a heritable component of variation in shell-lip thickness that is counter-gradient in terms of risk to crab predation. It remains unclear what environmental factor might favour thicker shells at exposed sites that are free of wave-tossed pebbles and boulders, such those studied here and by Palmer (1990). Exceptionally high shell thickness has been recorded in populations of N. lapillus occupying sites sheltered from severe wave action but experiencing strong tidal currents (Kitching et al. 1966; Currey and Hughes 1982). Such populations are characterized by exceptionally large adult size, implying that starvation cannot account for the greater relative shell mass. Strong water flow is common to tidal rips and heavy wave swash, perhaps selecting for shells with relatively greater mass and mechanical stability. Any such effect, however, would likely be miniscule compared with the effect of shell shape (Hughes and Taylor 1997) on drag reduction and foot size on grip (Etter 1988a).

Apparent costs of the induced defensive responses to crab-predation risk, mediated through constrained foraging and/or reallocation of energy, were expressed in reduced body growth and reproductive output, as found previously for Nucella spp. (Appleton and Palmer 1988; Palmer 1990; Rawlings 1994).

5.6. CONCLUSION

The experiments have indicated co-gradient plastic and genetic variation in shell shape corresponding to degree of exposure to wave action and counter-gradient variation in potential growth rate and predator-induced thickening of the shell lip. The above results support rapidly accumulating evidence that co-gradient variation usually, but by no means always, involves morphological traits, whereas counter-gradient variation tends to involve physiological or behavioural characters (Conover et al. 2009). Co-gradient variation in shell shape of N. lapillus will enhance resistance to dislodgement by waves at exposed sites and resistance to desiccation and crab predation at sheltered sites. Counter-gradient variation in growth rate will enhance growth at exposed sites by compensating lost foraging opportunity in times of heavy wave action. Counter-gradient variation of shell-lip thickening induced by perceived risk of crab predation is unexpected, but could enhance survival of individuals dispersing from exposed crab-free sites to sheltered crab-infested sites.

Comparable with the obtained results for N. lapillus, cogradient plastic and genetic components of phenotypic variation in shell shape have also been recorded in the infaunal bivalve Macoma balthica, which despite planktonic larval dispersal shows heritable variation in shell globosity between habitats (Luttikhuizen et al. 2003). M. balthica raised in a common-garden environment developed shell-shape variation similar to but not as pronounced as that of the parental forms. Luttikhizen et al. (2003) suggest that maintenance of local adaptation in M. balthica must depend on reduced gene flow, localized mating, or extremely strong selection, acting singly or in combination. The above similarity between species from two major taxa occupying widely different environments and exhibiting contrasting life histories suggests the combined influence of plasticity and inheritance on adaptive phenotypic variation may be quite general, maintained by a complex of factors whose detail depends on taxon and specific environment (Crispo 2008). In the case of N. lapillus, scales of dispersal and habitat variation are well documented but relatively little is known about local selection. It would be instructive to follow phenotypic and genetic variation among hatchlings through to adults in order to

assess the degree of selection across generations and the compromising effect of gene flow on genetic adaptation within sites.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through Ph.D. grant SFRH/BD/27711/2006 awarded to Sonia Pascoal. I would like to thank Dr Frederico Batista for helping with the setting up and maintenance of the common garden experiment, to Dr Niklas Tysklind for advice on microsatellite amplification and to Dr Delphine Lallias for guidance during paternity analysis.

5.7. REFERENCES

Adams, D. C., F. J. Rohlf, and D. E. Slice. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology 71:5-16.

Appleton, R. D., and A. R. Palmer. 1988. Water-Borne Stimuli Released by Predatory Crabs and Damaged Prey Induce More Predator-Resistant Shells in a Marine Gastropod. Proceedings of the National Academy of Sciences of the United States of America 85:4387-4391.

Bekkevold, D., C. Andre, T. G. Dahlgren, L. A. W. Clausen, E. Torstensen, H. Mosegaard, G. R. Carvalho, T. B. Christensen, E. Norlinder, and D. E. Ruzzante. 2005. Environmental correlates of population differentiation in Atlantic herring. Evolution 59:2656-2668.

Boulding, E. G., and T. K. Hay. 1993. Quantative genetics of shell form of an intertidal snail - constraints on short-term response to selection. Evolution 47:576-592.

Bourdeau, P. E. 2010. An inducible morphological defence is a passive by-product of behaviour in a marine snail. Proceedings of the Royal Society B-Biological Sciences 277:455-462.

Bradshaw, A. D. 1965. Evolutionary significance of phenotypic plasticity in plants. Advances in Genetics 13:115-155.

Brookes, J. I., and R. Rochette. 2007. Mechanism of a plastic phenotypic response: predator-induced shell thickening in the intertidal gastropod Littorina obtusata. Journal of Evolutionary Biology 20:1015-1027.

Burrows, M. T., and R. N. Hughes. 1990. Variation in growth and consuption among individuals and populations of dogwhelks, Nucella-lapillus - a link between foraging behavior and fitness. Journal of Animal Ecology 59:723-742.

Burrows, M. T., and R. N. Hughes. 1991. Optimal foraging decisions by dogwhelks, Nucella-lapillus (L) - influences of mortality risk and rate-constrained digestion. Functional Ecology 5:461-475.

Carvajal-Rodriguez, A., P. Conde-Padin, and E. Rolan-Alvarez. 2005. Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of littorina saxatilis. Journal of Molluscan Studies 71:313-318.

Carvajal-Rodriguez, A., J. Guerra-Varela, B. Fernandez, E. Rolan, and E. Rolan-Alvarez. 2006. An example of the application of geometric morphometric tools to the morphological diagnosis of two sibling species in Nassarius (Mollusca, Prosobranchia). Iberus 24:81-88.

Carvajal-Rodriguez, A., and M. G. Rodriguez. 2005. MODICOS: morphometric and distance computation software oriented for evolutionary studies. Online Journal of Bioinformatics 6:34-41.

Cavalcanti, M. J., L. R. Monteiro, and P. R. D. Lopes. 1999. Landmark-based morphometric analysis in selected species of serranid fishes (Perciformes : Teleostei). Zoological Studies 38:287-294.

Colson, I., J. Guerra-Varela, R. N. Hughes, and E. Rolan-Alvarez. 2006. Using molecular and quantitative variation for assessing genetic impacts on Nucella lapillus populations after local extinction and recolonization. Integrative Zoology 1:104-107.

Colson, I., and R. N. Hughes. 2004. Rapid recovery of genetic diversity of dogwhelk (Nucella lapillus L.) populations after local extinction and recolonization contradicts predictions from life-history characteristics. Molecular Ecology 13:2223-2233.

Conover, D. O., T. A. Duffy, and L. A. Hice. 2009. The Covariance between Genetic and Environmental Influences across Ecological Gradients Reassessing the Evolutionary Significance of Countergradient and Cogradient Variation. Year in Evolutionary Biology 2009 1168:100-129.

Coombs, V. A. 1973. Desiccation and age as factors in vertical distribution of dog-whelk, Nucella-lapillus. Journal of Zoology 171:57-66.

Crispo, E. 2008. Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. Journal of Evolutionary Biology 21:1460-1469.

Crothers, J. H. 1985. Dog-whelks an introduction to the biology of Nucella-lapillus. Field Studies 6:291-360.

Currey, J. D., and R. N. Hughes. 1982. Strength of the dogwhelk Nucella-lapillus and the winkle Littorina-littorea from different habitats. Journal of Animal Ecology 51:47-56.

Day, A. J. 1990. Microgeographic variation in allozyme frequencies in relation to the degree of exposure to wave action in the dogwelk Nucella lapillus (L) (Prosobranchia, Muricacea). Biological Journal of the Linnean Society 40:245-261.

Day, A. J., and B. L. Bayne. 1988. Allozyme variation of the dos-whelk Nucella lapillus (Prosobranchia, Muricacea) from the South West peninsula of England. Marine Biology 99:93-100.

DeWitt, T. J. 1998. Costs and limits of phenotypic plasticity: Tests with predator-induced morphology and life history in a freshwater snail. Journal of Evolutionary Biology 11:465-480.

Edgell, T. C., and R. Rochette. 2008. Differential snail predation by an exotic crab and the geography of shell-claw covariance in the northwest Atlantic. Evolution 62:1216-1228.

Endler, J. A. 1973. Gene flow and population differentiation. Science 179:243-250.

Endler, J. A. 1986. Monographs in population biology 21. Natural selection in the wild. Endler, J. a. Monographs in Population Biology, 21. Natural Selection in the Wild. Xiii+336p. Princeton University Press: Princeton, N.J., USA. Illus:XIII+336P.

Etter, R. J. 1988a. Asymmetrical Developmental Plasticity in an Intertidal Snail. Evolution 42:322-334.

Etter, R. J. 1988b. Physiological stress and color polymorphism in the intertidal snail Nucella-lapillus. Evolution 42:660-680.

Etter, R. J. 1996. The effect of wave action, prey type, and foraging time on growth of the predatory snail Nucella lapillus (L). Journal of Experimental Marine Biology and Ecology 196:341-356.

Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394-407.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Goudet, J., T. Demeeus, A. J. Day, and C. J. Gliddon. 1994. The different levels of populations structuring of the dogwhelk, Nucella lapillus, along the South Devon Coast. Genetics and Evolution of Aquatic Organisms:81-95.

Gould, S. J. 1966. Allometry and size in ontogeny and phylogeny. Biological Reviews of the Cambridge Philosophical Society 41:587-640.

Guerra-Varela, J., I. Colson, T. Backeljau, K. Breugelmans, R. N. Hughes, and E. Rolan-Alvarez. 2009. The evolutionary mechanism maintaining shell shape and molecular differentiation between two ecotypes of the dogwhelk Nucella lapillus. Evolutionary Ecology 23:261-280.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium series 41:95-98.

Hellberg, M. E., R. S. Burton, J. E. Neigel, and S. R. Palumbi. 2002. Genetic assessment of connectivity among marine populations. Bulletin of Marine Science 70:273-290.

Hollander, J. 2008. Testing the grain-size model for the evolution of phenotypic plasticity. Evolution 62:1381-1389.

Hori, M. 1993. Frequency-dependent natural-selection in the handedness of scale-eating cichlid fish. Science 260:216-219.

Hoskin, M. G. 2000. Effects of the East Australian Current on the genetic structure of a direct developing muricid snail (Bedeva hanleyi, Angas): variability within and among local populations. Biological Journal of the Linnean Society 69:245-262.

Hughes, R. N. 1972. Annual production of 2 Nova Scotian populations of Nucella-lapillus (L). Oecologia 8:356-370.

Hughes, R. N., and R. W. Elner. 1979. Tactics of a predator, Carcinus-maenas, and morphological responses of the prey, Nucella-lapillus. Journal of Animal Ecology 48:65-78.

Hughes, R. N., and M. J. Taylor. 1997. Genotype-environment interaction expressed in the foraging behaviour of dogwhelks, Nucella lapillus (L), under simulated environmental hazard. Proceedings of the Royal Society of London Series B-Biological Sciences 264:417-422.

Janson, K. 1982. Genetic and environmental effects on the growth-rate of Littorina saxatilis. Marine Biology 69:73-78.

Johannesson, B., and K. Johannesson. 1996. Population differences in behaviour and morphology in the snail Littorina saxatilis: Phenotypic plasticity or genetic differentiation? Journal of Zoology 240:475-493.

Johnson, M. S., and R. Black. 1998. Effects of isolation by distance and geographical discontinuity on genetic subdivision of Littoraria cingulata. Marine Biology 132:295-303.

Kawai, K., R. N. Hughes, and O. Takenaka. 2001. Isolation and characterization of microsatellite loci in the marine gastropod Nucella lapillus. Molecular Ecology Notes 1:270-272.

Kingsolver, J. G., and R. B. Huey. 1998. Evolutionary analyses of morphological and physiological plasticity in thermally variable environments. American Zoologist 38:545-560.

Kirby, R. R. 2000a. An ancient transpecific polymorphism shows extreme divergence in a multitrait cline in an intertidal snail (Nucella lapillus (L.)). Molecular Biology and Evolution 17:1816-1825.

Kirby, R. R. 2000b. Cloning and primary structure of putative cytosolic and mitochondrial malate dehydrogenase from the mollusc Nucella lapillus (L.). Gene 245:81-88.

Kirby, R. R. 2004. Genetic variation in a cline in a living intertidal snail arose in the neogastropoda over 100 million years ago. Journal of Molecular Evolution 58:97-105.

Kirby, R. R., B. L. Bayne, and R. J. Berry. 1994. Penotypic variation along a cline in allozyme and karyotype frequencies, and its relationship with habitat, in the dog-welk Nucella-lapillus, L. Biological Journal of the Linnean Society 53:255-275.

Kirby, R. R., R. J. Berry, and D. A. Powers. 1997. Variation in mitochondrial DNA in a cline of allele frequencies and shell phenotype in the dog-whelk Nucella lapillus (L.). Biological Journal of the Linnean Society 62:299-312.

Kitching, J. A. 1976. Distribution and changes in shell form of Thais spp. (Gastropoda) near Bamfield, BC. Journal of Experimental Marine Biology and Ecology 23:109-126.

Kitching, J. A. 1977. Shell form and niche occupation in Nucella-lapillus (L) (Gastropoda) Journal of Experimental Marine Biology and Ecology 26:275-287.

Kitching, J. A., L. Muntz, and F. J. Ebling. 1966. Ecology of Lough Ine. 15. Ecological significance of shell and body forms in Nucella. Journal of Animal Ecology 35:113-126.

Koehn, R. K., R. I. E. Newell, and F. Immermann. 1980. Maintenance of an aminopeptidase allele frequency cline by natural-selection. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 77:5385-5389.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, NJ.

Luttikhuizen, P. C., J. Drent, W. van Delden, and T. Piersma. 2003. Spatially structured genetic variation in a broadcast spawning bivalve: quantitative vs. molecular traits. Journal of Evolutionary Biology 16:260-272.

Marshall, D. J., R. M. Allen, and A. J. Crean. 2008. The ecological and evolutionary importance of maternal effects in the sea. Oceanography and Marine Biology: an Annual Review, Vol 46 46:203-250.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.

McInerney, C. E., A. L. Allcock, M. P. Johnson, and P. A. Prodohl. 2009. Understanding marine reserve function in a seascape genetics context: Nucella lapillus in Strangford Lough (Northern Ireland) as an example. Aquatic Biology 7:45-58.

Menge, B. A. 1978. Predation intensity in a rocky inter-tidal community - relation between predator foraging activity and environmental harshness. Oecologia 34:1-16.

Moran, N. A. 1992. The evolutionary maintenance of alternative phenotypes. American Naturalist 139:971-989.

Padilla, D. K., and S. C. Adolph. 1996. Plastic inducible morphologies are not always adaptive: The importance of time delays in a stochastic environment. Evolutionary Ecology 10:105-117.

Palmer, A. R. 1990. Effect of Crab Effluent and Scent of Damaged Conspecifics on Feeding, Growth, and Shell Morphology of the Atlantic Dogwhelk Nucella-Lapillus (L). Hydrobiologia 193:155-182.

Palumbi, S. R. 1984. Tactics of acclimation - morphological-changes of sponges in an unpredictable environment. Science 225:1478-1480.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics 25:547-572.

Pardo, L. M., and L. E. Johnson. 2005. Explaining variation in life-history traits: growth rate, size, and fecundity in a marine snail across an environmental gradient lacking predators. Marine Ecology-Progress Series 296:229-239.

Pascoe, P. L. 2006. Chromosomal polymorphism in the Atlantic dog-whelk, Nucella lapillus (Gastropoda : Muricidae): nomenclature, variation and biogeography. Biological Journal of the Linnean Society 87:195-210.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Perez-Figueroa, A., F. Cruz, A. Carvajal-Rodriguez, E. Rolan-Alvarez, and A. Caballero. 2005. The evolutionary forces maintaining a wild polymorphism of Littorina saxatilis: model selection by computer simulations. Journal of Evolutionary Biology 18:191-202.

Pigliucci, M. 2001. Syntheses in Ecology and Evolution. Phenotypic plasticity: Beyond nature and nuture.i-xvi, 1-328.

Rawlings, T. A. 1994. Effect of elevated predation risk on the metabolic-rate and spawning intensity of a rocky shore marine gastropod. Journal of Experimental Marine Biology and Ecology 181:67-79.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version-1.2) - Population-genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Rock, J., M. Eldridge, A. Champion, P. Johnston, and J. Joss. 1996. Karyotype and nuclear DNA content of the Australian lungfish, Neoceratodus forsteri (Ceratodidae: Dipnoi). Cytogenetics and Cell Genetics 73:187-189.

Rohlf, F. J. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Systematic Biology 47:147-158.

Rohlf, F. J., and F. L. Bookstein. 2003. Computing the uniform component of shape variation. Systematic Biology 52:66-69.

Sanford, E., and M. W. Kelly. 2011. Local adaptation in marine invertebrates Annual Reviews of Marine Science 3:509-535.

Sanford, E., and D. J. Worth. 2010. Local adaptation along a continuous coastline: Prey recruitment drives differentiation in a predatory snail. Ecology 91:891-901.

Scheiner, S. M. 1993. Genetics and evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 24:35-68.

Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: A reaction norm perspective. Phenotypic evolution: A reaction norm perspective:xii+387p.

Schmidt, P. S., M. D. Bertness, and D. M. Rand. 2000. Environmental heterogeneity and balancing selection in the acorn barnacle Semibalanus balanoides. Proceedings of the Royal Society of London Series B-Biological Sciences 267:379-384.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733-756.

Smith, T. B. 1993. Disruptive selection and the genetic-basis of bill size polymorphism in the African finch pyrenestes. Nature 363:618-620.

Stearns, S. C. 1989. The evolutionary significance of phenotypic plasticity - phenotypic sources of variation among organisms can be described by developmental switches and reaction norms Bioscience 39:436-445.

Svanback, R., M. Pineda-Krch, and M. Doebeli. 2009. Fluctuating Population Dynamics Promotes the Evolution of Phenotypic Plasticity. American Naturalist 174:176-189.

Thomas, M. L. H. 1986. A physically derived exposure index for marine shorelines. Ophelia 25:1-13.

Travis, J. 1994. Evaluating the adaptive role of morphological plasticity. Ecological morphology: Integrative organismal biology:99-122.

Trussell, G. C. 1997. Phenotypic plasticity in the foot size of an intertidal snail. Ecology 78:1033-1048.

Trussell, G. C. 2002. Evidence of countergradient variation in the growth of an intertidal snail in response to water velocity. Marine Ecology-Progress Series 243:123-131.

Trussell, G. C., and R. J. Etter. 2001. Integrating genetic and environmental forces that shape the evolution of geographic variation in a marine snail. Genetica 112:321-337.

Trussell, G. C., and M. O. Nicklin. 2002. Cue sensitivity, inducible defense, and trade-offs in a marine snail. Ecology 83:1635-1647.

Vadas, R. L., M. T. Burrows, and R. N. Hughes. 1994. Foraging strategies of dogwhelks, Nucella-lapillus (L) - interacting effects of age, diet and chemical cues to the threat of predation. Oecologia 100:439-450.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

Via, S., R. Gomulkiewicz, G. Dejong, S. M. Scheiner, C. D. Schlichting, and P. H. Vantienderen. 1995. Adaptive phenotypic plasticity - consensus and controversy. Trends in Ecology & Evolution 10:212-217.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57-63.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population-stucture. Evolution 38:1358-1370.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Yeap, K. L., R. Black, and M. S. Johnson. 2001. The complexity of phenotypic plasticity in the intertidal snail Nodilittorina australis. Biological Journal of the Linnean Society 72:63-76.

Zelditch, M., D. Swiderski, H. Sheets, and W. Fink. 2004. Geometric morphometrics for Biologists: A Primer. Elsevier Academic Press.

CHAPTER 6

General discussion

Understanding the molecular mechanisms of organismal adaptation to changing environments is a fundamental topic in modern evolutionary ecology (Stapley et al. 2010). Until recently, such studies have been mainly confined to classical model organisms (e.g Drosophila melanogaster, Caenorhabditis elegans). However, owing to recent advances in genomic technologies such as DNA sequencing (Margulies et al. 2005; Hudson 2008) and the development of downstream genomic tools (e.g. microarray, candidate gene identification, molecular marker development), the approach can now be applied feasibly to any species of interest. The development of genomics resources for non-model organisms, whose ecology and adaptation to different environments are particularly well understood, facilitates the study of a wider range of adaptive phenotypic traits than is possible with most model organisms (Wheat 2008). In particular, the production of gene expression tools (microarrays, digital transcriptomics), which provide invaluable genomic resources for a global overview of how the expression of every gene in the organism is responding to a particular stressor (Gracey 2007), is expected to solve hitherto intractable problems.

Application of the above emerging technologies to study gene-environment interactions in non-model organisms is therefore expected to contribute importantly to our understanding of how individuals and populations respond and adapt to environmental change. Consequently, in the present thesis this approach was applied to the ecologically well-studied species, Nucella lapillus, with the aim of investigating the response to anthropogenic environmental change (imposex) and response to natural environmental change (variation in shell morphology). Despite the extensive empirical and theoretical research on both endocrine disruption (imposex) and adaptive variation in shell morphology, the underlying mechanisms are still poorly understood. In order to enhance understanding of the above phenomena, the present study adopted a most needed interdisciplinary approach combining standard ecological and genetic methods (determination of imposex levels, reciprocal transplant and common garden experiments, morphometrics, and population genetics) with new genomic techniques (next-generation sequencing technologies applied to transcriptome and gene expression analysis using microarrays). The interdisciplinary approach and chosen methodology provided good

results and generated data yielding several new and valuable insights on the adaptive response in Nucella (a brief overview of the main findings is presented and discussed below).

Several similar studies addressing the new growing field of "ecological/adaptation genomics" applying emerging technologies in non-model species have recently emerged in the literature. For example, SNP restriction-site associated DNA (RAD) tags have been used to infer genome-wide patterns of parallel evolution in the three-spined stickleback Gasterosteus aculeatus (Hohenlohe et al. 2010); transcriptomic profiling (RNA-seq) using 454 Roche was used to identify genetic basis of phenotypic differentiation in the lake trout Salvelinus namaycush (Goetz et al. 2010); 454 Roche transcriptome sequencing was applied to detect outlier loci involved in local adaptation of the intertidal snail Littorina saxatilis (Galindo et al. 2010); combination of BAC sequencing and 454 Roche transcriptome sequencing was used to determine which and how many genes regulate variation in wing patterns of the butterfly Heliconius melpomene (Ferguson et al. 2010). Together, these examples, in addition to a growing number of other studies along with the present thesis work, are revealing new and important insights addressing some of the most fundamental questions in evolutionary genetics. Overarching foci include "i) Is adaptation the result of many loci of small effect or a few loci of large effect? ii) What type of genetic variation enables adaptation (i.e. point mutations, regulatory changes, inversions or gene duplications)? iii) What is the source of adaptive variation?" (Stapley et al. 2010).

6.1. RESPONSE TO TBT CONTAMINATION

To date, most information on the biological effects and mechanisms of endocrine disruptors (EDs) has been focused on vertebrates (Porte et al. 2006). However, studies on endocrine disruption in marine invertebrates have attracted some attention and are of great importance since invertebrates represent more than 95% of the known species in the animal kingdom and many taxa are of ecological relevance in the marine ecosystem

(Porte et al. 2006). At a population ecological level, the most thoroughly documented example of endocrine disruption in marine invertebrates is the TBT-induced imposex in gastropods. Despite recent advances in knowledge, however, the mechanism underlying imposex induction (Sternberg et al. 2010), hitherto have remained poorly understood, urgently warranting further exploration at a molecular level. Consequently, herein combinations of pyrosequencing and microarray technology were applied to disentangle the functional genomic mechanism of imposex using N. lapillus as a model organism.

Since at the outset of this study almost no genomic information was available for N. lapillus, we de novo sequenced the transcriptome using 454 Roche sequencing, generating a large EST resource that facilitated the reconstruction of a partial transcriptome with an estimated coverage of 18X. For gene annotation, searching the Nucella dataset against the NCBI and functional gene classification databases yielded a large proportion of sequences that have no similarity to published data, highlighting the lack of gene annotation in closely related species. However, the number of obtained matches (about 14% of the contigs) was in the same range of annotations of other non-model marine molluscs (e.g. Mytilus galloprovinciallis (Craft et al. 2010) and Laternula elliptica (Clark et al. 2010)) and indeed in non-model organisms in general (e.g. corals (Meyer et al. 2009) and butterflies (Vera et al. 2008)). Furthermore, in order to quantify levels of similarity between Nucella and a more close related species, more specific BLAST searches against the Lottia gigantea EST dataset (with no annotation) were performed and showed low levels of similarity between the two transcriptomes (approximately 12% of the contigs). However, it was within the normal range of homology assignments uncovered between other Gastropoda-Gastropoda comparisons (Feng et al. 2009) and it revealed over 4,000 putative orthologous genes between the two species.

Although the well-identified transcripts represented just a small proportion of the entire sequenced transcriptome, they document the first group of N. lapillus gene annotations, identifying a large set of genes of interest and make a substantial contribution towards the interpretation of Nucella genomic data (Fig. 1). Moreover, a significant resource of potential microsatellite molecular markers (Fig. 1) was produced, which following optimization and polymorphism testing may enable exploration of

diverse ecological and evolutionary questions concerning the intertidal environment (e.g. local adaptation) for gastropod species. Additionally, our data represent a valuable resource for comparative genome analysis and significantly augment transcriptomic knowledge within the class Gastropoda that hitherto has been poorly studied at the genomic level (chapter 2).

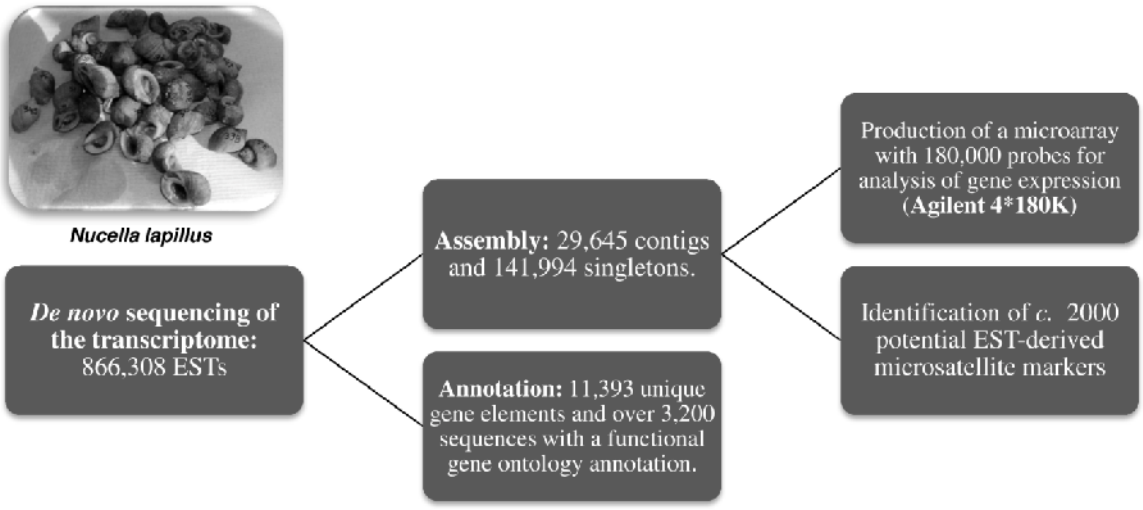


Figure 1. Schematic development of genomic resources for Nucella lapillus.

Three main hypotheses - steroid, neuroendocrine and retinoid - have been proposed to explain the chain of events leading to imposex development in gastropods. However, most of the proposed molecular targets have not been described in molluscs, and relationships between the proposed theories have been unclear (Castro et al. 2007; Sternberg et al. 2010). Taking advantage of advances and facilities in microarray design and production from Agilent (custom array design using eArray with no space limitation that allowed a novel approach by representing each gene with a sense and antisense probe), after transcriptome reconstruction and annotation, resulting data were used to generate a N. lapillus oligonucleotide array (Agilent 4*180,000-feature). This produced a highly functional genomic tool (Fig. 1) representing an unparallel genomic resource for N. lapillus, which has revealed new insights into the imposex mechanism itself and more

general responses to TBT. Microarray data were validated by qPCR and the same patterns of gene expression were obtained, confirming repeatability between techniques and hence the robustness of our findings. From the data, it is evident that the response to TBT contamination is a very complex phenomenon involving several molecular pathways and their interaction.

Microarray analyses for differential gene expression in response to TBT contamination support all three of the above hypotheses, indicating that they are not mutually exclusive. Additionally, from the transcription profile we identified new candidate TBT binders mapped by N. lapillus transcription factors and receptors that may be involved in the toxic pathway (e.g. RAR, ROR, Rev-Erba, AhR, ARNT, SRY and PPARγ). Several nuclear receptors, additional to steroid hormone receptors, have recently been identified as mediators of endocrine disruption (Iguchi and Katsu 2008). We therefore, anticipate that the new TBT binders that we have identified might be strong candidates as targets for the endocrine disruption mechanism. Impacts on the immune system, cell proliferation and apoptosis, DNA repair and tumour suppressors were evident and the possibility for a TBT-inhibited-transporter-based was also identified. Corroboratively, it has been proposed that the immune system may be susceptible to endocrine disruption (Swedenborg et al. 2009) and deregulation in transporter-based exchanges have been described both in programmed cell death in Drosophila (Yamada et al. 2008) and in gonad homeostasis in rats and humans (Suzuki et al. 2003). Possible effects of endocrine disruptors on functions other than reproduction, such as the immune response, may also help identifying specific targets for endocrine disruption in invertebrates (Porte et al. 2006). We predict that many further questions and studies may rise from our datasets, tools and findings (chapter 3).

Advances in sequencing technologies have facilitated access to the transcriptomes of non-model organisms, which can form the basis of microarray, or RNA-seq based interpretations of differentially expressed (DE) genes underpinning functional genomic responses. Many such studies yield a host of DE gene lists, accompanied by gene ontology classifications (e.g. Clark et al. 2010, Craft et al. 2010), but it is indeed rare to validate empirically the involvement of novel candidate genes and accompanying pathways by

inducing the focal mechanism in vivo in a non-model species. Consequently, in order to test and validate some of our molecular findings, we injected females with Rosiglitazone (a well known PPARγ ligand). Imposex was strongly induced (vas deferens development and penis growth at the same degree as for TBT itself) in the injected snails, and therefore we propose a new putative mechanism by which TBT triggers the imposex development in N. lapillus through the activation of the RXR:PPARγ heterodimer signalling pathway, hitherto not described in invertebrates, leading to the transcription of PPARγ target genes and generating a cascade of events that will ultimately cause the masculinisation of females. Although the RXR gene was clearly annotated within the present dataset, it was not delimited as a highly differentially expressed gene in any of the tissues and temporal sampling points during the imposex response. This might be explained by the simultaneous analysis of all the genes and experimental conditions in the microarray approach. While RXR presence in prosobranchs is well established, PPARγ and RAR-related homolog genes, among others, are novelties to this taxonomic group and considerably augment the understanding of nuclear receptors and transcription factors on N. lapillus imposex response (chapter 4).

At the beginning of this study, TBT exposures were performed to help elucidate the imposex mechanism and at the end, further exposures were made to test a new hypothesis. Our gene expression analysis yielded an improved understanding, not only of the imposex mechanism itself, but also of endocrine disruption in invertebrates more generally (Fig. 2). Briefly, imposex seems to result from a combination of steroid, neuroendocrine and retinoid (possibly through the RXR:PPARγ heterodimer) mechanisms along with an deregulation of energetics and transporter activity that are regulated by transcription factors and involve several lipophilic nuclear receptors. Despite there is still a long way to go in order to perceive the full cascade of events that lead to imposex development and to indeed fully explore the produced data; the identified new putative targets and mechanisms warrant future research regarding endocrine disruptions and highly contribute for an organismal-wide profile of response to TBT, previously not possible due to technological constraints and lack of genomic knowledge.

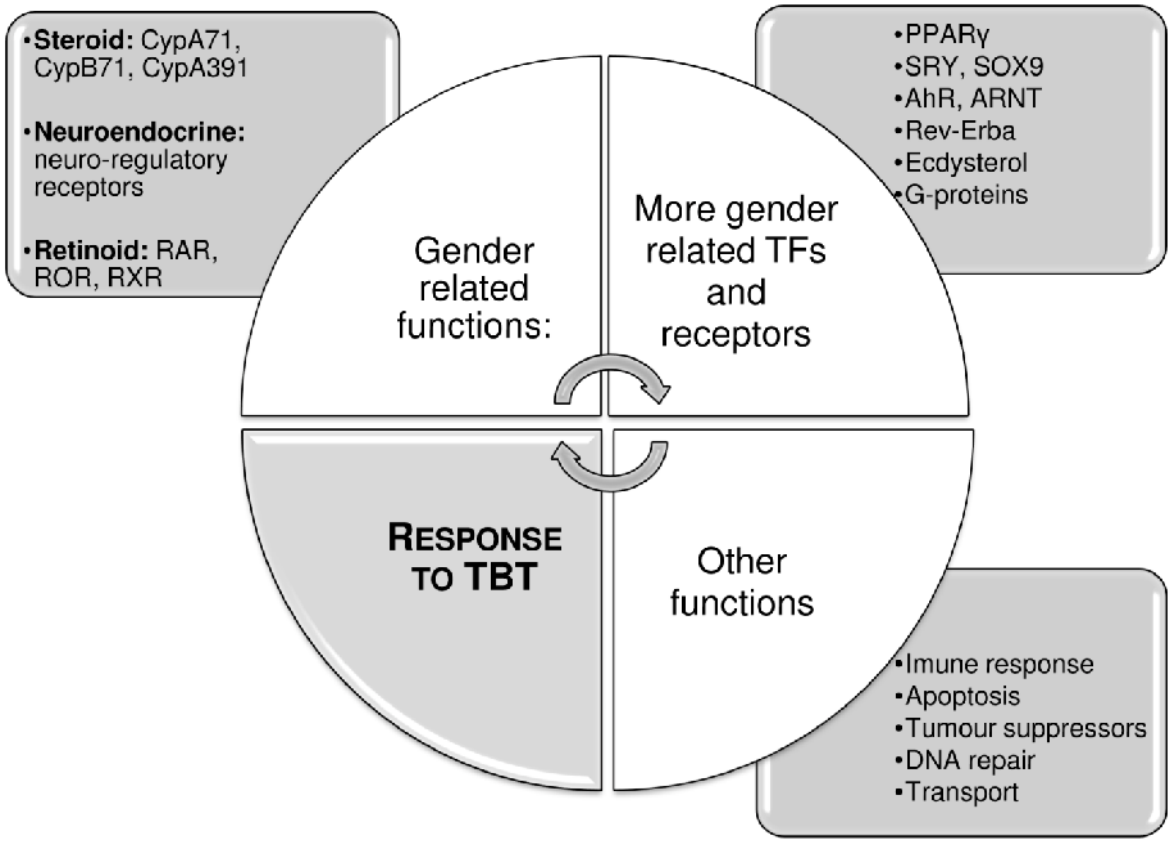


Figure 2. Gene expression analysis: main candidate TBT molecular targets for N. lapillus response to TBT contamination.

Regarding some conservation in transcription factors binding sites and pathways (Ren et al. 2001; Iguchi and Katsu 2008; and the present study), we anticipate that the present results may have broader application with regard to related pollutants and range of impacted taxa. Furthermore, to date, it has been reported that the environmental pollutant TBT is a ligand for RXR in Thais clavigera (Nishikawa et al. 2004; Horiguchi et al. 2007) and N. lapillus (Castro et al. 2007), and acts as a ligand for both the RXR and the PPARγ in the frog Xenopus laevis, mice and humans (Grun et al. 2006). Therefore, TBT, which induces imposex in marine snails and promotes adipogenesis in X. laevis and in mice, is an example of an environmental endocrine disrupter that promotes adverse effects through signalling pathways common to widely ranging taxa (Iguchi and Katsu 2008). Our new discovery that the PPARγ pathway may also be involved in the imposex induction in N. lapillus along with the identification of new potential targets of endocrine

disruption previously undescribed in gastropods, support the suggestion that the transcriptional mechanism for TBT action may be conserved across phyla.

Concerning technology, there is no doubt that the selected approach is leading edge and has generated huge amounts of data that revealed important findings for the TBT response and will also constitute a good starting point to address further questions using N. lapillus and its ability to adapt to the heterogeneous intertidal environment (e.g. SNPs development, mutations identification, gene structural variation (insertions, deletions, duplications), (Stapley et al. 2010)). However, promising advances in sequencing technologies with the soon expected release of the third generation sequencing, are presumed to produce cheaper, faster and more accurate sequencing with longer reads (Stapley et al. 2010). Similarly, recent advances in other transcriptome profiling approaches, RNA-Seq, other than the existing microarray technology, are offering several advantages such as very reduced, if any, background signals and absence of previous genomic knowledge that are particularly attractive for non-model organisms with genomic sequences that are yet to be determined (Wang et al. 2009). We therefore expect that such approaches will soon become popular and they will facilitate the further exploration of numerous biological questions primarily contributing for the understanding of how individuals adapt to changing environments.

Overall, this section of the thesis presents pioneering work using combinations of emerging molecular technologies to study imposex induction in N. lapillus; an ecologically well studied species for which there was almost no previous molecular data, producing large molecular datasets and genomic tools (Fig. 1.) crucial for gene-environment research. The transcriptomic data and functional genomic tools produced in this thesis provide a valuable resource for elucidating the functional genomic basis of the imposex mechanism and we anticipate that this resource will greatly facilitate exploration of environmental and evolutionary responses of N. lapillus, a sentinel organism, to challenges of the intertidal environment.

6.2. RESPONSE TO WAVE ACTION AND CRAB PREDATION

The ability of lineages and individuals to adapt to a particular habitat is a crucial topic in evolutionary biology, the exploitation of new niches being an important component of the speciation process (Etter 1988a; Nussey et al. 2007). The marine rocky intertidal zone is one of the most intrinsically variable habitats on earth (Gracey et al. 2008) and so provides a good "platform" for experimental exploration of adaptation to environment change.

Phenotypic plasticity of organisms in response to environmental variability may crucially influence the direction and rate of evolution (Trussell and Etter 2001; Pigliucci 2005) and is, therefore, a well developed concept in ecology and evolutionary biology. Plasticity is currently seen as one of the prime mechanisms by which organisms can respond adaptively to environmental change (Svanback et al. 2009; Ellers and Stuefer 2010), yet despite extensive empirical and theoretical research, the causes and consequences of plasticity are still poorly understood.

It has long been recognized that phenotypic variation reflects both genetic and environmental influences (Bradshaw 1965; Etter 1988a), and it has been suggested that genetic variation is favoured in stable environments (Hori 1993; Smith 1993), whereas phenotypic plasticity is favoured in unstable and fluctuating environments (Stearns 1989; Scheiner 1993; Svanback et al. 2009). The gastropod-wave exposure system that presents a strong correlation between phenotypes and distinct environmental gradient has been widely used for such studies (Kitching et al. 1966; Trussell and Etter 2001). Despite common trends, the results have varied sufficiently to require clarification.

The present study contributes to understanding of the evolution of adaptive character traits by combining common garden and reciprocal transplant experiments, complemented with genetic analysis, to assess the contributions of genotype, plasticity and their interaction to phenotypic expression in the ecologically well-studied intertidal gastropod Nucella lapillus in relation to the contrasted selection regimes on sheltered and exposed shores (Etter 1988a, 1988b; Kirby 2000; Guerra-Varela et al. 2009). Data mainly concur with previous findings; however, it also revealed new observations that augment the debate on shell shape phenotypic plasticity (Fig. 3). Overall, data show evidence of

both co- and counter-gradient variation in heritable and plastic traits and support that shell shape variation in N. lapillus is attributable both to plasticity and inheritance, which we have interpreted in the context of spatial scales of dispersal and habitat variation. Combining assessments of reciprocal transplants and common garden experiments in the same study is still quite rare and it greatly contributed for a better understanding of the variation in N. lapillus shell shape in response to wave action and crab predation.

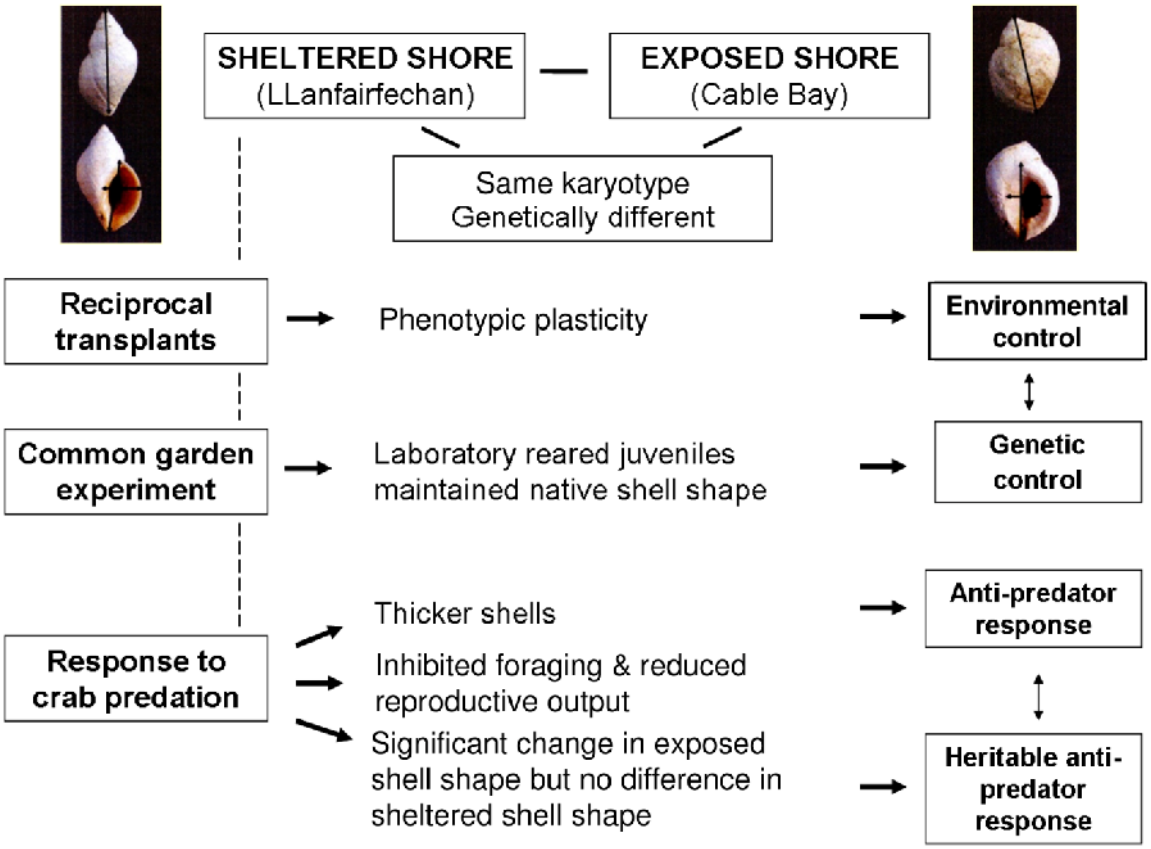


Figure 3. Common garden and reciprocal transplants results overview.

Previously, asymmetric phenotypic plasticity in N. lapillus, where plasticity is confined to transplants from sheltered to exposed sites, has been documented (Etter 1988a). However, our data showed that N. lapillus can also show phenotypic plasticity when transplanted from exposed to sheltered shores, highlighting a reciprocal phenotypic

plasticity between sites exposed to different degrees of wave action. Microsatellite analyses showed that the selected populations are genetically different (semi-isolated), with levels of divergence within the range of other marine taxa lacking pelagic larvae (Palumbi 1994; Hellberg et al. 2002), but no karyotype polymorphism was observed. Limited gene flow among populations is expected to favour local adaptation (Johnson and Black 1998; Hoskin 2000) and, indeed, the common-garden experiment indicated adaptive population genetic differentiation in the studied populations. Common garden experiments allowed us to follow F_1 and F_2 generations that provided key information on the contribution of the genetic component to phenotypic differentiation. Parental analyses were applied at this stage and the selected molecular markers revealed useful for such strategy, which as far as we know were not applied to N. lapillus before. Together, the data strongly support the conclusion that adaptive phenotypic differentiation between field populations is controlled genetically as well as by plasticity. Common garden data also contributed for the controversy of whether shell thickness results directly from risk to crab predation or through a passive result of starvation (Bourdeau 2010). Despite constrained foraging behaviour and reduced tissue growth, linear growth of the shell was undiminished compared with controls, and so, here we suggest that shell-lip thickness is a direct response to perceived risk of predation.

Furthermore, the obtained data suggest co-gradient variation in N. lapillus shell shape since a positive association between genetically based and environmentally induced differences was observed. Additionally, countergradient variation on growth rate and predator-induced changes in shell shape and increased thickening of the shell lip (Levins 1968; Crispo 2008; Conover et al. 2009) was also revealed. These findings support previous observations that co-gradient variation usually impacts morphology, whereas countergradient variation tends to affect physiology or behaviour (Conover et al. 2009). Co-gradient plastic and genetic components of phenotypic variation in shell shape documented here in N. lapillus have been also observed in species with planktonic larval dispersal (Luttikhuizen et al. 2003). We therefore suggest that the combined influence of plasticity and inheritance on adaptive phenotypic variation may be quite general and

should be maintained by a complex of factors whose detail depends on taxon and specific environment (Crispo 2008).

During life, an individual's phenotypic response to environmental cues need not be fixed. Often, the effect of environmental conditions is moderated by previous experience with such condition (e.g. heat and cold), (Ellers and Stuefer 2010). Similar conditioning can be observed for behavioural responses which can be modified through learning (Smid et al. 2007) or for immune responses, which are induced by previous contact with a pathogen (Schmid-Hempel 2005). Similarly, N. lapillus shell shape variation and thickening reflected the previous contact to risk of crab predation, abundant in sheltered shores and rare or inexistent in the exposed shore, which we have interpreted as a heritable anti-predator response. Similar traits have been documented in Littorina (Edgell and Rochette 2008) and Nucella lamellosa (Edgell and Neufeld 2008) responding differently to native and introduced predators.

Commonly, studies on plasticity concentrate on individual species ignoring the fact that species are part of complex interaction networks, in which species interactions may be condition-dependent. Therefore, the integration of the concept of plasticity into multitrophic relationships such as food webs or ecological communities is needed (Ellers and Stuefer 2010). However, this can be a challenging task given the multitude of potential species interactions within communities, as well as the numerous individual traits that can show phenotypic plasticity (Ellers and Stuefer 2010). The gastropod-wave action system may be a good model to address such debate since reciprocal transplants in the intertidal facilitates the simultaneous acquisition of contributions of local communities, risks of predation and other stressors on the shell morphology of N. lapillus representative of sheltered and exposed shores that may be easily complemented with common-garden experiments controlling individual variables.

Recent advances in large scale gene expression technology are facilitating the study of plasticity from a molecular perspective, and the generation of such data is expected to answer long-standing questions about this widespread phenomenon (Aubin-Horth and Renn 2009). It was originally planned to use the oligoarray produced in the first section of the thesis (endocrine disruption, chapter 3) to investigate the molecular mechanism underlying phenotypic plasticity in N. lapillus, but due to time constraints we did not achieve this aim. However, tissue samples were collected periodically from the reciprocal transplants and common garden experiments and are preserved in RNA-later for future use, mainly aiming to identify genes underpinning adaptation in the intertidal and to contribute for a better understanding of the mechanisms of phenotypic plasticity.

Bridging the topics covered in the first and second parts of the thesis, there is tentative evidence for a relationship between imposex and morphological variation of the shell in N. lapillus (Son and Hughes 2000). Shell size increased significantly with increasing degree of imposex and the authors suggested that this trend is possibly caused by a diversion of energy allocation from reproduction to shell growth when reproductive effort is blocked or disturbed by imposex (Son and Hughes 2000). We did not consider this variable in our study, but since the levels of TBT contamination are decreasing in the sampling area and are indeed very low, we assume that it is not influencing our results – at least not at a significant extent as the main variables wave action and crab predation.

6.3. CONCLUSIONS

Major conclusions that can be drawn from the present work are as follows:

i) Response to TBT exposure:

- This study generated a large EST resource that facilitated the reconstruction and annotation of a partial transcriptome for N. lapillus, an ecologically well known species for which there was almost no previous molecular data;
- From the transcriptomic data, a microarray was successfully designed providing an unparalleled environmental genomic resource for N. lapillus and, furthermore, potential molecular markers were identified;
- A group of 617 potential candidate genes responding to TBT exposure were identified for functional genomic investigation;
- TBT seems to be a multi-site binding compound and the contribution of several causal pathways and their crosstalk is evident;
- Several new candidate TBT targets were identified from gene expression analysis;
- A new hypothesis, that imposex may be induced through the RXR:PPARγ pathway, was formulated, tested and supported;
- A commonality in signalling of endocrine disruption along taxa was strengthened by the present study;
- 454 Roche sequencing and microarray are powerful technologies for gene-environment studies using non-model organisms and provided substantial contribution towards the interpretation of Nucella genomic data knowledge and the response to TBT in particular.

ii) Response to wave action and crab predation:

- Shell morphometric analyses revealed plasticity of shell shape in reciprocal transplants, but also the partial retention of parental shape by F_2 in common garden controls, indicating co-gradient plastic and genetic variation;
- Crab-predation odour stimulated the production of thicker shell lips, with greater response in exposed-site snails indicative of countergradient variation, and influenced shell shape in exposed-site but not sheltered-site snails;
- Combination of reciprocal transplant and common garden approaches highly contributed for a better understanding of shell shape plasticity in N. lapillus.

6.4. FURTHER DIRECTIONS

- Generated molecular resources comprise powerful tools for gene-environment exploration in the intertidal and can be used for numerous studies regarding this environment and ecological model species;
- Microsatellite markers optimization is required;
- SNP development can be easily achieved from the developed dataset and in combination with the identified microsatellite markers constitute powerful molecular markers for population genetics and as tools to address ecological and evolutionary questions;
- Several potential targets for TBT responses were identified and deserve further exploration;
- Further exploration of the produced dataset is desirable and may be used for comparative studies;
- Some of the plasticity results validated previous theories, however some others diverged from original thoughts and consequently deserve further testing;

- Molecular mechanism underlying phenotypic plasticity and local adaptation in N. lapillus may be explored using the produced microarray and the collected samples.

6.5. REFERENCES

Aubin-Horth, N., and S. C. P. Renn. 2009. Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. Molecular Ecology 18:3763-3780.

Bourdeau, P. E. 2010. An inducible morphological defence is a passive by-product of behaviour in a marine snail. Proceedings of the Royal Society B-Biological Sciences 277:455-462.

Bradshaw, A. D. 1965. Evolutionary significance of phenotypic plasticity in plants. Advances in Genetics 13:115-155.

Castro, L. F. C., D. Lima, A. Machado, C. Melo, Y. Hiromori, J. Nishikawa, T. Nakanishi, M. A. Reis-Henriques, and M. M. Santos. 2007. Imposex induction is mediated through the Retinoid X Receptor signalling pathway in the neogastropod Nucella lapillus. Aquatic Toxicology 85:57-66.

Clark, M. S., M. A. S. Thorne, F. A. Vieira, J. C. R. Cardoso, D. M. Power, and L. S. Peck. 2010. Insights into shell deposition in the Antarctic bivalve Laternula elliptica: gene discovery in the mantle transcriptome using 454 pyrosequencing. Bmc Genomics 11.

Conover, D. O., T. A. Duffy, and L. A. Hice. 2009. The Covariance between Genetic and Environmental Influences across Ecological Gradients Reassessing the Evolutionary Significance of Countergradient and Cogradient Variation. Year in Evolutionary Biology 2009 1168:100-129.

Craft, J. A., J. A. Gilbert, B. Temperton, K. E. Dempsey, K. Ashelford, B. Tiwari, T. H. Hutchinson, and J. K. Chipman. 2010. Pyrosequencing of Mytilus galloprovincialis cDNAs: Tissue-Specific Expression Patterns. Plos One 5.

Crispo, E. 2008. Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. Journal of Evolutionary Biology 21:1460-1469.

Edgell, T. C., and C. J. Neufeld. 2008. Experimental evidence for latent developmental plasticity: intertidal whelks respond to a native but not an introduced predator. Biology Letters 4:385-387.

Edgell, T. C., and R. Rochette. 2008. Differential snail predation by an exotic crab and the geography of shell-claw covariance in the northwest Atlantic. Evolution 62:1216-1228.

Ellers, J., and J. F. Stuefer. 2010. Frontiers in phenotypic plasticity research: new questions about mechanisms, induced responses and ecological impacts. Evolutionary Ecology 24:523-526.

Etter, R. J. 1988a. Asymmetrical Developmental Plasticity in an Intertidal Snail. Evolution 42:322-334.

Etter, R. J. 1988b. Physiological stress and color polymorphism in the intertidal snail Nucella-lapillus. Evolution 42:660-680.

Feng, Z. P., Z. Zhang, R. E. van Kesteren, V. A. Straub, P. van Nierop, K. Jin, N. Nejatbakhsh, J. I. Goldberg, G. E. Spencer, M. S. Yeoman, W. Wildering, J. R. Coorssen, R. P. Croll, L. T. Buck, N. I. Syed, and A. B. Smit. 2009. Transcriptome analysis of the central nervous system of the mollusc Lymnaea stagnalis. Bmc Genomics 10.

Ferguson, L., S. Lee, N. Chamberlain, N. Nadeau, M. Joron, S. Baxter, P. Wilkinson, A. Papanicolaou, S. Kumar, T. Kee, R. Clark, C. Davidson, R. Glithero, H. Beasley, H. Vogel, R. Ffrench-Constant, and C. Jiggins. 2010. Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. Molecular Ecology 19:240-254.

Galindo, J., J. W. Grahame, and R. K. Butlin. 2010. An EST-based genome scan using 454 sequencing in the marine snail Littorina saxatilis. Journal of Evolutionary Biology 23:2004-2016.

Goetz, F., D. Rosauer, S. Sitar, G. Goetz, C. Simchick, S. Roberts, R. Johnson, C. Murphy, C. Bronte, and S. Mackenzie. 2010. A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (Salvelinus namaycush). Molecular Ecology 19:176-196.

Gracey, A. Y. 2007. Interpreting physiological responses to environmental change through gene expression profiling. Journal of Experimental Biology 210:1584-1592.

Gracey, A. Y., M. L. Chaney, J. P. Boomhower, W. R. Tyburczy, K. Connor, and G. N. Somero. 2008. Rhythms of Gene Expression in a Fluctuating Intertidal Environment. Current Biology 18:1501-1507.

Grun, F., H. Watanabe, Z. Zamanian, L. Maeda, K. Arima, R. Cubacha, D. M. Gardiner, J. Kanno, T. Iguchi, and B. Blumberg. 2006. Endocrine-disrupting organotin compounds are potent inducers of adipogenesis in vertebrates. Molecular Endocrinology 20:2141-2155.

Guerra-Varela, J., I. Colson, T. Backeljau, K. Breugelmans, R. N. Hughes, and E. Rolan-Alvarez. 2009. The evolutionary mechanism maintaining shell shape and molecular differentiation between two ecotypes of the dogwhelk Nucella lapillus. Evolutionary Ecology 23:261-280.

Hellberg, M. E., R. S. Burton, J. E. Neigel, and S. R. Palumbi. 2002. Genetic assessment of connectivity among marine populations. Bulletin of Marine Science 70:273-290.

Hohenlohe, P., S. Bassham, P. Etter, N. Stiffler, E. Johnson, and W. Cresko. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. Plos Genetics 6.

Hori, M. 1993. Frequency-dependent natural-selection in the handedness of scale-eating cichlid fish. Science 260:216-219.

Horiguchi, T., T. Nishikawa, Y. Ohta, H. Shiraishi, and M. Morita. 2007. Retinoid X receptor gene expression and protein content in tissues of the rock shell Thais clavigera. Aquatic Toxicology 84:379-388.

Hoskin, M. G. 2000. Effects of the East Australian Current on the genetic structure of a direct developing muricid snail (Bedeva hanleyi, Angas): variability within and among local populations. Biological Journal of the Linnean Society 69:245-262.

Hudson, M. E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular Ecology Resources 8:3-17.

Iguchi, T., and Y. Katsu. 2008. Commonality in Signaling of Endocrine Disruption from Snail to Human. Bioscience 58:1061-1067.

Johannesson, B., and K. Johannesson. 1996. Population differences in behaviour and morphology in the snail Littorina saxatilis: Phenotypic plasticity or genetic differentiation? Journal of Zoology 240:475-493.

Johnson, M. S., and R. Black. 1998. Effects of isolation by distance and geographical discontinuity on genetic subdivision of Littoraria cingulata. Marine Biology 132:295-303.

Kirby, R. R. 2000. An ancient transpecific polymorphism shows extreme divergence in a multitrait cline in an intertidal snail (Nucella lapillus (L.)). Molecular Biology and Evolution 17:1816-1825.

Kitching, J. A., L. Muntz, and F. J. Ebling. 1966. Ecology of Lough Ine. 15. Ecological significance of shell and body forms in Nucella. Journal of Animal Ecology 35:113-126.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, NJ.

Luttikhuizen, P. C., J. Drent, W. van Delden, and T. Piersma. 2003. Spatially structured genetic variation in a broadcast spawning bivalve: quantitative vs. molecular traits. Journal of Evolutionary Biology 16:260-272.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Meyer, E., G. V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. K. Colbourne, B. L. Willis, and M. V. Matz. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. Bmc Genomics 10.

Nishikawa, J., S. Mamiya, T. Kanayama, T. Nishikawa, F. Shiraishi, and T. Horiguchi. 2004. Involvement of the retinoid X receptor in the development of imposex caused by organotins in gastropods. Environmental Science and Technology 38:6271-6276.

Nussey, D., A. Wilson, and J. Brommer. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20:831-844.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics 25:547-572.

Pigliucci, M. 2005. Evolution of phenotypic plasticity: where are we going now? Trends in Ecology & Evolution 20:481-486.

Porte, C., G. Janer, L. C. Lorusso, M. Ortiz-Zarragoitia, M. P. Cajaraville, M. C. Fossi, and L. Canesi. 2006. Endocrine disruptors in marine organisms: approaches and perspectives. Comparative Biochemistry and Physiology, part C 143:303-315.

Ren, L. I., H. H. Cheng, Y. Q. Guo, X. Huang, L. Liu, and R. J. Zhou. 2001. Evolutionary conservation of Dmrt gene family in amphibians, reptiles and birds. Chinese Science Bulletin 46:1992-1995.

Scheiner, S. M. 1993. Genetics and evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 24:35-68.

Schmid-Hempel, P. 2005. Evolutionary ecology of insect immune defenses. Annual Review Entomology 50:529-551.

Smid, H., G. Wang, T. Bukovinszky, J. Steidle, M. Bleeker, J. van Loon, and L. Vet. 2007. Species-specific acquisition and consolidation of long-term memory in parasitic wasps. Proceedings of the Royal Society B-Biological Sciences 274:1539-1546.

Smith, T. B. 1993. Disruptive selection and the genetic-basis of bill size polymorphism in the African finch pyrenestes. Nature 363:618-620.

Son, M. H., and R. N. Hughes. 2000. Relationship between imposex and morphological variation of the shell in Nucella lapillus (Gastropoda : thaididae). Estuarine Coastal and Shelf Science 50:599-606.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25:705-712.

Stearns, S. C. 1989. The evolutionary significance of phenotypic plasticity - phenotypic sources of variation among organisms can be described by developmental switches and reaction norms. Bioscience 39:436-445.

Sternberg, R. M., M. P. Gooding, A. K. Hotchkiss, and G. A. LeBlanc. 2010. Environmental-endocrine control of reproductive maturation in gastropods: implications for the mechanism of tributyltin-induced imposex in prosobranchs. Ecotoxicology 19:4-23.

Suzuki, T., T. Onogawa, N. Asano, H. Mizutamari, T. Mikkaichi, M. Tanemoto, M. Abe, F. Satoh, M. Unno, K. Nunoki, M. Suzuki, T. Hishinuma, J. Goto, T. Shimosegawa, S. Matsuno, S. Ito, and T. Abe. 2003. Identification and characterization of novel rat and human gonad-specific organic anion transporters. Molecular Endocrinology 17:1203-1215.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25:705-712.

Svanback, R., M. Pineda-Krch, and M. Doebeli. 2009. Fluctuating Population Dynamics Promotes the Evolution of Phenotypic Plasticity. American Naturalist 174:176-189.

Swedenborg, E., J. Ruegg, S. Makela, and I. Pongratz. 2009. Endocrine disruptive chemicals: mechanisms of action and involvement in metabolic disorders. Journal of Molecular Endocrinology 43:1-10.

Trussell, G. C., and R. J. Etter. 2001. Integrating genetic and environmental forces that shape the evolution of geographic variation in a marine snail. Genetica 112:321-337.

Vera, J. C., C. W. Wheat, H. W. Fescemyer, M. J. Frilander, D. L. Crawford, I. Hanski, and J. H. Marden. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636-1647.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57-63.

Wheat, C. W. 2008. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138:433-451.

Yamada, Y., K. Davis, and C. Coffman. 2008. Programmed cell death of primordial germ cells in Drosophila is regulated by p53 and the Outsiders monocarboxylate transporter. Development 135:207-216.

Supplementary information

CHAPTER 2.

- Appendix 2.1 - Bioinformatics details: in-house Perl scripts (Kevin Ashelford).

1. Data trimming

1) Created untrimmed fasta reads from original sff files:
In order to trim effectively it was necessary to work with raw untrimmed 454 read data (i.e., with 454 primers A and B still in place). For this, the following 454 software command was used:

```
sffinfo -s -n MID6_FW6YJSM01.sff > MID6_FW6YJSM01.untrimmed.fas
sffinfo -s -n MID7_FW6YJSM02.sff > MID7_FW6YJSM02.untrimmed.fas
```

2) Next created trimfiles from the raw untrimmed data:

Using a specially written script, trimfiles were created from each untrimmed fasta file. A trimfile is a simple space-delimited text file listing accession number and trim coordinates. Note that through experimentation we found that most adapter sequence could be found and trimmed away by searching for the last 11 bases of the adapter within the first and last 30 bases of the reads:

```
createTrimfile.pl -i MID7_FW6YJSM02.untrimmed.fas -5 CAACGCAGAGT -3 ACTCTGCGTTG -o MID7_FW6YJSM02.trimfile
createTrimfile.pl -i MID6_FW6YJSM01.untrimmed.fas -5 CAACGCAGAGT -3 ACTCTGCGTTG -o MID6_FW6YJSM01.trimfile
```

Perl script createTrimfile.pl:

```
#!/usr/bin/perl -w

use strict;
use warnings;
use Getopt::Std;
use File::Basename;

my %options;
getopts("i:5:3:o:", \%options);
my $infile          = $options{i} or &usage;
my $fivePrimeQuery  = $options{5} or &usage;
my $threePrimeQuery = $options{3} or &usage;
my $outfile         = $options{o} or &usage;
sub usage {
        die
                "Usage: " .
                basename($0) .
                "\n\t[-i 454 RAW read fasta infile (must be produced via sffinfo with -n flag)]" .
                "\n\t[-5 5' query sequence (e.g. AAGCAGTGGTATCAACGCAGAGT)]" .
                "\n\t[-3 3' query sequence (e.g. ACTCTGCGTTGATACCACTGCTT)]" .
                "\n\t[-o trim outfile]\n";
        }

# Read infile.
my $sequences = readFile($infile);
```

```
# Create outfile.
open(OUT, ">$outfile") or die "ERROR: Could not create $outfile.\n";

foreach my $id (sort keys %$sequences) {

	process(
		$id,				# Id
		$sequences->{$id},	# sequence
		*OUT,				# reference to filehandle
		$fivePrimeQuery,	# 5' search string
		$threePrimeQuery	# 3' search string.
		);

	}

close(OUT) or die "\nERROR: Could not close $outfile: $!";

################################################################################

sub readFile {

	my $infile = shift;

	my %sequences;
	my $header;
	my $sequence = "";
	open (IN, $infile) or die "ERROR: could not open $infile.\n";
	while (<IN>) {
		next if /(^#)|(^\s*$)/;
		if (/^>(\S+)/) {
			if (defined $header) {
				$sequences{$header} = $sequence;
				}
			$header = $1;
			$sequence = "";
			}
		elsif (/^(\w+)$/) {
			$sequence .= uc($1);
			$sequence =~ s/\s//g;
			}
		else {
			die "ERROR: Unexpected line '$_'.\n";
			}
		}
	close (IN) or die "ERROR: could not close $infile.\n";

	$sequences{$header} = $sequence;

	return \%sequences;

	} # End of method.

################################################################################
sub process {
```

```
	my $id                          = shift;
	my $sequence                    = shift;
	my $FILEHANDLE                  = shift;
	my $fivePrimeQuery   = shift;
	my $threePrimeQuery= shift;

	print "-->$id\n$sequence\n";

	my $start = 0;
	my $end = 0;

	if ($sequence =~ /^\w{0,30}$fivePrimeQuery/i) {
		$start = $+[0]+1;
		}

	if ($sequence =~ /$threePrimeQuery\w{0,30}/i) {
		$end = $-[0];
		}

	print $FILEHANDLE "$id $start-$end\n";

	} # End of method.
##############################################################################################
```

3) Next, again using 454 software, we created new sff files representing the trimmed data:

sfffile -t MID7_FW6YJSM02.trimfile -o MID7_FW6YJSM02.trimmed.sff MID7_FW6YJSM02.sff
sfffile -t MID6_FW6YJSM01.trimfile -o MID6_FW6YJSM01.trimmed.sff MID6_FW6YJSM01.sff

4) In order to check effectiveness of trimming procedure, we generated fasta files from the new sff files (this time trimmed versions).

sffinfo -s MID7_FW6YJSM02.trimmed.sff > MID7_FW6YJSM02.trimmed.fas
sffinfo -s MID6_FW6YJSM01.trimmed.sff > MID6_FW6YJSM01.trimmed.fas

In order to check that the adapter had been removed we used the following commands (here using fasta file MID6_FW6YJSM01.fas as an example):

Forward:
perl -ne 'print if /^\w{0,30}CAACGCAGAGT/i' MID6_FW6YJSM01.fas | wc -l

Reverse:
perl -ne 'print if /ACTCTGCGTTG\w{0,30}$/i' MID6_FW6YJSM01.fas | wc -l

Thus:

	Before trimming		After trimming	
run	forward	reverse	forward	reverse
MID7_FW6YJSM02	209,311	12,431	113	0
MID6_FW6YJSM01	516,168	41,984	280	0

2. BLAST

The full blast command is as follows:

1) For blastn:

blastall -p blastn -i 454AllContigs.fna -d /data/db/blastdb/nt-200906 -m 8 -a 4 -o blastn.22-07-2009.out -e 1e-6

In detail:

blastall is the program one invokes to run blast. The type of blast being dependent on the parameter choices made. Thus:

```
-p blastn           # -p selects blast program, in this case blastn.
-i 454AllContigs.fna  # -i identifies the input fasta file, in this case our
                        #   contigs.
-d nt-200906          # -d defines the precompiled blast database (here, the
                        #   full nucleotide sequence database downloaded from
                        #   NCBI and installed locally).
-m 8                  # -m optionally specifies the output format, where 8
                        #   signifies a tabular output.
-a 4                  # -a optionally specifies the number of processors to
                        #   use (in this case 4).
-o blastn.out         # -o defines the outfile to write to
-e 1e-10            # -e optionally defines the minimum e value allowed.
```

2) For blast x:

blastall -p blastx -i 454AllContigs.fna -d /data/db/blastdb/nr -m 8 -a 4 -o blastx.22-07-2009.out -e 1e-6

3) For gene finding blast:

blastall -p blastx -i <query input file> -d <database name> -F "m S" -U -f 14 -b 10000 -v 10000 -e 100 -m 8

In detail;
-F "m S" switches on soft masking. By default, blast ignores low-complexity regions which is normally a good idea but can lead to matches being broken up due to low-complexity regions. A compromise is to use this flag which switches on soft masking which means that low complexity regions are only masked during the word seeding phase and not during subsequent extension. This should result in more hits.

-U This switches on lower case filtering - i.e. ignores regions in the query sequence that are lower case (which with 454 contigs indicates lower quality base calls). You may or may not wish to switch this on.

-f 14 This recommendation is to increase the word threshold score used with the default BLOSUM62 matrix (default for blastx is 12). The purpose is to increase matching speed without reducing sensitivity too much - I'm not sure this is necessary and if speed is not an issue you again may which to ignore this setting. Basically the longer the word the fewer initial seeding matches.

-b 10000
-v 10000 These two flags effectively mean the same thing for -m 8 generated output - namely, they specify the maximum number of hits reported per query - the default is set low (I think to 250) so you may lose

matches due to certain protein families, with lots of representatives, swamping the output. Setting both to a high figure will ensure this i snot a problem.

-e 100 The recommendation is to set e high to capture low-scoring alignments.

3. BLAST filtering

Used script for filtering blast output. To run it from the command line, place script in the same directory as your blastn.out outfile and then use the following command:

perl screen_blast_output.pl blastn.out > bestlines.txt

This command runs the provided perl script, passing the file blastn.out to the script and capturing all stdout from script to an outfile called bestlines.txt. As written, the script takes a blast outfile as input and prints for each query the line with the highest alignment length. Script can be modified to screen for another parameter besides alignment length or a combination of parameters.

Perl script screen_blast_output.pl:

```
#!/usr/bin/perl -w

use strict;
use warnings;

my %bestLines;
my %bestAlignment;

# While loop reads through provided file, one line at a time...
while (<>) {

        # Regular expression matches alignment line, capturing all contained information.
         if
(/^(contig\d+)\t(\S+)\t(\d+\.{0,1}\d*)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\d+)\t(\S+)\s+(\d+\.{0,1}\d*
)$/) {

                # For clarity, informatively-named variables initialised from captured data:
                my $queryId = $1;
                my $subjectId = $2;
                my $percentIdentity = $3;
                my $alignmentLength = $4;
                my $mismatches = $5;
                my $gapOpenings = $6;
                my $queryStart = $7;
                my $queryEnd = $8;
                my $subjectStart = $9;
                my $subjectEnd = $10;
                my $eValue = $11;
                my $bitScore = $12;


                # NEW CODE:
```

```
                        # For current query id, checks whether current alignment length is greater
                        # than the previous best for this id.  If so store line, else ignore line.

                        # Specifically: for current query id, if no best alignment has yet to be stored
                        # or if best alignment is less than current alignment, store current alignment
                        # as best alignment and store current line as best line.

                        if (!exists $bestAlignment{$queryId} || $bestAlignment{$queryId} < $alignmentLength) {
                                $bestLines{$queryId} = $_;
                                $bestAlignment{$queryId} = $alignmentLength;
                                }

                        }

                # The above line should match all lines in blast outfile, but just in case, kill
                # the script with an appropriate message if an unexpected line is encountered.
                else {
                        die "ERROR: unrecognised line:\n$_\n";
                        }

                }


# Finally, work through each (sorted) query id in turn and print out best line.
foreach my $id (sort keys %bestLines) {
        print "$bestLines{$id}";
        }
```

4. Adding annotation to BLAST results

Take an –m 8 generated tabular blast output such as the ones produced above. Use perl script modify_blast.pl to append additional column containing definition information extracted from a separate database containing accessions and corresponding definitions.

perl modify_blast.pl -i blastn.out -r definitions_lookup_table.txt -o modified_blastn.out

perl script modify_blast.pl:

```
#!/usr/bin/perl -w

        use strict;
        use warnings;
        use Getopt::Std;
        use File::Basename;
        my %options;
        getopts ("i:o:r:", \%options);
        my $infile      = $options{i} or &usage;
        my $outfile     = $options{o} or &usage;
        my $reffile     = $options{r} or &usage;
        my %data;
        open (IN, $infile) or die "\nERROR: Could not open $infile - $!";
```

```
	while (<IN>) {
		chomp;
		if (/^(contig\d+)\t(\S+)/) {
	    my $subjectId = $2;

	    push(@{$data{$subjectId}}, $_);

		}


	else {

		die "\nERROR: Line not accounted for: $_.\n";

		}

	}

close (IN) or die "\nERROR: Could not close $infile - $!";
open (OUT, ">$outfile") or die "\nERROR: Could not create $outfile - $!";

open (REF, $reffile) or die "\nERROR: Could not open $reffile - $!";

while (<REF>) {

	chomp;

	if (/^(\S+)\s+(.+)$/) {

		my $subjectId = $1;
		foreach (@{$data{$subjectId}}) {
			print OUT "$_\t$2\n" if exists $data{$subjectId};


			}
		}


    else {
		die "\nERROR: Line not accounted for: $_.\n";
		}
}


close (REF) or die "\nERROR: Could not close $reffile - $!";
close (OUT) or die "\nERROR: Could not close $outfile - $!";
sub usage {

	die "\n" .  basename($0) . " [-i blast tabular output as infile] [-r reference file] [-o outfile]\n\n";


	}
```

- Appendix 2.2 - Summary statistics for all BLAST analysis of N. lapillus 454 Roche contigs and singletons.

Contigs	NCBI NT	NCBI NR	NCBI NR*	Lottia NT	Lottia NR	Lottia NR*
N significant hits with annotation	1519	4191	24588	296	3353	29584
mean % identity	87.65	41.85	30.57	85.74	51.39	29.72
Mean best hit alignment length	154.9	138.6	99.1	141.5	111.2	78.7
Mean mismatches	22.37	75.04	62.64	23.96	54.42	51.34
Mean gap openings	0.45	2.78	2.68	0.11	1.66	1.99
Mean query start	196.80	288.72	230.44	269.30	290.15	223.56
Mean query end	350.22	308.64	228.86	409.73	311.70	226.12
Mean e value	7.49E-08	8.35E-08	2.25E+01	5.76E-08	5.16E-08	1.48E+01
Mean bitscore	120.83	89.33	38.99	88.84	100.78	33.16
Singletons	NCBI NT	NCBI NR	NCBI NR*	Lottia NT	Lottia NR	Lottia NR*
N significant hits with annotation	2690	8578	82161	633	3997	132396
mean % identity	90.23	43.75	34.05	92.27	52.18	35.31
Mean best hit alignment length	96.44	86.29	72.24	59.48	72.07	49.78
Mean mismatches	10.11	46.76	44.88	6.59	34.90	32.08
Mean gap openings	0.76	1.34	1.80	0.19	0.87	1.10
Mean query start	95.48	166.42	147.86	130.36	158.85	117.94
Mean query end	189.79	170.88	144.08	188.75	164.71	115.45
Mean e value	8.24E-08	9.32E-08	2.54E+01	1.07E-07	7.07E-08	2.45E+01
Mean bitscore	95.04	67.38	35.04	62.80	71.10	25.83

* **$e=10^2$**

CHAPTER 3.

- Appendix 3.1: Candidate genes annotations from BLASTx searches.

cluster	Gene name	Genefinding annotation
1	contig14874	PREDICTED: similar to UDP glycosyltransferase 2 family, polypeptide A1 [Strongylocentrotus purpuratus]
1	contig09883_antisense	NA (hypothetical)
1	contig24044	PREDICTED: similar to pulmonary surfactant protein A [Monodelphis domestica]
1	contig03623_antisense	ATPase 2
1	contig15073	acetyl-CoA carboxylase [Homo sapiens]
2	FW6YJSM01A2X05_antisense	PREDICTED: similar to zinc finger protein [Ciona intestinalis]
2	FW6YJSM02RO3GQ_antisense	fimbrial biogenesis outer membrane usher protein [Escherichia coli O157:H7 str. EC4024]
2	FW6YJSM01CNC4D_antisense	diaminobutyrate
2	FW6YJSM02RJGDY	ammonium transporter [Rhodobacter sphaeroides ATCC 17025]
2	FW6YJSM01EQ8BD	phospho-N-acetylmuramoyl-pentapeptide-transferase [Borrelia hermsii DAH]
2	contig03398	gluconate permease, putative [Burkholderia mallei ATCC 23344]
2	contig12665_antisense	PREDICTED: similar to double homeobox, 4 [Pan troglodytes]
2	contig23054	carbamoyl-phosphate synthase small subunit [Methylophilales bacterium HTCC2181]
2	contig10249_antisense	Zinc finger, RING-type; RINGv [Medicago truncatula]
2	contig03398	REPEATED
2	FW6YJSM01AICIJ	PREDICTED: similar to synaptobrevin [Acyrthosiphon pisum]
2	contig03379	amino acid permease, unknown 10 [Schizosaccharomyces pombe 972h-]
2	contig18942	exodeoxyribonuclease III [Burkholderia thailandensis E264]
2	FW6YJSM02RZD8U	similar to squamous cell carcinoma antigen recognised by T cells [Xenopus laevis]
2	contig26976_antisense	NA (hypothetical)
2	FW6YJSM01D6LL8	NADH dehydrogenase subunit 5 [Corvus frugilegus]
2	FW6YJSM01D48VA_antisense	integral membrane protein [Theileria annulata]
2	contig03379	REPEATED
9	contig03588_antisense	cathepsin L-like tick cysteine proteinase B [Haemaphysalis longicornis]
9	contig26465_antisense	PREDICTED: similar to Col protein [Strongylocentrotus purpuratus]
9	FW6YJSM01BHKY5_antisense	still life, putative [Aedes aegypti]
9	FW6YJSM01A5SZ9_antisense	GCN1; translational activator of GCN4 [Pichia stipitis CBS 6054]
9	FW6YJSM01ANQ64_antisense	HAE1 family efflux transporter [Candidatus Methanoregula boonei 6A8]
9	contig12995	CCAAT-binding transcription factor [Plasmodium knowlesi strain H]
9	FW6YJSM02PLQQL_antisense	multi-sensor hybrid histidine kinase [Desulfonatronospira thiodismutans ASO3-1]
9	contig17362_antisense	cartilage matrix protein
9	contig17531_antisense	putative ATP-dependent RNA helicase [Vibrio splendidus 12B01]
9	contig05232	keratin associated protein 10-2 [Bos taurus]
9	FW6YJSM01AWN33	FAD linked oxidase domain-containing protein [Actinobacillus succinogenes 130Z]
9	FW6YJSM01EX1DN_antisense	PE-PGRS family protein [Mycobacterium marinum M]

9	contig19578_antisense	nucleotidyl transferase [Ralstonia phage RSL1]
9	FW6YJSM01BOOWR	calpain-like cysteine peptidase [Trypanosoma cruzi strain CL Brener]
9	FW6YJSM01DLAJD	delta-aminolevulinic acid dehydratase [Zymomonas mobilis subsp. mobilis ZM4]
9	contig08106_antisense	chemokine receptor CXCR4 [Mandrillus leucophaeus]
9	FW6YJSM02P79DZ_antisense	PREDICTED: similar to putative elicitin protein RAM6 [Gallus gallus]
9	FW6YJSM01AR0E1_antisense	PREDICTED: similar to Xeroderma pigmentosum group A-like CG6358-PA [Apis mellifera]
9	FW6YJSM01DX0ST	protein-export membrane protein SecD [Polynucleobacter necessarius subsp. asymbioticus QLW-P1DMWA-1]
9	contig08915	type VI secretion protein, family [Burkholderia pseudomallei 576]
9	FW6YJSM01DIM0W_antisense	T-cell receptor beta chain [Homo sapiens]
9	contig11274	tetratricopeptide TPR_2 [Trichodesmium erythraeum IMS101]
9	FW6YJSM01ESAFH_antisense	olfactory receptor, family 10, subfamily Q, member 1 [Homo sapiens]
9	FW6YJSM01CTB1X_antisense	PREDICTED: similar to cutA divalent cation tolerance homolog [Sus scrofa]
9	FW6YJSM01D40G3_antisense	membrane protein [Capnocytophaga sputigena Capno]
9	contig14887	putative dissimilatory membrane-bound nitrate reductase [uncultured bacterium]
9	FW6YJSM02TX2FM_antisense	ABC transporter related [Pseudomonas putida F1]
9	contig21520	phosphatidylinositol-4-phosphate 5-kinase family protein [Arabidopsis thaliana]
9	FW6YJSM01BDSKQ	PREDICTED: similar to G protein-coupled receptor 20 [Danio rerio]
9	FW6YJSM01E3OCS_antisense	PREDICTED: similar to Col protein [Strongylocentrotus purpuratus]
9	FW6YJSM01EXJ9S	PREDICTED: similar to Scavenger receptor class F member 2 precursor (Scavenger receptor expressed by endothelial cells 2 protein) (SREC-II) (SRECRP-1) [Bos taurus]
9	contig14410_antisense	histidine kinase [Delftia acidovorans SPH-1]
9	FW6YJSM02TWATK	PREDICTED: similar to zinc finger protein 544 [Danio rerio]
9	FW6YJSM02QNAGV	PREDICTED: similar to growth-associated polypeptide [Gallus gallus]
9	contig12904	F-box family protein [Arabidopsis thaliana]
9	contig08375	major facilitator superfamily MFS_1 [Verrucomicrobium spinosum DSM 4136]
9	FW6YJSM01DVCPN	PREDICTED: ephrin receptor EphB3 [Pan troglodytes]
9	contig09406	RNA binding protein [Homo sapiens]
9	contig19561_antisense	immunoglobulin light chain [Acipenser baerii]
9	FW6YJSM01ARS1Z	unnamed protein product [Tetraodon nigroviridis]
9	contig12249	prephenate dehydratase, putative [Ricinus communis]
9	FW6YJSM01BNQTA	pol-like protein [Ciona intestinalis]
9	FW6YJSM02THC1G	Bkm-like sex-determining region hypothetical protein CS314 - fruit fly (Drosophila melanogaster) (fragment)
9	FW6YJSM02TL8LC_antisense	WD domain containing protein [Entamoeba histolytica HM-1:IMSS]
9	FW6YJSM02TUY4S	2-oxoglutarate dehydrogenase E1 component [Thalassiosira pseudonana CCMP1335]
9	FW6YJSM02QAZU3	putative branched chain amino acid ABC transporter substrate-binding protein [Symbiobacterium thermophilum IAM 14863]
9	FW6YJSM02T0YTC	maturase K [Carlephyton glaucophyllum]
9	FW6YJSM02PO750_antisense	putative transcriptional regulatory protein [Croceibacter atlanticus HTCC2559]
9	FW6YJSM01DO7ZA_antisense	glutamine-binding periplasmic protein/glutamine transport system permease protein [Campylobacter concisus 13826]

9	contig16506	regulator of kdp operon (transcriptional effector) [Acinetobacter sp. ADP1]
9	FW6YJSM02P40HD_antisense	EGF-like module containing, mucin-like, hormone receptor-like 1 [Rattus norvegicus]
9	contig10272_antisense	putative Phosphoglycolate phosphatase (PGPase) (PGP) [Candidatus Cloacamonas acidaminovorans]
9	contig03910	tRNA pseudouridine synthase D TruD [Methanococcus aeolicus Nankai-3]
9	FW6YJSM01BLHEM_antisense	putative branched chain amino acid ABC transporter substrate-binding protein [Symbiobacterium thermophilum IAM 14863]
9	FW6YJSM01AKOSL	PREDICTED: similar to death receptor interacting protein [Monodelphis domestica]
9	FW6YJSM01CEUNC_antisense	Nitrogenase [Geobacter sp. FRC-32]
9	FW6YJSM02PY7SM_antisense	hypothetical protein Plav_2289 [Parvibaculum lavamentivorans DS-1]
9	contig17542	RNA polymerase II associated Paf1 complex (predicted) [Schizosaccharomyces pombe]
9	FW6YJSM02TJ345_antisense	PREDICTED: similar to ADAM metallopeptidase with thrombospondin type 1 motif, 12 preproprotein [Bos taurus]
9	FW6YJSM01A1WDT_antisense	iduronate sulfatase
9	FW6YJSM02ROOSP	PREDICTED: similar to G-type lysozyme [Ciona intestinalis]
9	FW6YJSM01A235L	PREDICTED: similar to hydroxysteroid (17-beta) dehydrogenase 2 [Canis familiaris]
9	FW6YJSM01BQB1H_antisense	monosaccharid transporter [Nicotiana tabacum]
9	FW6YJSM01BQKIV_antisense	DNA primase polypeptide 2 [Gallus gallus]
9	contig01653_antisense	zinc finger protein 64 isoform d [Homo sapiens]
9	FW6YJSM02S5236_antisense	DNA helicase/exodeoxyribonuclease V, subunit B [Magnetococcus sp. MC-1]
9	contig10756	PREDICTED: suppression of tumorigenicity 5 isoform 1 [Pan troglodytes]
9	FW6YJSM01DBTHM	nuclear acid binding protein, putative [Ricinus communis]
9	FW6YJSM02RJ4R3	DNA polymerase III subunit delta [Helicobacter acinonychis str. Sheeba]
9	FW6YJSM01DATMF_antisense	nuclear acid binding protein, putative [Ricinus communis]
9	contig23302_antisense	putative inner membrane protein [Vibrio shilonii AK1]
9	FW6YJSM01AJTUH_antisense	PREDICTED: similar to cysteine and tyrosine-rich protein 1 [Ornithorhynchus anatinus]
9	FW6YJSM01B89IB	PE-PGRS family protein [Mycobacterium marinum M]
9	contig00743_antisense	DEAD box helicase [Plasmodium falciparum 3D7]
9	contig13072	lacto-N-biosidase [Lactococcus lactis subsp. lactis Il1403]
9	FW6YJSM01C8879_antisense	rCG63475 [Rattus norvegicus]
9	contig05232	PREDICTED: similar to keratin associated protein 9.2 [Bos taurus]
9	contig08778_antisense	PREDICTED: similar to mannose receptor, C type 2 [Strongylocentrotus purpuratus]
9	FW6YJSM01EMY1H	RecName: Full=Pol polyprotein; Contains: RecName: Full=Reverse transcriptase/ribonuclease H; Short=RT; Contains: RecName: Full=Integrase; Short=IN
9	FW6YJSM02PGDGJ_antisense	peptidoglycan synthetase [Nephroselmis olivacea]
11	contig08019	sterol regulatory element binding transcription factor 1 [Bos taurus]
11	FW6YJSM02TWSWH	cysteine-type endopeptidase, putative [Ricinus communis]
11	FW6YJSM01EDTKC	Tetratricopeptide TPR_2 repeat protein [Desulfatibacillum alkenivorans AK-01]
11	FW6YJSM02RWD13	endonuclease/reverse transcriptase [Branchiostoma floridae]
11	contig00657	ORF2-encoded protein [Danio rerio]
11	contig22311	hypothetical protein RUMOBE_04205 [Ruminococcus obeum ATCC 29174]

11	FW6YJSM01BWOEO_antisense	RecName: Full=DEP domain-containing protein 7
11	FW6YJSM01CLBR5	hydroxyproline-rich glycoprotein family protein [Arabidopsis thaliana]
11	contig16185	ABC transporter family protein [Tetrahymena thermophila SB210]
11	contig01860	periplasmic protein [Neisseria meningitidis 053442]
11	FW6YJSM02SAJM9_antisense	60S ribosomal protein L5 [Salmo salar]
11	FW6YJSM02P5N4R_antisense	RNA recognition motif domain containing protein [Entamoeba histolytica HM-1:IMSS]
11	contig09843_antisense	PREDICTED: similar to scavenger receptor class F, member 1 isoform 1 precursor [Canis familiaris]
11	contig08019_antisense	protein kinase family protein [Arabidopsis thaliana]
11	FW6YJSM02S7SZH_antisense	PREDICTED: similar to chromosome condensation protein G [Apis mellifera]
11	contig27555_antisense	glutathione peroxidase 7 [Mus musculus]
11	contig01680	zinc ion binding [Arabidopsis thaliana]
11	FW6YJSM01APWKF	melibiose:sodium symporter [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
11	contig12607_antisense	PREDICTED: similar to SRY (sex determining region Y)-box 30 [Monodelphis domestica]
11	contig26079_antisense	protein kinase, putative [Toxoplasma gondii GT1]
11	FW6YJSM01BX5VV_antisense	PREDICTED: similar to stonustoxin alpha-subunit [Monodelphis domestica]
11	FW6YJSM02P72CQ	PREDICTED: similar to reverse transcriptase (put.); putative [Acyrthosiphon pisum]
11	FW6YJSM01BJ7KL	unnamed protein product [Candida glabrata]
11	contig05327	FACT complex component Spt16 (predicted) [Schizosaccharomyces pombe 972h-]
11	contig27095_antisense	PREDICTED: similar to marapsin, partial [Strongylocentrotus purpuratus]
11	contig09190	putative ankyrin [Arabidopsis thaliana]
11	FW6YJSM01EL76O_antisense	PREDICTED: similar to Tigger transposable element-derived protein 6 [Acyrthosiphon pisum]
11	contig00661	PREDICTED: similar to reverse transcriptase-like protein [Strongylocentrotus purpuratus]
11	contig13706_antisense	glycoside hydrolase, family 31 [Victivallis vadensis ATCC BAA-548]
11	contig11493	histone acetyltransferase HPA2 [Vibrio harveyi HY01]
11	FW6YJSM01AWIL1	PREDICTED: similar to Homeobox protein engrailed-1 (Hu-En-1) [Macaca mulatta]
11	FW6YJSM02S5I9T	PREDICTED: similar to Probable RNA-directed DNA polymerase from transposon BS (Reverse transcriptase), partial [Hydra magnipapillata]
11	FW6YJSM02S3NYQ	L-arabinose isomerase [Flavobacterium sp. MED217]
11	FW6YJSM01ATNRQ	PREDICTED: similar to G protein-coupled receptor 65 [Macaca mulatta]
11	FW6YJSM01DHW76_antisense	60S ribosomal protein L5 [Salmo salar]
11	FW6YJSM02PEHS8	NA
11	FW6YJSM01AXBR4	Collectin sub-family member 12 [Homo sapiens]
11	FW6YJSM01BG59A_antisense	NADH dehydrogenase [Shewanella sp. W3-18-1]
11	FW6YJSM01BRXT2_antisense	catalase [Pleurotus ostreatus]
11	contig00657	60S ribosomal protein L5 [Salmo salar]
11	contig12437_antisense	zinc finger (C2H2 type) protein (WIP5) [Arabidopsis thaliana]
11	contig02236	extensin-like protein Dif54 [Solanum lycopersicum]
11	FW6YJSM02R22W2_antisense	PREDICTED: similar to mitogen-activated protein kinase kinase kinase 4 [Acyrthosiphon pisum]
11	FW6YJSM01APIM1	SdhA, substrate of the Dot/Icm system [Legionella pneumophila str. Lens]

11	FW6YJSM01C6TJ8	COG1519: 3-deoxy-D-manno-octulosonic-acid transferase [Magnetospirillum magnetotacticum MS-1]
11	contig25031	putative membrane protein [Clostridium botulinum C str. Eklund]
11	contig01860	REPEATED
11	contig00622_antisense	DNA-binding transcriptional repressor [Carnobacterium sp. AT7]
11	FW6YJSM01BGCO3_antisense	Glycosyl transferase, group 1 [Crocosphaera watsonii WH 8501]
11	FW6YJSM02QL8B9_antisense	thermosome subunit [Culex quinquefasciatus]
11	FW6YJSM01BKT2L_antisense	MarR family transcriptional regulator [Picrophilus torridus DSM 9790]
11	contig29290	DNA-directed RNA polymerase beta' subunit [Mariprofundus ferrooxydans PV-1]
11	contig02472_antisense	peroxisome proliferator-activated receptor gamma [Oncorhynchus keta]
11	contig00825_antisense	olfactory receptor, family 13, subfamily G, member 1 [Homo sapiens]
11	FW6YJSM02Q6SQY	probable glycosyltransferase [Synechococcus sp. BL107]
11	FW6YJSM02P8OMU_antisense	TonB-dependent receptor [Gluconacetobacter diazotrophicus PAl 5]
11	FW6YJSM01AGGOI	DNA repair protein RecO [Anaplasma phagocytophilum HZ]
11	FW6YJSM01CLUAS_antisense	PREDICTED: similar to Rho GTPase activating protein 17 [Gallus gallus]
11	FW6YJSM01A803B	NADH dehydrogenase subunit 6 [Echinococcus canadensis]
11	FW6YJSM02SN0NT	aldehyde dehydrogenase [Ralstonia metallidurans CH34]
11	FW6YJSM02TFWY5_antisense	ADL165Cp [Ashbya gossypii ATCC 10895]
11	FW6YJSM01DUM9T	PREDICTED: similar to Cornifin B (Small proline-rich protein 1B) (SPR1B) (SPR1 B) [Rattus norvegicus]
11	FW6YJSM01EINUJ_antisense	ras interacting protein [Dictyostelium discoideum AX4]
11	FW6YJSM01B5XAR	cleavage stimulation factor subunit 3 isoform 1 [Homo sapiens]
11	FW6YJSM02SKROR	hypothetical protein MED217_09547 [Flavobacterium sp. MED217]
11	FW6YJSM01DI7QG	putative suppressor for copper-sensitivity B precursor [Photobacterium sp. SKA34]
11	FW6YJSM01D8JYT_antisense	multidrug resistance protein MexA [Rhodopirellula baltica SH 1]
11	FW6YJSM02TVC8Y_antisense	PREDICTED: similar to Na+/H+ exchanger domain containing 2 [Bos taurus]
11	FW6YJSM01D9HRX	reverse transcriptase-like protein [Paralichthys olivaceus]
11	contig00395	PREDICTED: similar to glucose dehydrogenase [Acyrthosiphon pisum]
11	FW6YJSM02RO9DK_antisense	TonB-dependent receptor [Colwellia psychrerythraea 34H]
11	contig08290_antisense	PREDICTED: similar to chloride channel 7, partial [Strongylocentrotus purpuratus]
11	contig00450	26S proteasome regulatory subunit S1 [Dictyostelium discoideum AX4]
11	FW6YJSM01BBOOW	CNPV016 C-type lectin-like protein [Canarypox virus]
11	FW6YJSM02S094G	nucleoporin 43kDa [Xenopus laevis]
11	contig12011_antisense	hypothetical protein lin1655 [Listeria innocua Clip11262]
11	contig13191	PREDICTED: similar to Solute carrier family 44, member 4 [Strongylocentrotus purpuratus]
11	FW6YJSM01BXJXS_antisense	DNA polymerase/reverse transcriptase [Hepatitis B virus]
11	contig00661	REPEATED
11	FW6YJSM01BREH5_antisense	putative arginine/ornithine antiporter transporter [Escherichia fergusonii ATCC 35469]
11	FW6YJSM01DOE1H_antisense	alpha amylase [Crassostrea gigas]
11	contig00395	REPEATED
11	FW6YJSM01EUF18_antisense	reverse transcriptase-like protein [Paralichthys olivaceus]
11	FW6YJSM02Q5ZFJ_antisense	sterol regulatory element binding protein Scp1 [Schizosaccharomyces

		pombe]
11	contig01680	REPEATED
11	FW6YJSM01DE02K	APM2/PEX13 (ABERRANT PEROXISOME MORPHOLOGY 2); protein binding [Arabidopsis thaliana]
11	FW6YJSM02PVP22_antisense	PREDICTED: similar to Cytochrome P450 39A1 (Oxysterol 7-alpha-hydroxylase) (hCYP39A1) [Canis familiaris]
11	contig16892_antisense	trypsin proteinase inhibitor precursor [Nicotiana miersii]
11	FW6YJSM02QS3FV_antisense	pseudouridine synthase D [Helicobacter pylori P12]
11	FW6YJSM01BB6NO	PREDICTED: similar to leucine zipper protein [Hydra magnipapillata]
11	contig00697_antisense	probable phage resistance protein [Planctomyces maris DSM 8797]
11	FW6YJSM01A8KN2	NADH dehydrogenase subunit 5 [Saturnia boisduvalii]
11	FW6YJSM02RCVN5	PREDICTED: similar to Cytokine receptor common subunit beta precursor (GM-CSF/IL-3/IL-5 receptor common beta-chain) (CD131 antigen) (CDw131) [Sus scrofa]
11	FW6YJSM02SCPJC	cytochrome cd1 nitrate reductase [uncultured organism]
11	contig25850_antisense	polyketide synthase [Dictyostelium discoideum AX4]
11	contig21401_antisense	PREDICTED: similar to Solute carrier family 7 (cationic amino acid transporter, y+ system), member 5, partial [Strongylocentrotus purpuratus]
11	FW6YJSM01DT86X	zonadhesin [Mus musculus]
11	FW6YJSM01DK0JU	ABC transporter related [Methylobacterium radiotolerans JCM 2831]
11	FW6YJSM02TGW47	leucine rich repeat (in FLII) interacting protein 2 [Danio rerio]
11	FW6YJSM02P97LF_antisense	acyl-CoA dehydrogenase-like [Burkholderia cenocepacia AU 1054]
11	contig23509	TRAP transporter, DctM family [Klebsiella pneumoniae 342]
11	FW6YJSM01DO5IP	acetyl-CoA acetyltransferase [Thermus thermophilus HB8]
11	FW6YJSM02S8NGG_antisense	putative 3,4-dihydroxy-2-butanone 4-phosphate/shikimate 5-dehydrogenase fusion protein [Bacteroides fragilis NCTC 9343]
11	FW6YJSM01D0EBT	protein kinase, putative [Toxoplasma gondii GT1]
11	FW6YJSM01AKCDR_antisense	estrogen receptor beta [Danio rerio]
4	contig00034	sugar ABC transporter, ATP-binding protein [Roseobacter sp. MED193]
4	contig24254	acetyl-coA carboxylase protein [Cryptosporidium muris RN66]
4	FW6YJSM01CAKWG_antisense	hypothetical protein Emin_0408 [Elusimicrobium minutum Pei191]
4	contig00034	REPEATED
4	FW6YJSM01D4JW3_antisense	hypothetical protein [Monosiga brevicollis MX1]
4	contig29591	Rad50 DNA repair/recombination protein [Phaeodactylum tricornutum CCAP 1055/1]
4	FW6YJSM02RZ4AA_antisense	Type 2C protein phosphatase (PP2C); inactivates the osmosensing MAPK cascade by dephosphorylating Hog1p; mutation delays mitochondrial inheritance; deletion reveals defects in precursor tRNA splicing, sporulation and cell separation; Ptc1p [Saccharomyces cerevisiae]
4	contig16104	PREDICTED: solute carrier family 13 (sodium/sulfate symporters), member 4 isoform 1 [Macaca mulatta]
4	contig07538	paternally expressed 3 [Mus musculus]
4	contig00272_antisense	nuclear transcription factor, X-box binding, putative [Ricinus communis]
4	contig28123	short chain dehydrogenase [Chlorobium tepidum TLS]
4	contig03938_antisense	gag-pol polyprotein
4	FW6YJSM01ECFX6	predicted protein [Nematostella vectensis]
4	contig08394	RecName: Full=C-X-C chemokine receptor type 5; Short=CXC-R5; Short=CXCR-5; AltName: Full=Burkitt lymphoma receptor 1 homolog; AltName: CD_antigen=CD185

4	FW6YJSM02P2J1I_antisense	aminotransferase, class I and II [Roseobacter sp. CCS2]
4	contig13401_antisense	diaminopimelate epimerase [Aurantimonas sp. SI85-9A1]
4	contig21989	PREDICTED: similar to dpy-19-like 1, like [Canis familiaris]
4	FW6YJSM01EK6QE_antisense	PREDICTED: similar to N-acetyltransferase UNQ2771/PRO7155 homolog (GNAT acetytransferase) [Ciona intestinalis]
4	contig03872_antisense	NA
4	FW6YJSM02TBN5H	phenylalanyl-tRNA synthetase, beta subunit, putative [Campylobacter upsaliensis RM3195]
4	FW6YJSM02ROS4B_antisense	monooxygenase FAD-binding [Burkholderia ambifaria MC40-6]
4	contig21123_antisense	G protein-coupled receptor PGR28 [Homo sapiens]
4	FW6YJSM02RJJNF	jumonji family transcription factor, putative [Neosartorya fischeri NRRL 181]
4	contig20968_antisense	GK20092 [Drosophila willistoni]
4	contig10995	Ste24 endopeptidase [Methylobacillus flagellatus KT]
4	FW6YJSM01CL64W	G protein-coupled receptor 22 [Rattus norvegicus]
4	contig21526	translational activator gcn1 [Aedes aegypti]
4	contig25108	ammonia monooxygenase/particulate methane monooxygenase subunit A [uncultured bacterium]
4	FW6YJSM02RX0DL	flagellar protein export ATPase FliI [Methylobacterium sp. 4-46]
4	FW6YJSM01CSX64_antisense	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [Hahella chejuensis KCTC 2396]
4	contig26436	glycosyl transferase family protein [Pseudoalteromonas atlantica T6c]
4	FW6YJSM02SW5P5	UDP-N-acetylenolpyruvoylglucosamine reductase [Bifidobacterium animalis subsp. lactis HN019]
4	contig11132	intermediate subunit of galactose lectin 1 [Entamoeba dispar]
4	FW6YJSM02PDR8B_antisense	cytochrome b [Acanthopagrus taiwanensis]
4	FW6YJSM02SB0JJ_antisense	hypothetical protein [Plasmodium berghei strain ANKA]
4	contig00271_antisense	Ser/Thr protein kinase [Cryptosporidium parvum Iowa II]
4	FW6YJSM01D4N0L_antisense	aconitate hydratase, mitochondrial precursor [Aspergillus terreus NIH2624]
4	FW6YJSM01BC6NK_antisense	hypothetical protein HpylHP_12645 [Helicobacter pylori HPKX_438_AG0C1]
4	contig23921	PREDICTED: similar to Cytochrome P450 7B1 (Oxysterol 7-alpha-hydroxylase) [Canis familiaris]
4	FW6YJSM02P3D31	Ser/Thr protein phosphatase family protein [Myxococcus xanthus DK 1622]
4	FW6YJSM01DIVQY	transducin / WD-40 repeat protein family [Cryptosporidium hominis TU502]
4	FW6YJSM01BF38Y	PREDICTED: similar to Snf2-related CBP activator protein [Macaca mulatta]
4	FW6YJSM01D3YWD_antisense	protein binding protein, putative [Ricinus communis]
4	contig22447_antisense	PREDICTED: similar to sodium-dependent phosphate transporter [Tribolium castaneum]
4	contig26305_antisense	HECT type ubiquitin ligase, putative [Toxoplasma gondii GT1]
4	FW6YJSM01D62U4_antisense	mCG141070 [Mus musculus]
3	contig01761	guanine deaminase [Neosartorya fischeri NRRL 181]
3	contig17624	component of SCAR regulatory complex [Dictyostelium discoideum AX4]
3	FW6YJSM01D2FPS_antisense	PREDICTED: erythrocyte protein band 4.1-like 4b [Mus musculus]
3	FW6YJSM02Q3ZU0	GM16709 [Drosophila sechellia]
3	FW6YJSM01B6E6C_antisense	amine oxidase family protein [Musa acuminata]
3	FW6YJSM01BQXFS	potassium-transporting ATPase subunit A [Flavobacterium psychrophilum JIP02/86]
3	contig22937_antisense	membrane-bound lytic murein transglycosylase D [Photobacterium profundum SS9]

3	contig02393	PREDICTED: similar to N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase (Phosphatidylinositol-glycan biosynthesis class L protein) (PIG-L) [Equus caballus]
3	FW6YJSM01AGJ3E	thiol:disulfide interchange protein [Vibrio shilonii AK1]
3	FW6YJSM01AVQNC	NA
3	FW6YJSM02RDW6P	N-6 DNA methylase [Natranaerobius thermophilus JW/NM-WN-LF]
3	FW6YJSM02SZX6J	TRAP-type uncharacterized transport system periplasmic component-like protein [Methylobacterium nodulans ORS 2060]
3	contig19431	formamidopyrimidine-DNA glycosylase [Mesoplasma florum L1]
3	FW6YJSM01CZWPB_antisense	ORF2
3	contig11970_antisense	GL23987 [Drosophila persimilis]
3	contig20675	ribosomal protein L22 [Saccharophagus degradans 2-40]
3	contig15031	progestin and adipoQ receptor family member III [Mus musculus]
3	FW6YJSM01APL8A	acetolactate synthase small subunit [Candidatus Sulcia muelleri str. Hc (Homalodisca coagulata)]
3	contig05999	RNA-editing complex protein [Leishmania major strain Friedlin]
3	contig03217_antisense	transmembrane regulator [Bordetella avium 197N]
3	FW6YJSM01DHG6J_antisense	coiled-coil domain containing 93 [Danio rerio]
3	contig07366_antisense	lethal(2)giant larvae gene
3	contig26555	sperm mitochondrial-associated cysteine-rich protein [Peromyscus polionotus subgriseus]
3	contig02393	REPEATED
3	contig17321	immunoglobulin gamma heavy chain variable region [Ornithorhynchus anatinus]
3	contig14655_antisense	glycosyl transferase [Haemophilus influenzae Rd KW20]
3	contig12618	ACYL-COA DEHYDROGENASE, SHORT-CHAIN SPECIFIC; ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT; RUBREDOXIN [Fusobacterium nucleatum subsp. vincentii ATCC 49256]
3	FW6YJSM01D6RNI	PREDICTED: similar to Histone-like transcription factor, putative, partial [Hydra magnipapillata]
3	FW6YJSM02SWY0K	PREDICTED: similar to V1R pheromone receptor-like protein [Ornithorhynchus anatinus]
3	FW6YJSM01EHXTN_antisense	methyl-accepting chemotaxis protein [Borrelia duttonii Ly]
3	contig16590	solute carrier family 39 (zinc transporter), member 10 (predicted), isoform CRA_a [Rattus norvegicus]
3	contig23849	diacylglycerol kinase, putative [Toxoplasma gondii GT1]
3	contig05999	REPEATED
3	FW6YJSM01B66MT	Hypothetical protein [Oryza sativa]
3	contig03933_antisense	GGDEF domain-containing protein [Rhodopseudomonas palustris BisB5]
3	FW6YJSM02PW5V2_antisense	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein [Sulfolobus acidocaldarius DSM 639]
3	contig24058_antisense	PREDICTED: similar to olfactory receptor MOR31-4 [Equus caballus]
3	FW6YJSM01CWGPQ	envelope glycoprotein [Human immunodeficiency virus type 1]
5	FW6YJSM01EPOIT_antisense	3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase [Pyrenophora tritici-repentis Pt-1C-BFP]
5	FW6YJSM02SOXXR_antisense	transporter [Plasmodium vivax SaI-1]
5	FW6YJSM02S7KDJ_antisense	multidrug transport protein (MATE family) [Alteromonadales bacterium TW-7]
5	contig20806	inositol polyphosphate-1-phosphatase [Xenopus (Silurana) tropicalis]
5	contig11406_antisense	PREDICTED: similar to eukaryotic translation elongation factor isoform 1 [Strongylocentrotus purpuratus]

5	contig03899_antisense	putative molybdate-binding ABC-transporter periplasmic binding-protein [Bordetella petrii DSM 12804]
5	FW6YJSM02SQST0	protein tyrosine phosphatase receptor type A precursor [Lethenteron japonicum]
5	contig05924	glycosyl transferase [Cryptosporidium muris RN66]
5	FW6YJSM02PXVUI_antisense	PREDICTED: similar to ankyrin 2,3/unc44 [Strongylocentrotus purpuratus]
5	FW6YJSM01DJ5Q8_antisense	WD repeat domain 8 [Gallus gallus]
5	FW6YJSM01ETP03_antisense	hypothetical protein DP2850 [Desulfotalea psychrophila LSv54]
5	FW6YJSM01DIG7A	thiol:disulfide interchange protein DsbD [Candidatus Vesicomyosocius okutanii HA]
5	FW6YJSM01CMSKM	PREDICTED: similar to multidrug resistance protein 2; MRP2 [Monodelphis domestica]
5	FW6YJSM01EKDG0_antisense	alpha-glucan phosphorylase [Thermus aquaticus Y51MC23]
5	contig12590_antisense	ABC transporter family protein [Tetrahymena thermophila SB210]
5	FW6YJSM01D1INP	RecName: Full=Dynein heavy chain; Short=DYHC
5	contig15945_antisense	PREDICTED: similar to Probable RNA-directed DNA polymerase from transposon X-element (Reverse transcriptase) [Danio rerio]
5	FW6YJSM01CXXK0	euchromatic histone-lysine N-methyltransferase 1 [Xenopus (Silurana) tropicalis]
5	FW6YJSM01D7BZU	putative FAD-linked oxidoreductase [Corynebacterium diphtheriae NCTC 13129]
5	contig25887	NAD-dependent aldehyde dehydrogenase [Lactobacillus acidophilus NCFM]
5	FW6YJSM01AE18E_antisense	Zinc finger, C2H2 type family protein [Brugia malayi]
5	FW6YJSM01DQGGU	breast and ovarian cancer susceptibility protein variant BRCA1-delta 11b [Bos taurus]
5	FW6YJSM02PYCJ4_antisense	hypothetical protein LOC419817 [Gallus gallus]
5	contig10944	peroxisome proliferator-activated receptor gamma [Oncorhynchus keta]
5	contig02914	extracellular solute-binding protein [Staphylothermus marinus F1]
5	contig03969	Fis family transcriptional regulator [Burkholderia cenocepacia MC0-3]
5	contig00862	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
5	contig05924	REPEATED
5	contig25924	kinesin-associated protein, putative [Aedes aegypti]
5	FW6YJSM01A7K1C	hypothetical protein DP0866 [Desulfotalea psychrophila LSv54]
5	FW6YJSM02TZ8NP	ATPase family protein associated with various cellular activities [Mycobacterium smegmatis str. MC2 155]
5	FW6YJSM01E16UF_antisense	possible transmembrane protein [Algoriphagus sp. PR1]
5	contig09015	ABC transporter permease [Clostridium kluyveri DSM 555]
5	FW6YJSM02QJYVU	acyl-CoA dehydrogenase domain-containing protein [Salinispora tropica CNB-440]
5	FW6YJSM02QGGBL_antisense	NADH dehydrogenase subunit F [Bejaria zamorae]
5	FW6YJSM02RRNQF	DNA nuclease [Streptococcus suis]
5	FW6YJSM01DFKXF_antisense	Paralichthys olivaceus TBT-bp 1 gene for Tributyltin binding protein type 1, complete cds
5	FW6YJSM02R39C6	glycoprotein G1 and G2 precursor; envelope glycoprotein precursor [Pergamino virus]
5	FW6YJSM01CIPU3_antisense	cytochrome b/b6 domain-containing protein [Syntrophobacter fumaroxidans MPOB]
5	contig22675_antisense	endoarabinanase [Penicillium chrysogenum]
5	FW6YJSM01DGCDL	hypothetical protein DP0866 [Desulfotalea psychrophila LSv54]

5	contig06586	NADH dehydrogenase subunit 5 [Naegleria gruberi]
5	FW6YJSM02PVZW8	Membrane-fusion protein, contains peptidase family M50 domain [Methylokorus infernorum V4]
5	contig03969	REPEATED
5	FW6YJSM02RK069	protein of unknown function DUF323 [Cyanothece sp. PCC 8801]
5	contig06586	REPEATED
5	contig17463	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20 [Homo sapiens]
6	FW6YJSM01AKOSP	protein RtcB [Capnocytophaga sputigena Capno]
6	contig07147_antisense	cytochrome c oxidase III
6	FW6YJSM01CAJNA	ABC transporter, putative [Toxoplasma gondii GT1]
6	FW6YJSM01CHQSJ_antisense	PREDICTED: similar to Homeobox protein slou (S59/2) (Protein slouch) (Homeobox protein NK-1) [Apis mellifera]
6	FW6YJSM01B3SRE	granulin b [Danio rerio]
6	FW6YJSM02S1Z9A_antisense	PREDICTED: similar to olfactory receptor OR20, partial [Taeniopygia guttata]
6	contig01619_antisense	PREDICTED: similar to neural stem cell-derived dendrite regulator [Acyrthosiphon pisum]
6	FW6YJSM02S7M0C_antisense	predicted protein [Populus trichocarpa]
6	FW6YJSM02QZEAK	heat shock protein Hsp20 domain-containing protein [Dictyostelium discoideum AX4]
6	contig07860_antisense	spermidine/putrescine ABC transporter, permease protein [Lyngbya sp. PCC 8106]
6	FW6YJSM01B0TDR	seven transmembrane helix receptor [Homo sapiens]
6	contig14582	UDP-glucose/GDP-mannose dehydrogenase [Methanococcus maripaludis C5]
6	FW6YJSM01CZKVY	eukaryotic translation initiation factor 3, subunit 6, [Plasmodium berghei strain ANKA]
6	FW6YJSM01BD8JD	cytochrome C oxidase subunit III
6	contig01302_antisense	Chaperone protein dnaJ, putative [Ricinus communis]
6	FW6YJSM02SN4TK_antisense	seven transmembrane helix receptor [Homo sapiens]
6	FW6YJSM01BRF4J	unnamed protein product [Tetraodon nigroviridis]
6	contig25389_antisense	Kinesin motor domain containing protein [Tetrahymena thermophila SB210]
6	FW6YJSM02PTR8S	transcription factor 19 [Oryzias dancena]
6	FW6YJSM01CV0Y1_antisense	DEAH-box RNA helicase [Chlamydomonas reinhardtii]
6	FW6YJSM01DE9Y4	COG0437: Fe-S-cluster-containing hydrogenase components 1 [Magnetospirillum magnetotacticum MS-1]
6	FW6YJSM02SUGI6	uncharacterized protein involved in outer membrane biogenesis [Lawsonia intracellularis PHE/MN1-00]
6	contig09130_antisense	adenylate cyclase type ix [Aedes aegypti]
6	FW6YJSM02RY4SH_antisense	interleukin 17 receptor D [Danio rerio]
6	contig02846	DnaJ (Hsp40) homolog, subfamily B, member 7 [Rattus norvegicus]
6	FW6YJSM01BPEIS_antisense	glycosyl transferase family protein [Thiomicrospira crunogena XCL-2]
6	FW6YJSM01D2AEA	transcription factor MYB75 [Arabidopsis thaliana]
6	FW6YJSM02S43MG_antisense	PREDICTED: similar to Transformation/transcription domain-associated protein (350/400 kDa PCAF-associated factor) (PAF350/400) (STAF40) (Tra1 homolog) isoform 2 [Canis familiaris]
6	FW6YJSM02QUSIL	shikimate 5-dehydrogenase [Cyanothece sp. PCC 7424]
6	FW6YJSM01AM2NJ	5-aminolevulinate synthase [Dictyostelium discoideum AX4]
6	contig16441	hypothetical protein [Plasmodium falciparum 3D7]
6	contig18150	diguanylate cyclase [Thermosipho melanesiensis BI429]

6	FW6YJSM01BB8T5	acetolactate synthase III large subunit [Lyngbya sp. PCC 8106]
6	FW6YJSM02TSECE	Tctex1 domain containing 4 [Homo sapiens]
6	FW6YJSM01EZUB3_antisense	NADH dehydrogenase subunit 6 [Echinococcus canadensis]
6	FW6YJSM01DXM6P	anti-mullerian hormone [Rattus norvegicus]
6	contig08250_antisense	PREDICTED: similar to Protein kinase domain containing protein [Hydra magnipapillata]
6	FW6YJSM02R7ATA	porin [Synechococcus sp. WH 8102]
6	FW6YJSM02S6PXO	F-box only protein 23 [Rattus norvegicus]
6	contig20922_antisense	Myosin-2 heavy chain, non muscle, putative [Ricinus communis]
6	FW6YJSM02TQU5L_antisense	PREDICTED: similar to exportin 5, partial [Strongylocentrotus purpuratus]
6	contig00449	Na+/proline symporter [Flavobacteria bacterium BBFL7]
6	contig03787	hypothetical protein BACUNI_02464 [Bacteroides uniformis ATCC 8492]
6	contig17838_antisense	RasGEF domain containing protein [Trichomonas vaginalis G3]
6	FW6YJSM02SV7PA	glycosyl transferase, group 1 family protein [Oceanicaulis alexandrii HTCC2633]
6	FW6YJSM02SNTDY	NA
6	FW6YJSM02QJDKX	RecName: Full=Zinc finger protein 672
6	contig03151	putative dehydratase/racemase [Bordetella bronchiseptica RB50]
6	FW6YJSM01BUZ6S	LysR, substrate-binding [Enterococcus faecium DO]
6	contig00449	repeated
6	FW6YJSM01EXTJO_antisense	NADH dehydrogenase subunit 2 [Cafeteria roenbergensis]
6	FW6YJSM01BSNW1_antisense	zinc finger protein 628 [Mus musculus]
6	FW6YJSM02RU1OL_antisense	major Facilitator superfamily protein [Zea mays]
6	FW6YJSM02RTTQB	GA25715 [Drosophila pseudoobscura pseudoobscura]
6	contig11498	histone acetyltransferase HPA2 [Vibrio harveyi HY01]
6	contig02846	REPEATED
6	FW6YJSM02P2286_antisense	Tubulin-tyrosine ligase family protein [Tetrahymena thermophila SB210]
6	FW6YJSM01BHFMP	large Forked protein [Drosophila melanogaster]
7	FW6YJSM01BZFUT	hypothetical protein LELG_03221 [Lodderomyces elongisporus NRRL YB-4239]
7	contig21613	acetyltransferase [Leuconostoc citreum KM20]
7	FW6YJSM02Q5N12	molybdenum cofactor sulfurylase [Rhodobacter sphaeroides 2.4.1]
7	contig02607	exodeoxyribonuclease V, alpha chain [Nostoc punctiforme PCC 73102]
7	contig22797	glutathione S-transferase domain-containing protein [Ochrobactrum anthropi ATCC 49188]
7	FW6YJSM01CULJC	neuroendocrine convertase 1 precursor [Brugia malayi]
7	FW6YJSM02TQ9N6	PREDICTED: similar to sorting nexin 8 [Danio rerio]
7	FW6YJSM02R48FV_antisense	transposon protein, putative, CACTA, En/Spm sub-class [Oryza sativa (japonica cultivar-group)]
7	FW6YJSM02TEUGG	UDP-glucose 4-epimerase [marine gamma proteobacterium HTCC2080]
7	FW6YJSM02S0PIJ_antisense	Metaxin 1 [Mus musculus]
7	contig20263_antisense	PREDICTED: similar to Cytochrome P450 7A1 (CYPVII) (Cholesterol 7-alpha-monooxygenase) (Cholesterol 7-alpha-hydroxylase) [Ciona intestinalis]
7	FW6YJSM01BI2SU_antisense	NADH dehydrogenase subunit 2 [Vasdavidius concursus]
7	contig19955	amino acid ABC transporter permease [Sinorhizobium meliloti 1021]
7	contig19314	acetyl-coenzyme A synthetase [Gramella forsetii KT0803]

7	contig11585	PREDICTED: similar to transposase [Strongylocentrotus purpuratus]
7	contig24447	nitrite/nitrate transporter [Chlamydomonas reinhardtii]
7	FW6YJSM01AQDEV	MAP kinase-interacting serine/threonine kinase 1 [Mus musculus]
7	contig18935_antisense	reverse transcriptase [Drosophila ananassae]
7	FW6YJSM02QX6DF_antisense	hypothetical protein K07A12.5 [Caenorhabditis elegans]
7	contig29182_antisense	erythrocyte membrane protein 1 (PfEMP1) [Plasmodium falciparum 3D7]
7	FW6YJSM02R8PVA_antisense	RecName: Full=Type 1 phosphatases regulator YPI1
7	FW6YJSM01B14N0	high-affinity potassium uptake transporter [Debaryomyces occidentalis]
7	FW6YJSM02R67GS	Misexpression suppressor of KSR, putative [Aedes aegypti]
7	contig28170_antisense	transporter [Trypanosoma cruzi strain CL Brener]
7	contig05184	putative secreted protein [Erythrobacter sp. SD-21]
7	FW6YJSM02PWYMT	cytochrome c oxidase subunit III [Argiope bruennichi]
7	contig00923_antisense	Neuropeptide-Like Protein family member (nlp-32) [Caenorhabditis elegans]
7	FW6YJSM01C7F5S	diguanylate cyclase [Methylobacterium extorquens PA1]
7	contig03903_antisense	olfactory receptor 1066 [Mus musculus]
7	contig12747	testis-specific RNP-type RNA binding protein [Drosophila melanogaster]
7	contig29281	meiotically upregulated gene Mug57 [Schizosaccharomyces pombe 972h-]
7	FW6YJSM02QFJUB_antisense	PREDICTED: similar to RhoGTPase regulating protein variant ARHGAP20-1ad [Strongylocentrotus purpuratus]
7	FW6YJSM02P7Q9L	PREDICTED: similar to Achaete-scute complex-like 1 (Drosophila) [Strongylocentrotus purpuratus]
7	contig06405_antisense	PREDICTED: similar to Y26D4A.9, partial [Ciona intestinalis]
7	contig05184	serine/threonine protein kinase with PASTA sensor(s) [Exiguobacterium sibiricum 255-15]
7	contig05835_antisense	NADH dehydrogenase subunit 4 [Chaetosoma scaritides]
7	contig08702	Mid1 interacting protein 1 (gastrulation specific G12-like (zebrafish)) [Mus musculus]
7	FW6YJSM02SD35K	NADH dehydrogenase subunit 2 [Bemisia tabaci]
7	FW6YJSM02Q299X	putative oligopeptidase A [Synechococcus sp. WH 5701]
7	contig16191_antisense	membrane protein, putative [Microscilla marina ATCC 23134]
7	FW6YJSM02PO9XF	PREDICTED: similar to helicase [Acyrthosiphon pisum]
7	FW6YJSM01C3UXV	5-oxoprolinase (ATP-hydrolyzing) [Acidovorax sp. JS42]
7	FW6YJSM02SWS09	glycosyl transferase [Desulfotalea psychrophila LSv54]
7	FW6YJSM01BIHGZ_antisense	CYCLOIDEA-like group 1B protein [Lupinus rivularis]
7	FW6YJSM02TGCE9	PREDICTED: similar to bucentaur, partial [Strongylocentrotus purpuratus]
7	FW6YJSM01CABET_antisense	PREDICTED: RNA binding motif protein 15 isoform 1 [Pan troglodytes]
7	contig01133_antisense	NADH dehydrogenase subunit 2 [Portunus trituberculatus]
7	contig28856_antisense	peroxisomal acyl-CoA thioesterase 2 [Mus musculus]
7	contig02667_antisense	PREDICTED: similar to endonuclease/reverse transcriptase [Strongylocentrotus purpuratus]
7	FW6YJSM02P48FW_antisense	asparagine synthase (glutamine-hydrolyzing) [Lactobacillus gasseri ATCC 33323]
7	FW6YJSM01EB79Y	Tetraspanin family protein [Tetrahymena thermophila SB210]
7	contig19265	PREDICTED: similar to Zinc finger X-chromosomal protein [Gallus gallus]
7	FW6YJSM01C1ALS	ATP-dependent DNA helicase RecG [Sulfurihydrogenibium sp. YO3AOP1]
7	FW6YJSM02PTHT6	RecName: Full=Neuroendocrine convertase 1; Short=NEC 1; AltName: Full=Prohormone convertase 1; AltName: Full=Proprotein convertase 1;

		Short=PC1; AltName: Full=PC3; AltName: Full=Furin homolog; AltName: Full=Propeptide-processing protease; Flags: Precursor
7	contig02607	NA
7	FW6YJSM01A35U6_antisense	NA
7	contig10812_antisense	PREDICTED: similar to olfactory receptor MOR172-2 [Ornithorhynchus anatinus]
7	contig25999	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
7	FW6YJSM01E144N_antisense	NADH dehydrogenase subunit 2 [Vasdavidius concursus]
8	contig00009	regulator of nonsense transcripts [Plasmodium falciparum 3D7]
8	contig19849	hemocyanin isoform 1 [Nucula nucleus]
8	FW6YJSM01DE5LQ	hypothetical protein ANACAC_00192 [Anaerostipes caccae DSM 14662]
8	FW6YJSM01C0VVD_antisense	hypothetical protein BRAFLDRAFT_96376 [Branchiostoma floridae]
8	contig24089	retrotransposon protein, putative, Ty3-gypsy subclass [Oryza sativa (japonica cultivar-group)]
8	FW6YJSM01CWXIK_antisense	hypothetical 2-hydroxyacid dehydrogenase [Coccidioides immitis RS]
8	FW6YJSM02RIX32	peptidyl-prolyl cis-trans isomerase [Clostridium botulinum E3 str. Alaska E43]
8	FW6YJSM02POTHY	NADH dehydrogenase subunit 5 [Exorista japonica]
8	contig21483_antisense	putative DEAD/DEAH box helicase [Colwellia psychrerythraea 34H]
8	contig17165_antisense	D amino acid oxidase (DAO) family protein [Flavobacteria bacterium BAL38]
8	contig03538	glycine cleavage T-protein (aminomethyl transferase) domain containing protein [Babesia bovis T2Bo]
8	FW6YJSM01ARY2I_antisense	T-cell receptor beta chain ANA 11 [Brugia malayi]
8	contig07709	PREDICTED: similar to cytosine-5-methyltransferase 3-like protein [Monodelphis domestica]
8	contig00009	repeated
8	FW6YJSM02ST5JK_antisense	PREDICTED: similar to zinc finger protein 534 [Strongylocentrotus purpuratus]
8	contig06438	oxygen-dependent protoporphyrinogen IX oxidase [Toxoplasma gondii]
8	contig06938	PREDICTED: similar to lipoprotein receptor-related protein 6 [Strongylocentrotus purpuratus]
8	FW6YJSM02R0AZN	signal recognition particle-docking protein FtsY [Treponema denticola ATCC 35405]
8	FW6YJSM01BD3DI_antisense	ABC transporter, periplasmic binding protein [Mesorhizobium loti MAFF303099]
8	contig05472	hemocyanin [Haliotis tuberculata]
8	FW6YJSM01DZ3A0	RNA pseudouridylate synthase, putative [Toxoplasma gondii GT1]
8	FW6YJSM01DJTMO	cell wall surface anchor family protein [Enterococcus faecalis V583]
8	FW6YJSM02R4RDS_antisense	cytochrome c oxidase-like protein [Glyptapanteles flavicoxis]
8	contig16957_antisense	beta-carotene C4 oxygenase [Brevundimonas aurantiaca]
8	contig06934	potential chromatin-associated protein [Candida albicans SC5314]
8	contig24524	hypothetical protein HH1713 [Helicobacter hepaticus ATCC 51449]
8	FW6YJSM01DA1AL	hypothetical protein [Cryptosporidium hominis TU502]
8	contig03538	repeated
8	FW6YJSM01BZN95_antisense	UDP-2,3-diacylglucosamine hydrolase [Burkholderia oklahomensis EO147]
8	contig05472	repeated
8	FW6YJSM01C7FLM	PREDICTED: similar to olfactory receptor Olr633 [Equus caballus]
8	contig25631	oligopeptide transport integral membrane protein [Clavibacter

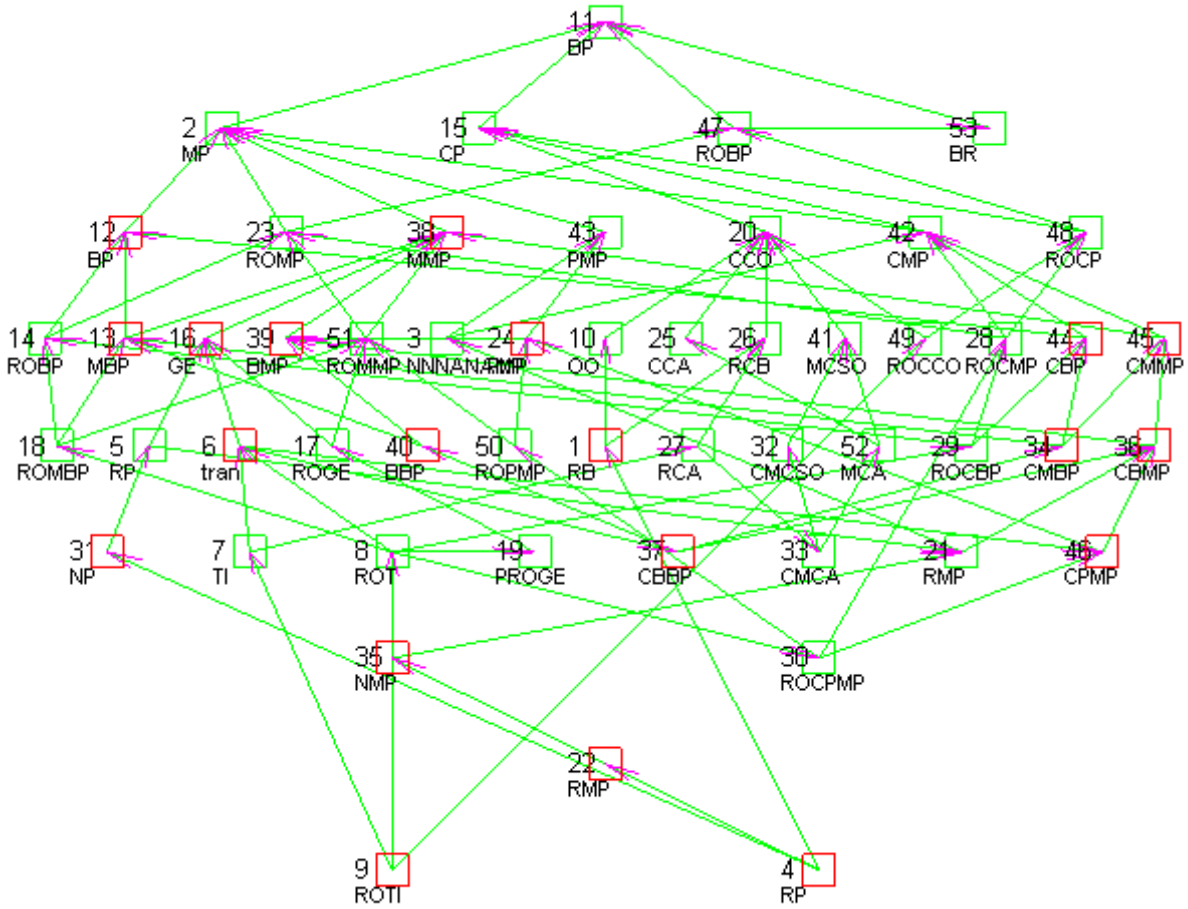
		michiganensis subsp. sepedonicus]
8	contig28388	GTPase protein [Mycoplasma mobile 163K]
8	FW6YJSM01A56WI	SULfate Permease family member (sulp-2) [Caenorhabditis elegans]
8	FW6YJSM02S1QSF_antisense	chloramphenicol-sensitive protein RarD [Salmonella enterica subsp. enterica serovar Typhi str. CT18]
8	FW6YJSM02TZ6MX_antisense	mitogen-inducible gene 6 protein [Bos taurus]
8	FW6YJSM02SWG3C	lysophospholipase 3, isoform CRA_b [Mus musculus]
8	FW6YJSM01EE0LJ_antisense	interleukin 27 [Mus musculus]
8	FW6YJSM01B9YCK	cutaneous T-cell lymphoma tumor antigen se70-2 [Brugia malayi]
8	contig06934	repeated
8	contig06938	repeated
8	FW6YJSM01B7TFM_antisense	Putative Sorghum bicolor 22 kDa kafirin cluster [Oryza sativa]
8	contig20558	PREDICTED: similar to oxysterol 7alpha-hydroxylase [Monodelphis domestica]
8	FW6YJSM02QP0DZ	PREDICTED: similar to dishevelled-associated activator of morphogenesis 1, partial [Ciona intestinalis]
8	FW6YJSM02RZ2S0	FGGY carbohydrate kinase domain containing [Rattus norvegicus]
8	FW6YJSM02S8R1Q_antisense	centaurin beta [Aedes aegypti]
8	contig01229	cytochrome c oxidase subunit III [Tetranychus urticae]
8	contig06438	X-box binding protein [Haliotis discus discus]
8	contig28297_antisense	NA
8	contig29033	c3h4-type ring finger protein [Plasmodium falciparum 3D7]
8	contig07416	binding-protein-dependent transport systems inner membrane component [Burkholderia cenocepacia AU 1054]
8	contig00828_antisense	PREDICTED: similar to ATP citrate lyase [Danio rerio]
8	contig00872	NA
8	FW6YJSM01DWB5G_antisense	PREDICTED: similar to GTPase, IMAP family member 7 [Danio rerio]
8	FW6YJSM01AMZOE	nicotinate phosphoribosyltransferase [Lodderomyces elongisporus NRRL YB-4239]
8	FW6YJSM01EJKKL	novel protein similar to vertebrate glutamate receptor, ionotropic, N-methyl D-aspartate 2D (GRIN2D) [Danio rerio]
8	contig28387	glycosyl transferase [Clostridium butyricum 5521]
8	FW6YJSM02TU476	putative type III secretion protein YscQ [Hahella chejuensis KCTC 2396]
8	FW6YJSM01BLY0M_antisense	two-component response regulator [Finegoldia magna ATCC 29328]
8	FW6YJSM02RE9HT	cell surface antigen I/II precursor [Lactococcus lactis subsp. cremoris MG1363]
8	contig16061_antisense	NA
8	FW6YJSM01BP6FD_antisense	ecdysone-induced protein 75A [Aedes aegypti]
8	FW6YJSM01AKP1A	PREDICTED: similar to WD domain containing protein like, GAstrulation Defective GAD-1 (69.0 kD) (gad-1), partial [Danio rerio]
8	FW6YJSM01EKONT_antisense	ubiquitin domain-containing protein [Dictyostelium discoideum AX4]
8	FW6YJSM01EJ989_antisense	PREDICTED: similar to polyhomeotic-like 2 (Drosophila), partial [Ornithorhynchus anatinus]
8	FW6YJSM02RHT7H	PREDICTED: similar to alpha-1 type XI collagen [Strongylocentrotus purpuratus]
8	FW6YJSM02RC4Z8	hypothetical protein NEMVEDRAFT_v1g149903 [Nematostella vectensis]
8	contig16649_antisense	Protein kinase domain containing protein [Tetrahymena thermophila SB210]
8	FW6YJSM01A74Q4	hypothetical protein [Monosiga brevicollis MX1]

8	FW6YJSM02Q3MNP	integral membrane protein [Streptococcus suis 05ZYH33]
8	FW6YJSM01ALQ3M_antisense	glycosyl transferase [Haemophilus influenzae Rd KW20]
8	FW6YJSM02QWJYA	chordin-like protein [Hydra magnipapillata]
8	FW6YJSM01EJXJ3	mCG15855, isoform CRA_a [Mus musculus]
8	contig26097_antisense	integral membrane protein [Theileria annulata strain Ankara]
8	FW6YJSM01E0T50	zinc finger protein [Ciona intestinalis]
10	contig08129_antisense	retinoblastoma binding protein 9 [Danio rerio]
10	FW6YJSM02PQJDY	Solute carrier family 35, member F3 [Homo sapiens]
10	contig26021	NADH dehydrogenase (quinone) [Methylobacterium populi BJ001]
10	contig00924_antisense	basic helix-loop-helix (bHLH) family protein [Arabidopsis thaliana]
10	FW6YJSM01DU69T_antisense	glucose-specific phosphotransferase system enzym II, factor IIB [Staphylococcus carnosus subsp. carnosus TM300]
10	FW6YJSM02RN0GX_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	FW6YJSM01DXACP_antisense	PREDICTED: similar to neuronal nicotinic acetylcholine receptor beta4 subunit [Monodelphis domestica]
10	FW6YJSM01D4HJS_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	FW6YJSM02TXROJ	unnamed protein product [Tetraodon nigroviridis]
10	contig24601	MFS transporter, putative [Penicillium marneffei ATCC 18224]
10	FW6YJSM02TR1DT_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	contig10359	PREDICTED: similar to pol polyprotein [Danio rerio]
10	FW6YJSM02RWVEK	tubulin-tyrsoine ligase-like protein [Leishmania infantum JPCM5]
10	FW6YJSM01C5DMD_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02SPRYH	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	FW6YJSM01DI7N6_antisense	Catenin (cadherin-associated protein), alpha 3 [Bos taurus]
10	FW6YJSM02TW4LA_antisense	Catenin (cadherin-associated protein), alpha 3 [Bos taurus]
10	contig01961	Fc receptor, IgE, low affinity II, alpha polypeptide isoform b [Rattus norvegicus]
10	FW6YJSM01CDLMM_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	contig22689_antisense	ABC transporter ATP-binding protein [Bacillus pumilus SAFR-032]
10	FW6YJSM01CH0NR	PREDICTED: similar to Homeobox protein slou (S59/2) (Protein slouch) (Homeobox protein NK-1) [Apis mellifera]
10	contig17084	ABC transporter related [Prosthecochloris aestuarii DSM 271]
10	FW6YJSM01E0S6B	transcriptional regulator [Hahella chejuensis KCTC 2396]
10	FW6YJSM02SZ36Z	protein serine/threonine kinase [Dictyostelium discoideum AX4]
10	FW6YJSM01BRFYF	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02QY767_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM01B86NB_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	contig15595	RecName: Full=Mitotic checkpoint serine/threonine-protein kinase BUB1 beta; AltName: Full=MAD3/BUB1-related protein kinase; Short=BubR1; AltName: Full=Mitotic checkpoint kinase MAD3L
10	contig10986_antisense	translation initiation factor IF-2 [Escherichia coli O157:H7 str. Sakai]
10	contig29329	histone acetylase complex subunit Paf400, putative [Neosartorya fischeri

		NRRL 181]
10	FW6YJSM02REAVR_antisense	DNA-binding protein elg, putative [Aedes aegypti]
10	contig24964_antisense	NA
10	FW6YJSM02SIN4E	homeobox protein HoxC8bb [Salmo salar]
10	FW6YJSM02PVTE7	G protein-coupled receptor 35 [Rattus norvegicus]
10	contig20420_antisense	PREDICTED: similar to RNA binding motif protein, X-linked 2 isoform 1 [Bos taurus]
10	FW6YJSM02QZM2E_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	contig00789	HMG-box variant [Schizosaccharomyces pombe]
10	FW6YJSM02TEHVW_antisense	PREDICTED: similar to Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E isoform 3 [Pan troglodytes]
10	FW6YJSM02QHWAR	PREDICTED: similar to Vitamin K epoxide reductase complex, subunit 1 [Pan troglodytes]
10	FW6YJSM01BVZVH	hypothetical protein NP1886A [Natronomonas pharaonis DSM 2160]
10	contig02805_antisense	hypothetical protein [Plasmodium yoelii yoelii str. 17XNL]
10	FW6YJSM01B2CP2	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02SO72B	RecName: Full=DNA-directed RNA polymerase subunit alpha; Short=PEP; AltName: Full=Plastid-encoded RNA polymerase subunit alpha; Short=RNA polymerase subunit alpha
10	FW6YJSM02PTH8A	glycosyl transferase family 2 [Streptococcus suis 89/1591]
10	FW6YJSM01BJD5A	CCAAT-box DNA binding protein subunit B [Plasmodium falciparum 3D7]
10	contig05006	predicted protein [Nematostella vectensis]
10	FW6YJSM01BRCBE_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	contig27256	maturase K [Eurystyles sp. Szlachetko s.n.]
10	FW6YJSM01A2YKM_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02S5SVQ_antisense	PREDICTED: similar to neuronal nicotinic acetylcholine receptor beta4 subunit [Monodelphis domestica]
10	contig26423_antisense	PREDICTED: similar to putative pheromone receptor (Go-VN5) [Rattus norvegicus]
10	contig27390	maturase K [Eurystyles sp. Szlachetko s.n.]
10	FW6YJSM02RX104_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	contig15533	PREDICTED: similar to zinc transporter 5 [Monodelphis domestica]
10	FW6YJSM02PZLWZ	amidase [Frankia sp. CcI3]
10	FW6YJSM02SKAEW_antisense	AAAP amino acid permease [Laccaria bicolor S238N-H82]
10	contig21745	methyl-accepting chemotaxis protein McpH [Roseovarius sp. HTCC2601]
10	contig00789	HMG-box variant [Schizosaccharomyces pombe]
10	FW6YJSM02QDZG7	PREDICTED: similar to tyrosine kinase receptor, partial [Hydra magnipapillata]
10	contig19935	NADH dehydrogenase subunit 2 [Erinaceus europaeus]
10	FW6YJSM02QRVDS_antisense	PREDICTED: similar to glutathione S-transferase class-alpha, partial [Ornithorhynchus anatinus]
10	FW6YJSM01AE9EH_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02TYT6B	Protein kinase domain containing protein [Tetrahymena thermophila SB210]
10	FW6YJSM01CCTH9_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]

10	contig08369_antisense	Chitin-inducible gibberellin-responsive protein, putative [Ricinus communis]
10	contig25190	hypothetical protein [Tetrahymena thermophila SB210]
10	contig06970_antisense	glyceraldehyde-3-phosphate dehydrogenase GAPDH [Sus scrofa]
10	FW6YJSM01AUOS5_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	contig05006	envelope glycoprotein [Feline immunodeficiency virus]
10	FW6YJSM01BLMSO	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM01BNTAW	NADH dehydrogenase subunit 6 [Laminaria ephemera]
10	contig16211_antisense	protein kinase [Dictyostelium discoideum AX4]
10	FW6YJSM02RP4KR_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM02TI12T_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	FW6YJSM01EZQ47_antisense	PREDICTED: similar to endonuclease-reverse transcriptase, partial [Strongylocentrotus purpuratus]
10	contig28385_antisense	glycosyl transferase [Clostridium butyricum 5521]
10	FW6YJSM02TUJXC	RecName: Full=Taste receptor type 2 member 1; Short=T2R1
10	FW6YJSM01AS0Z8_antisense	NADH dehydrogenase subunit 5 [Gyrodactylus thymalli]
10	FW6YJSM02RW00U	light-independent protochlorophyllide reductase, B subunit [Roseobacter sp. AzwK-3b]
10	contig18440	methyl-accepting chemotaxis sensory transducer [Desulfatibacillum alkenivorans AK-01]
10	FW6YJSM02S5VET_antisense	PREDICTED: similar to endonuclease-reverse transcriptase [Strongylocentrotus purpuratus]
10	contig24915	ubiquitin-transferase, putative [Toxoplasma gondii RH]
10	contig00800_antisense	L-aminoadipate-semialdehyde dehydrogenase large subunit [Aspergillus terreus NIH2624]

- Appendix 3.2. A focused view at the significant Go term; diagram of significant GO terms and their ancestor terms.

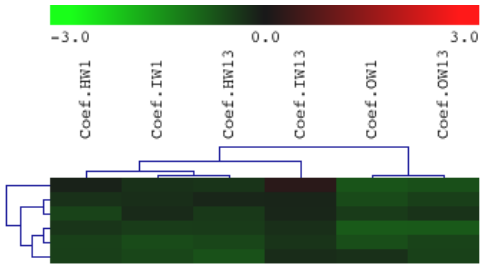


1. ribosome biogenesis
2. metabolic process
3. nucleobase, nucleoside, nucleotide and nucleic acid metabolic
4. rRNA processing
5. RNA processing
6. translation
7. translational initiation
8. regulation of translation
9. regulation of translational initiation
10. organelle organization
11. biological-process
12. biosynthetic process
13. macromolecule biosynthetic process
14. regulation of biosynthetic process
15. cellular process
16. gene expression
17. regulation of gene expression
18. regulation of macromolecule biosynthetic process
19. posttranscriptional regulation of gene expression
20. cellular component organization
21. RNA metabolic process
22. rRNA metabolic process
23. regulation of metabolic process
24. protein metabolic process
25. cellular component assembly
26. ribonucleoprotein complex biogenesis
27. ribonucleoprotein complex assembly
28. regulation of cellular metabolic process
29. regulation of cellular biosynthetic process
30. regulation of cellular protein metabolic process
31. ncRNA processing
32. cellular macromolecular complex subunit organization
33. cellular macromolecular complex assembly
34. cellular macromolecule biosynthetic process
35. ncRNA metabolic process
36. cellular biopolymer metabolic process
37. cellular biopolymer biosynthetic process
38. macromolecule metabolic process
39. biopolymer metabolic process
40. biopolymer biosynthetic process
41. macromolecular complex subunit organization
42. cellular metabolic process
43. primary metabolic process
44. cellular biosynthetic process
45. cellular macromolecule metabolic process
46. cellular protein metabolic process
47. regulation of biological process
48. regulation of cellular process
49. regulation of cellular component organization
50. regulation of protein metabolic process
51. regulation of macromolecule metabolic process
52. macromolecular complex assembly
53. biological regulation

- Appendix 3.3: Expression profile analysis for PPAR network and receptors; interesting clusters details and projected functions

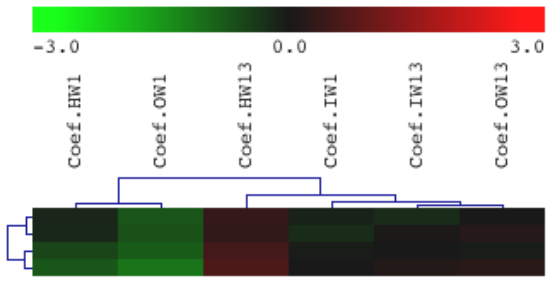
1. PPAR network

Cluster5:



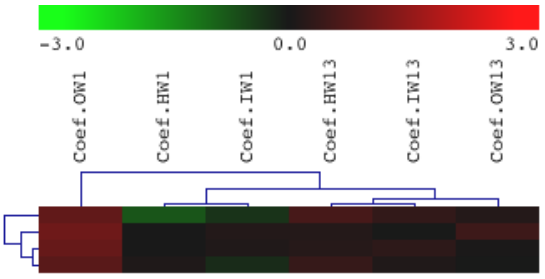
"; K08762 DBI, ACBP; diazepam-binding inhibitor (GABA receptor modulator, acyl-CoA-binding protei
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K08764 SCP2, SCPX; sterol carrier protein 2 [EC:2.3.1.176]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]; long"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]; long

Cluster4:



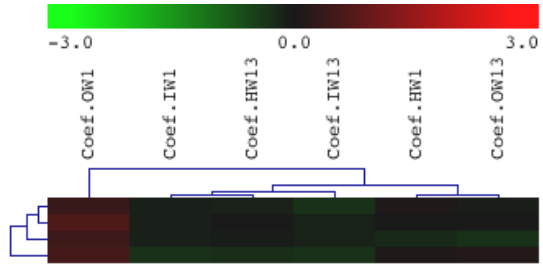
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K00232 E1.3.3.6, ACOX1, ACOX3; acyl-CoA oxidase [EC:1.3.3.6]"
; K01388 MMP1; matrix metalloproteinase-1 (interstitial collagenase) [EC:3.
; K08771 AQP7; aquaporin-7; long

Cluster10:



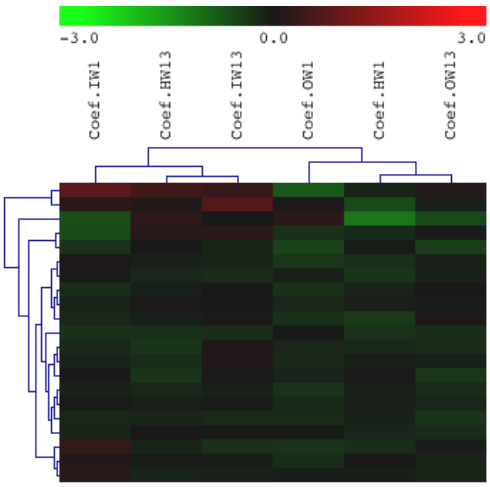
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"

Cluster9:



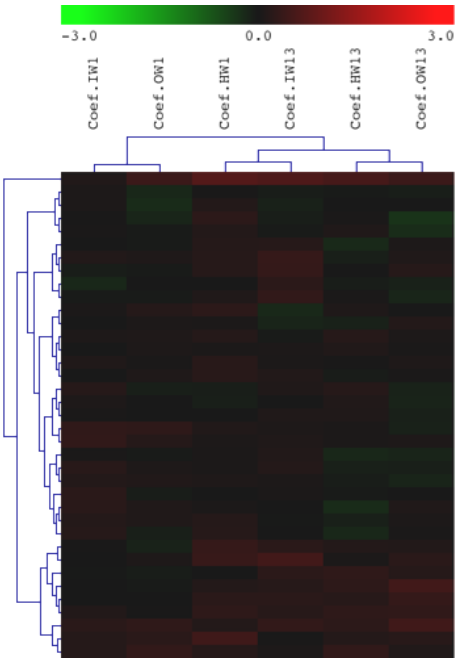
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]
; K06276 PDPK1; 3-phosphoinositide dependent protein kinase-1 [EC:2.7.11.1]

Cluster1:



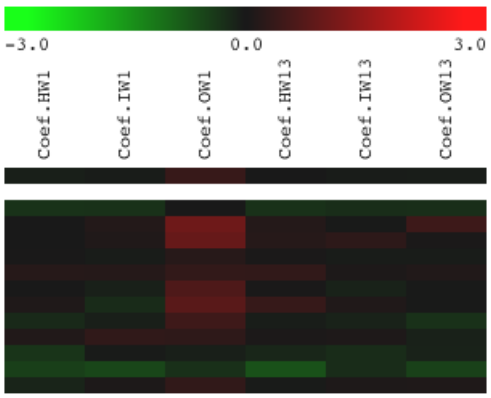
"; K07425 CYP4A; cytochrome P450, family 4, subfamily A [EC:1.14.15.3]"
; K06259 CD36; CD36 antigen: long
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K08745 SLC27A1_4, FATP1, FATP4; solute carrier family 27 (fatty acid transporter), member 1/4
"; K08764 SCP2, SCPX; sterol carrier protein 2 [EC:2.3.1.176]; long"
"; K08767 ANGPTL4, PGAR; angiopoietin-like 4"
"; K07425 CYP4A; cytochrome P450, family 4, subfamily A [EC:1.14.15.3]"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]
"; K01596 E4.1.1.32, pckA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]"
; K09025 APOA5; apolipoprotein A-V; long
; K08757 APOA1; apolipoprotein A-I
"; K07425 CYP4A; cytochrome P450, family 4, subfamily A [EC:1.14.15.3]"
"; K00249 E1.3.99.3, ACADM, acd; acyl-CoA dehydrogenase [EC:1.3.99.3]"
; K06259 CD36; CD36 antigen
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
; K08763 OLR1; oxidised low-density lipoprotein receptor 1
"; K01596 E4.1.1.32, pckA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]"
"; K00232 E1.3.3.6, ACOX1, ACOX3; acyl-CoA oxidase [EC:1.3.3.6]"
"; K08530 NR1C3, PPARG; peroxisome proliferator-activated receptor gamma"
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]

Cluster8:



"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]; long"
"; K08749 SLC27A6, FATP6; solute carrier family 27 (fatty acid transporter), member 6"
"; K00249 E1.3.99.3, ACADM, acd; acyl-CoA dehydrogenase [EC:1.3.99.3]"
"; K08745 SLC27A1_4, FATP1, FATP4; solute carrier family 27 (fatty acid transporter), member 1/4 [EC:6.2.1.-]"
"; K08767 ANGPTL4, PGAR; angiopoietin-like 4"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K08748 SLC27A5, FATP5; solute carrier family 27 (fatty acid transporter), member 5 [EC:6.2.1.7]"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]
"; K08764 SCP2, SCPX; sterol carrier protein 2 [EC:2.3.1.176]"
"; K01596 E4.1.1.32, pckA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K00232 E1.3.3.6, ACOX1, ACOX3; acyl-CoA oxidase [EC:1.3.3.6]"
"; K00232 E1.3.3.6, ACOX1, ACOX3; acyl-CoA oxidase [EC:1.3.3.6]"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K00249 E1.3.99.3, ACADM, acd; acyl-CoA dehydrogenase [EC:1.3.99.3]"
; K07296 ACDC; adiponectin
"; K08748 SLC27A5, FATP5; solute carrier family 27 (fatty acid transporter), member 5 [EC:6.2.1.7]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]
"K08745 SLC27A1_4, FATP1, FATP4; solute carrier family 27 (fatty acid transporter), member 1/4 [EC:6.2.1.-]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K00864 E2.7.1.30, glpK; glycerol kinase [EC:2.7.1.30]"
"; K08526 NR2B3, RXRG; retinoid X receptor gamma"
; K08761 PLTP; phospholipid transfer protein
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]; long"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
" K08745 SLC27A1_4, FATP1, FATP4; solute carrier family 27 (fatty acid transporter), member 1/4 [EC:6.2.1.-]"
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]
"; K00489 CYP7A1; cytochrome P450, family 7, subfamily A (cholesterol 7alpha-monooxygenase) [EC:1.14.13.17]"
"; K00232 E1.3.3.6, ACOX1, ACOX3; acyl-CoA oxidase [EC:1.3.3.6]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"

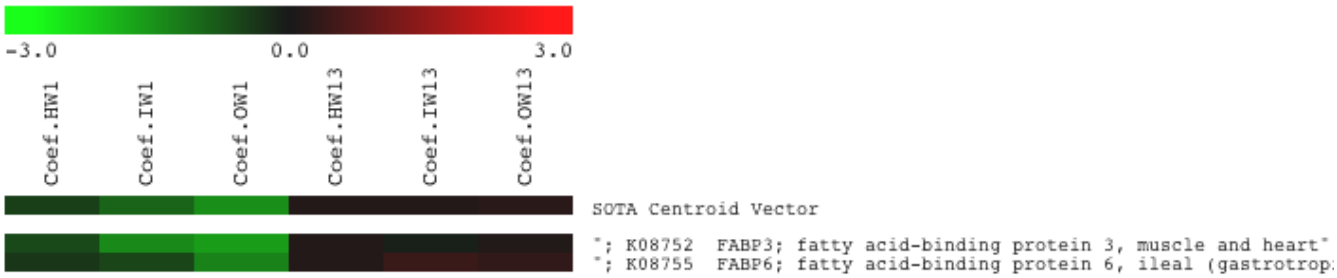
Cluster 11:



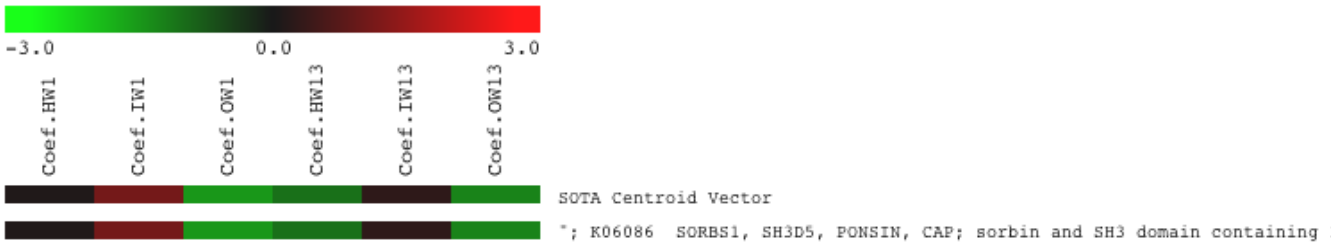
SOTA Centroid Vector

; K09025 APOA5; apolipoprotein A-V; long
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K00507 SCD, desC; stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
"; K01897 ACSL, fadD; long-chain acyl-CoA synthetase [EC:6.2.1.3]"
; K08765 CPT1; carnitine O-palmitoyltransferase 1 [EC:2.3.1.21]
"; K00249 E1.3.99.3, ACADM, acd; acyl-CoA dehydrogenase [EC:1.3.99.3]"
"; K08767 ANGPTL4, PGAR; angiopoietin-like 4"
; K06272 ILK; integrin-linked kinase [EC:2.7.11.1]; long
"; K01596 E4.1.1.32, pckA, PEPCK; phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]"

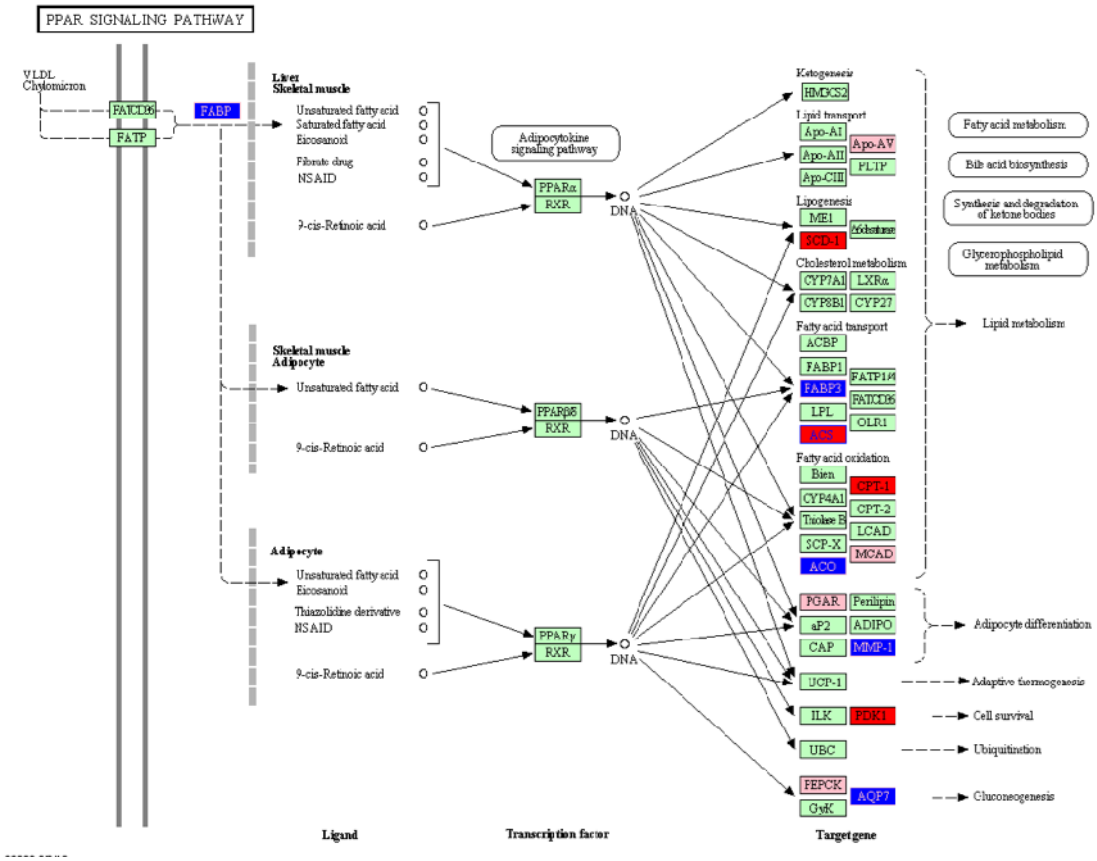
Cluster5:

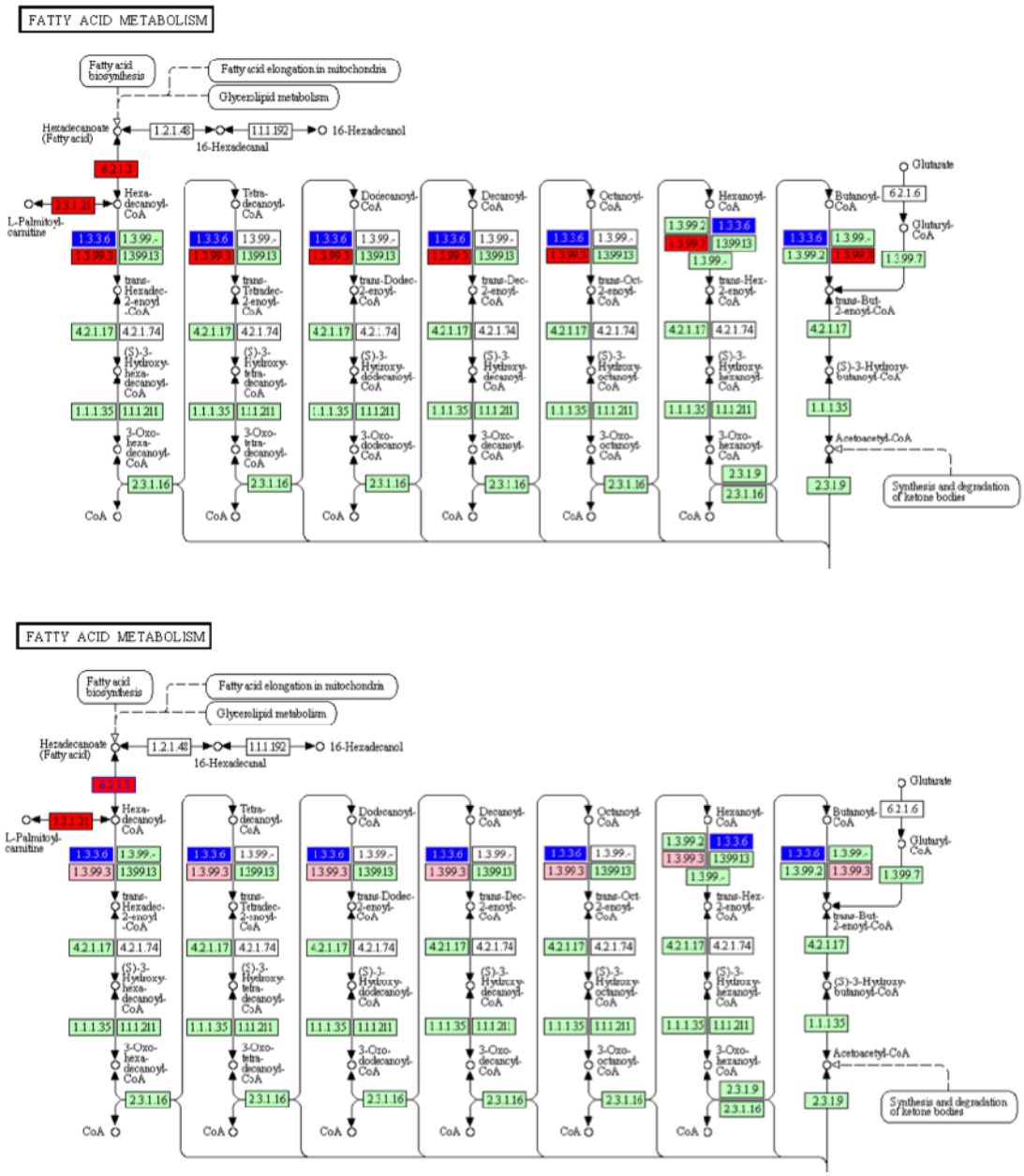


Cluster6:

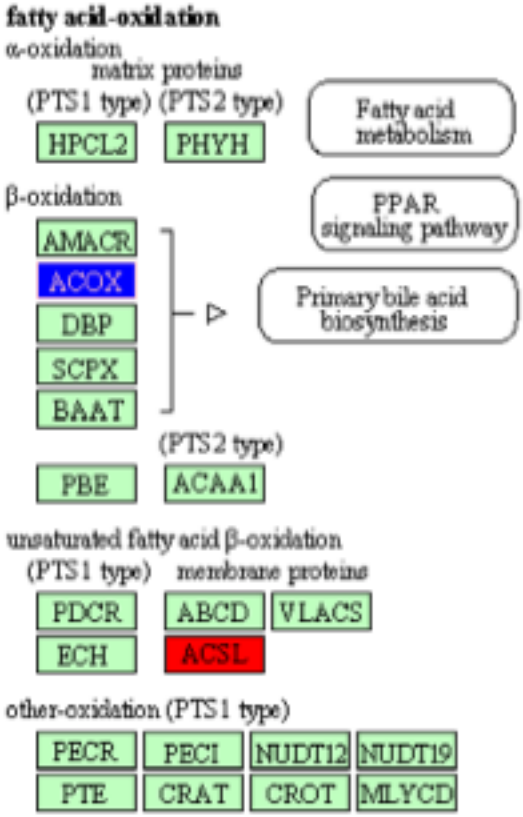


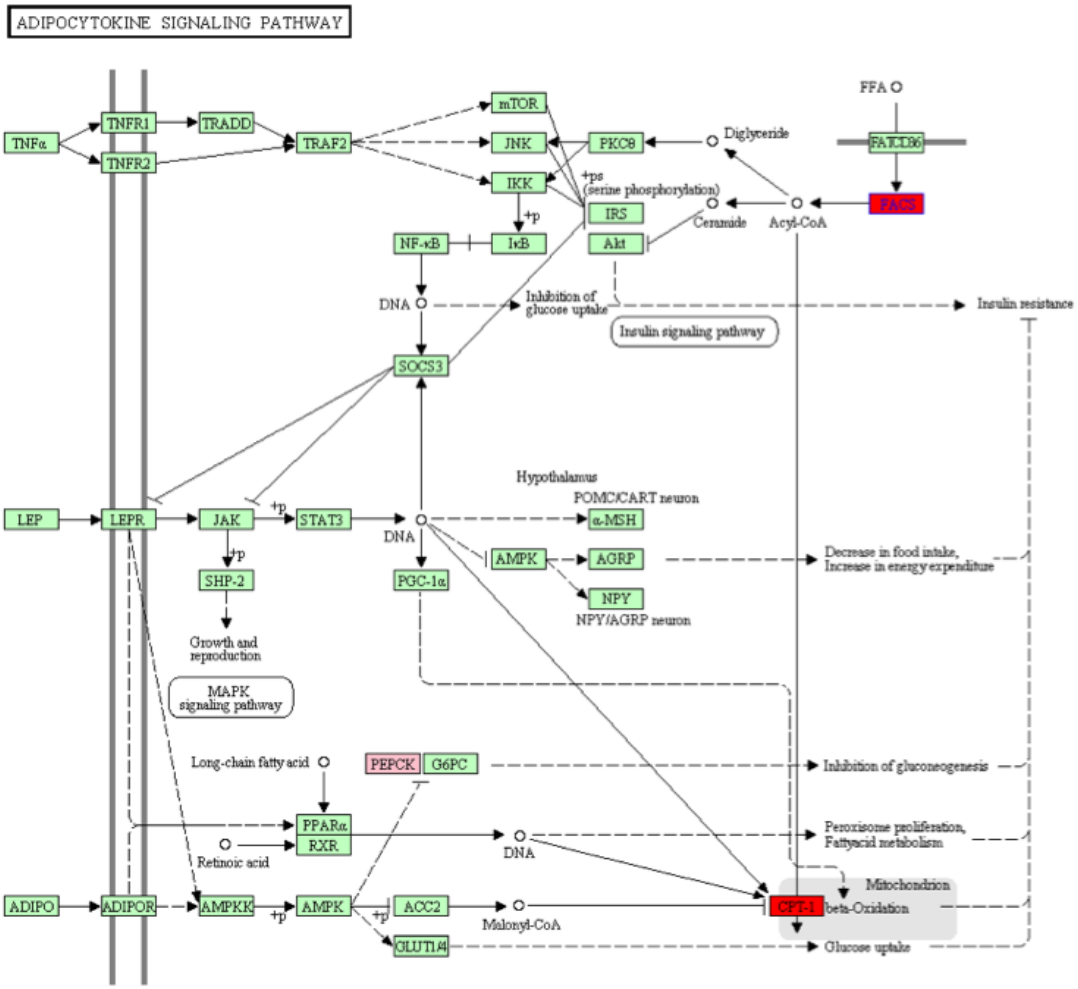
Transcription factors profiles correlations: projected functions which are differentially expressed in OW1 on the KEGG maps. Blue: down-regulation, red: up-regulation, pink: up-regulated, other cluster and less fold change. It seems that PPARγ has a central function and that different PPARs are interacting in this regulation and might crosstalk with other factors.




FATTY ACID METABOLISM
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Glycerolipid metabolism
Hexadecanoate (Fatty acid)
1.2.1.48
16-Hexadecanal
1.1.1.192
16-Hexadecanol
6.2.1.3
2.3.1.21
L-Palmitoyl-carnitine
Hexadecanoyl-CoA
Tetradecanoyl-CoA
Dodecanoyl-CoA
Decanoyl-CoA
Octanoyl-CoA
Hexanoyl-CoA
Butanoyl-CoA
Glutarate
6.2.1.6
Glutaryl-CoA
1.3.99.7
trans-Hexadec-2-enoyl-CoA
trans-Tetradec-2-enoyl-CoA
trans-Dodec-2-enoyl-CoA
trans-Dec-2-enoyl-CoA
trans-Oct-2-enoyl-CoA
trans-Hex-2-enoyl-CoA
trans-But-2-enoyl-CoA
4.2.1.17
4.2.1.74
(S)-3-Hydroxyhexadecanoyl-CoA
(S)-3-Hydroxytetradecanoyl-CoA
(S)-3-Hydroxydodecanoyl-CoA
(S)-3-Hydroxydecanoyl-CoA
(S)-3-Hydroxyoctanoyl-CoA
(S)-3-Hydroxyhexanoyl-CoA
(S)-3-Hydroxybutanoyl-CoA
1.1.1.35
1.1.1.211
3-Oxohexadecanoyl-CoA
3-Oxotetradecanoyl-CoA
3-Oxododecanoyl-CoA
3-Oxodecanoyl-CoA
3-Oxooctanoyl-CoA
3-Oxohexanoyl-CoA
Acetoacetyl-CoA
2.3.1.16
2.3.1.9
Synthesis and degradation of ketone bodies
CoA

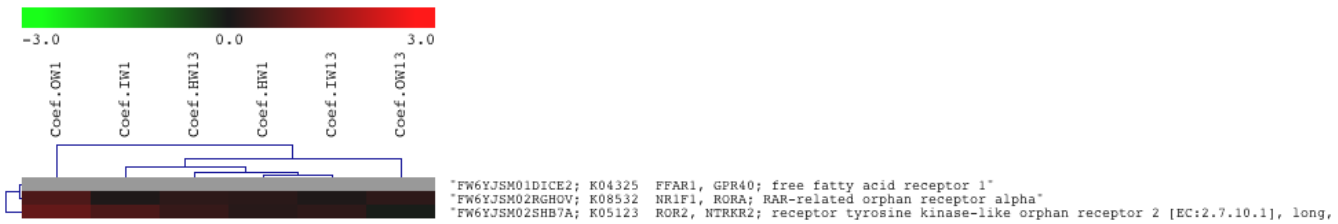
FATTY ACID METABOLISM
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Glycerolipid metabolism
Hexadecanoate (Fatty acid)
1.2.1.48
16-Hexadecanal
1.1.1.192
16-Hexadecanol
6.2.1.3
2.3.1.21
L-Palmitoyl-carnitine
Hexadecanoyl-CoA
Tetradecanoyl-CoA
Dodecanoyl-CoA
Decanoyl-CoA
Octanoyl-CoA
Hexanoyl-CoA
Butanoyl-CoA
Glutarate
6.2.1.6
Glutaryl-CoA
1.3.99.7
trans-Hexadec-2-enoyl-CoA
trans-Tetradec-2-enoyl-CoA
trans-Dodec-2-enoyl-CoA
trans-Dec-2-enoyl-CoA
trans-Oct-2-enoyl-CoA
trans-Hex-2-enoyl-CoA
trans-But-2-enoyl-CoA
4.2.1.17
4.2.1.74
(S)-3-Hydroxyhexadecanoyl-CoA
(S)-3-Hydroxytetradecanoyl-CoA
(S)-3-Hydroxydodecanoyl-CoA
(S)-3-Hydroxydecanoyl-CoA
(S)-3-Hydroxyoctanoyl-CoA
(S)-3-Hydroxyhexanoyl-CoA
(S)-3-Hydroxybutanoyl-CoA
1.1.1.35
1.1.1.211
3-Oxohexadecanoyl-CoA
3-Oxotetradecanoyl-CoA
3-Oxododecanoyl-CoA
3-Oxodecanoyl-CoA
3-Oxooctanoyl-CoA
3-Oxohexanoyl-CoA
Acetoacetyl-CoA
2.3.1.16
2.3.1.9
Synthesis and degradation of ketone bodies
CoA


fatty acid-oxidation
α-oxidation
matrix proteins
(PTS1 type)
(PTS2 type)
HPCL2
PHYH
Fatty acid metabolism
β-oxidation
PPAR signaling pathway
AMACR
ACOX
DBP
SCPX
BAAT
Primary bile acid biosynthesis
(PTS2 type)
PBE
ACAA1
unsaturated fatty acid β-oxidation
(PTS1 type)
membrane proteins
PDCR
ECH
ABCD
ACSL
VLACS
other-oxidation (PTS1 type)
PECR
PTE
PECI
CRAT
NUDT12
CROT
NUDT19
MLYCD

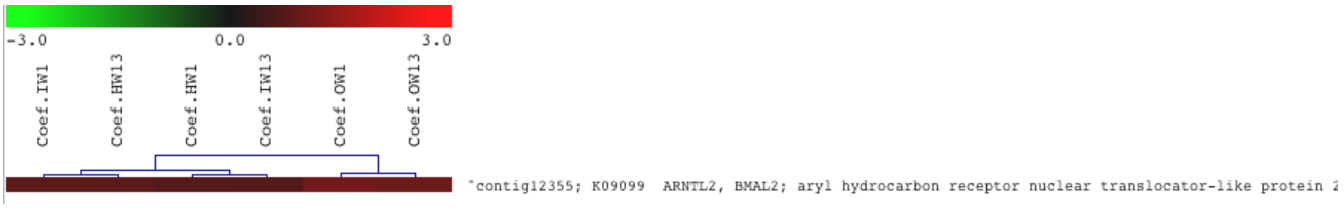

ADIPOCYTOKINE SIGNALING PATHWAY
TNFα
TNFR1
TNFR2
TRADD
TRAF2
mTOR
JNK
IKK
PKCθ
Diglyceride
FFA
FATP1/6
FACL
+ps
(serine phosphorylation)
+p
IRS
Ceramide
Acyl-CoA
NF-κB
IκB
Akt
DNA
Inhibition of
glucose uptake
Insulin resistance
Insulin signaling pathway
SOCS3
Hypothalamus
POMC/CART neuron
LEP
LEPR
JAK
STAT3
α-MSH
AMPK
AGRP
NPY
NPY/AGRP neuron
Decrease in food intake,
Increase in energy expenditure
SHP-2
PGC-1α
Growth and
reproduction
MAPK
signaling pathway
Long-chain fatty acid
PEPCK
G6PC
Inhibition of gluconeogenesis
PPARα
RXR
Retinoic acid
Peroxisome proliferation,
Fattyacid metabolism
Mitochondrion
ADIPO
ADIPOR
AMPKK
ACC2
Malonyl-CoA
CPT-1
beta-Oxidation
GLUT1/4
Glucose uptake
INSULIN SIGNALING PATHWAY
Flotillin
CAP
Cbl
CrkII
GRF2
TC10
Exo70
CIP4/2
CLUT4 vesicle
translocation
to PM
GLUT4
Glucose
Glucose uptake
Glucose
homeostasis
SHIP
SKIP
SOCS
PTP1B
JNK
IKK
aPKC
SREBP-1c
AMPK
ACC
FAS
Lipogenesis
Fatty acid
biosynthesis (path 1)
PYK
GK
Glycolysis
PGC-1α
FOXO1
DNA
G6PC
FBP
PEPCK
Glycolysis /
Gluconeogenesis
INS
INSR
IRS
PI3K
PIP3
PDK1/2
Akt
GSK-3β
GYS
Glycogen
Glycogenesis
Starch and sucrose
metabolism
PP1
PHK
PYG
Phosphatidyl inositol
signaling system
TSC2
TSC1
PDE3
cAMP
PKA
HSL
Antilipolysis
+ps : serine phosphorylation
+py : tyrosine phosphorylation
-py : tyrosine dephosphorylation
LAR
Rheb
BAD
Antapoptosis
Apoptosis
Lipid
homeostasis
MAPK signaling
pathway
mTOR
Raptor
p70S6K
S6
4EBP1
eIF4E
Protein synthesis
SHC
GRB2
SOS
Ras
Raf
MEK1/2
ERK1/2
MNK
Elk1
Proliferation,
differentiation
DNA

2. Receptors

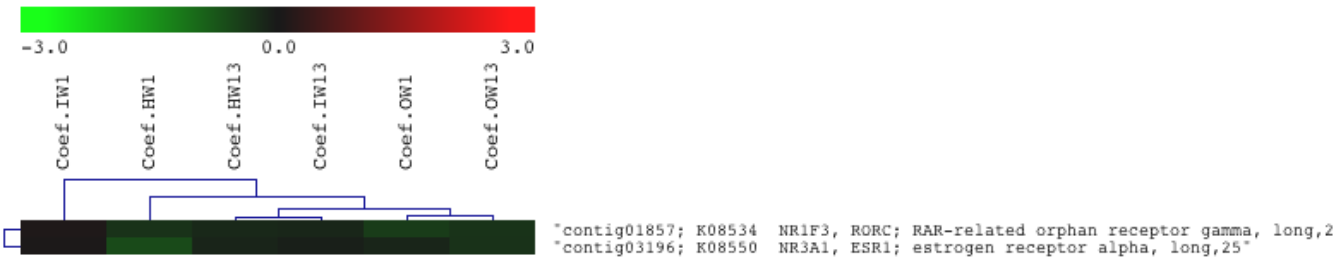
Cluster1: RORA, ROR2, GPR40-day1, OW,IW



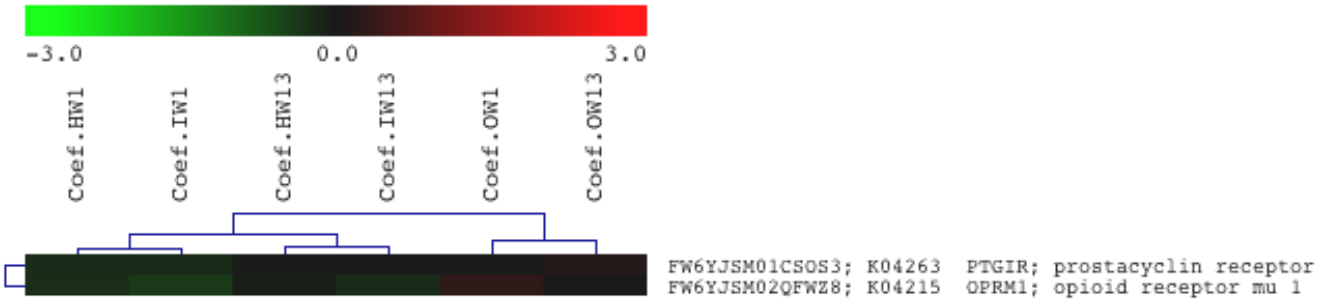
Cluster4: ARNTL2: in all -up



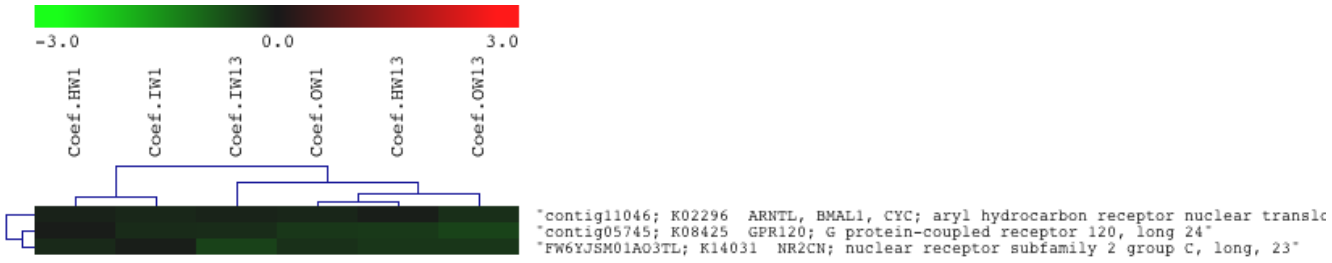
Cluster11: ER-down , RARG



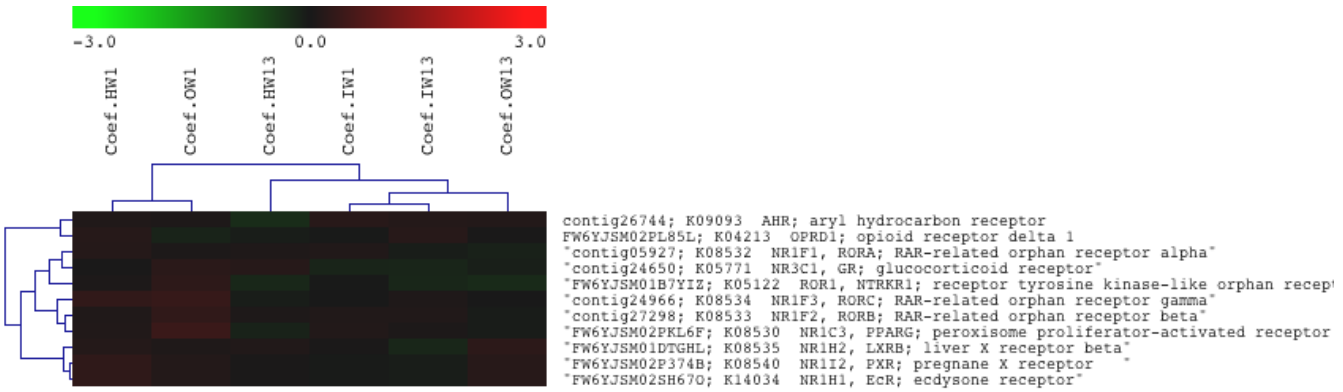
Custer 8: Prostacyclin and OPRM1-up in Ovary only



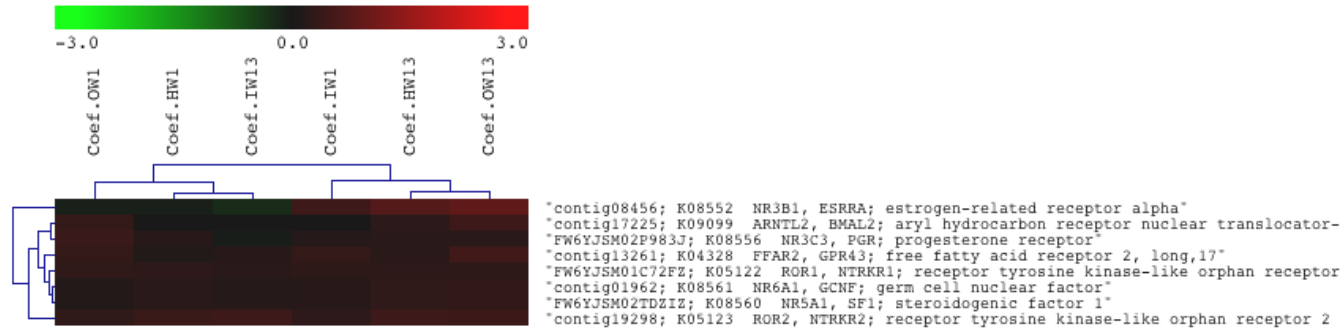
Cluster 10: late down



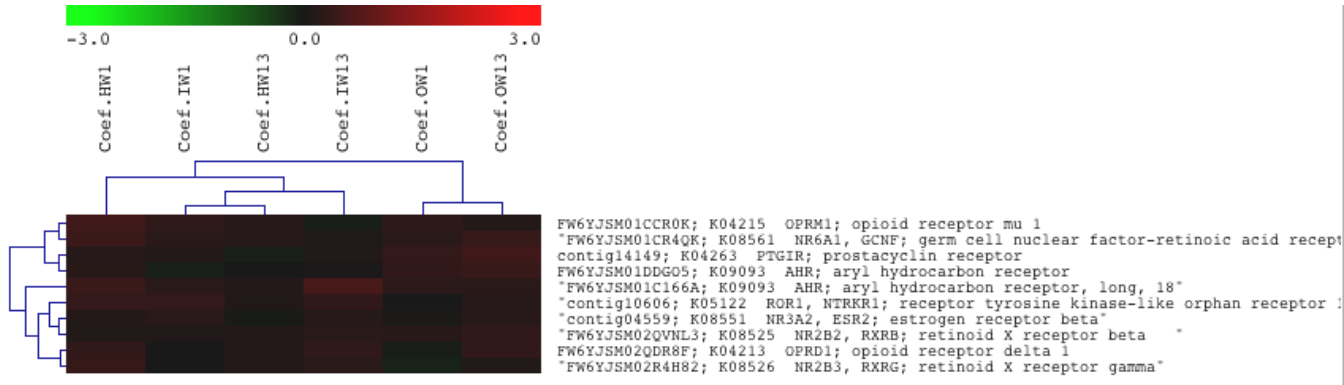
Cluster 7: RARA, AHR, RORC, RORB, ROR1, LXRB, PXR, ECR (ecdysone): some regulated oppositely in early and late for O and H. Especially O. AHR, Ecdyson, LXRB, PXR- increasingly up-regulated



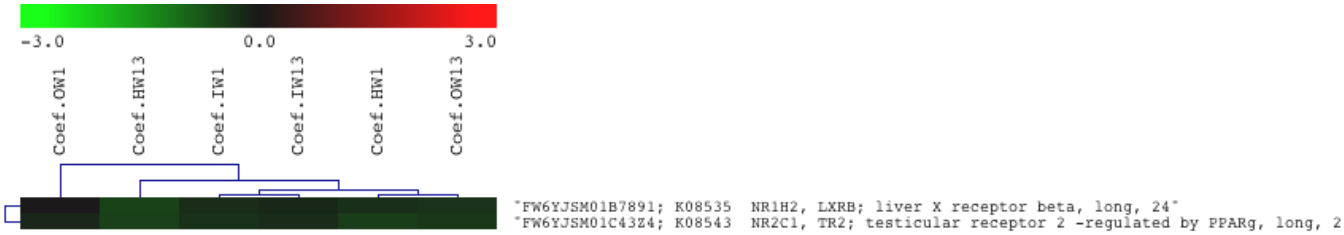
Cluster 6: ROR1, ROR2, ESRRA, ARNTL2, PGR, FFAR2, SF1, NR6A1: Increasingly up-regulated with time



Cluster5: Same, but more constant up-regulation with time (PPARG, ESR2, AHR, PTGIR…



Cluster9: LXR, TR2-down



- Appendix 3.4: Representation of homologous to SRY that are found in Ciona; red code for SOX9.

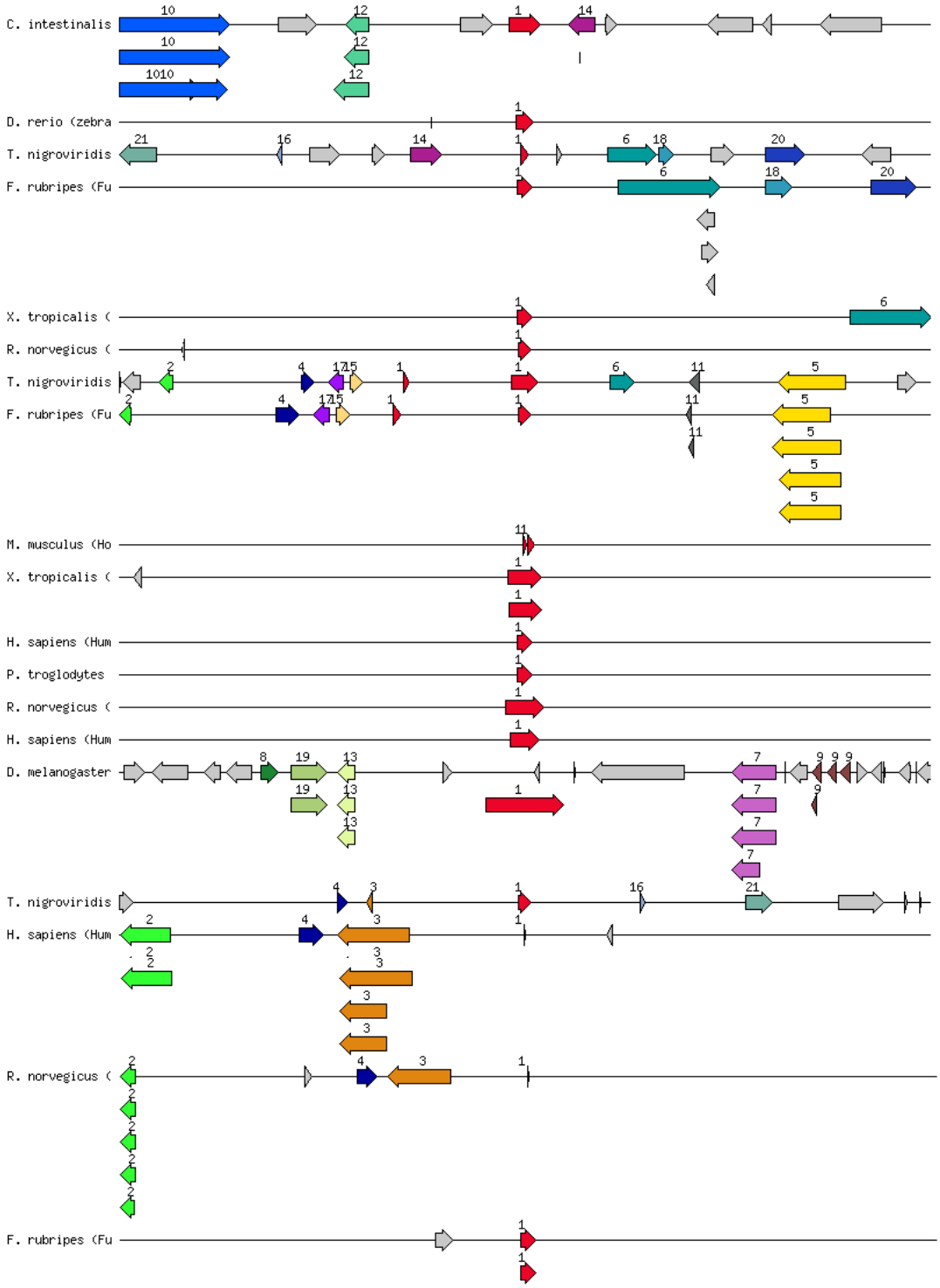


Fig1. Red_1-SOX9, 20-peroxisomal protein, 2-opioid receptor, 4-regulator of G-protein signaling.

- Appendix 3.5: Conservative motifs in the promoter region of TF binding sites of SOX9.

Transcription factor binding sites:

Coservative motifs in 500bp promoter region (MEME http://meme.nbcr.net/meme4_5_0/cgi-bin/meme.cgi)

Motif1: n the '**GTTTT**' repeat that abolishes SRY **binding sites** !!!

2	+	80	9.41e-07	GTCACATGTG GTTTTCCCCT CAAACCAGTG
rat	+	122	9.41e-07	GCTCCAGCCA GTTTTCCCCT GGCAGGTGTA
3	+	435	3.03e-06	GCAGAGACAA GGTTTCCCCT TAGGT
4	-	223	3.88e-06	TATTCCGTTA GTTTTTCCCT TTGTCGGCGG
1	+	342	9.12e-06	CCAGCAATAC GCTTGCCCCT ACAACCTAAC

Click on any row to highlight sequence in all motifs.

Name	Strand	Start	*p*-value	Sites
2	+	80	9.41e-07	GTCACATGTG GTTTTCCCCT CAAACCAGTG
rat	+	122	9.41e-07	GCTCCAGCCA GTTTTCCCCT GGCAGGTGTA
3	+	435	3.03e-06	GCAGAGACAA GGTTTCCCCT TAGGT
4	-	223	3.88e-06	TATTCCGTTA GTTTTTCCCT TTGTCGGCGG
1	+	342	9.12e-06	CCAGCAATAC GCTTGCCCCT ACAACCTAAC

http://www.pnas.org/content/90/3/1097.full.pdf

rat,_1-ciona,_2-Tetradion,_3-xenopus,_4-drosophila

Motif2: PPARs?? (TGA..TGA) Should be TGACCT for PPARG, 5'-AGGTCAAAGGTCA-3'-fr PPAR/RXR binding..

Name	Strand	Start	*p*-value	Sites
3	+	185	5.40e-12	CCAAAACCAT TGGTGGTAGATATTATTTCAGCTACTGA AGCCTAAATT
2	+	173	1.53e-11	CAGTTGGAGT TTTTAGGTGTTATTTAGCCAAAACCTGA CTTTGATTCG
4	-	333	7.53e-11	GTTGTTGTAG TTGTTGCTGTTGTTTTTGCTGTTGCTGA CGTCGGTGGT
1	-	456	1.17e-10	GCTTTGGCTG TTTCCGTTGATTTTCTGCTACGTACTGA AGTATTATTG
rat	+	319	1.18e-09	ACTGGGTAGA TCTCCAGAGCTATTAACGCAGCAGGTTA GGACTAACGG

Name	Strand	Start	*p*-value	Sites
3	+	185	5.40e-12	CCAAAACCAT TGGTGGTAGATATTATTTCAGCTACTGA AGCCTAAATT
2	+	173	1.53e-11	CAGTTGGAGT TTTTAGGTGTTATTTAGCCAAAACCTGA CTTTGATTCG
4	-	333	7.53e-11	GTTGTTGTAG TTGTTGCTGTTGTTTTTGCTGTTGCTGA CGTCGGTGGT
1	-	456	1.17e-10	GCTTTGGCTG TTTCCGTTGATTTTCTGCTACGTACTGA AGTATTATTG
rat	+	319	1.18e-09	ACTGGGTAGA TCTCCAGAGCTATTAACGCAGCAGGTTA GGACTAACGG

MOtif 3

Name	Strand	Start	*p*-value	Sites
3	+	386	4.64e-08	AGACGCTGAT TGGTTGGTGCAGTTA TAGAGGCCTA
rat	-	9	1.30e-07	TACTAGTCCA AGGTTGTTCACGTCA ACATACTAG
1	-	256	1.83e-07	TATGGTGACG TGGTTGATGACGTAA CAGTAGAGTA
4	-	385	3.68e-07	GGTGACGTTG ATGTTGTTGCAGTTG CAGTTGTTGC
2	-	108	4.68e-07	CCCGGACTAC AGGCTGCTGCTGTCA GATGTCAGCA

Name	Strand	Start	*p*-value	Sites
3	+	386	4.64e-08	AGACGCTGAT TGGTTGGTGCAGTTA TAGAGGCCTA
rat	-	9	1.30e-07	TACTAGTCCA AGGTTGTTCACGTCA ACATACTAG
1	-	256	1.83e-07	TATGGTGACG TGGTTGATGACGTAA CAGTAGAGTA

4	-	385	3.68e-07	GGTGACGTTG ATGTTGTTGCAGTTG CAGTTGTTGC
2	-	108	4.68e-07	CCCGGACTAC AGGCTGCTGCTGTCA GATGTCAGCA

Motif4 (steroid receptor TR2-like?)

Name	Strand	Start	*p*-value	Sites
4	-	66	2.18e-04	CCACGGCATG TGTGAA TTTGCGTCCA
3	+	96	2.18e-04	ATGGGAATAC TGTGAA GGTTCTTATC
2	+	138	2.18e-04	CCGGGTAGTG TGTGAA CGCCACTGGG
1	-	207	2.18e-04	TGAGACCAGC TGTGAA AATGGACGGA
rat	-	35	2.18e-04	TATTGCTCAC TGTGAA GTACTAGTCC

MOtif5 Name	Strand	Start	*p*-value	Sites
4	+	294	1.03e-10	AGCCAACAAC AACAACAACAACACCAATTT CAACCTGCCA
1	+	390	1.68e-09	TCTTGCAGTA AACGACCAAACCAGCTATTT ATCCAGCTAT
3	-	127	1.43e-08	TAGCCTTGGC AACAGCAAGTGCACTTATAT GCCATGATCT
rat	+	44	9.84e-08	CTTCACAGTG AGCAATATCTCCAGACATTT GAATTTCAGC
2	+	476	2.07e-07	AGACCTGCAC AACGCAGAGCTCAGCAAAAT GCTG

Name	Strand	Start	*p*-value	Sites
1	+	12	6.04e-10	GCCTAATAAT GCGGGTAGCGGTACCTTGCCC TACACGATAC
rat	-	414	4.74e-09	GGACCCTGAG GCGGGGATGGGAAGCTTGCCT GGAGATGCCC
2	+	323	5.24e-09	TGAGTGTCCT CCGGGAGACGGAGATTTGCCC GACGGACACG
4	-	111	2.34e-08	CTGATGTCTG CCGGGCCGAGCATTTTTGCCC AGTTACAGTT
3	-	359	5.84e-08	ACCAATCAGC GTCTGGAGCGGAGTCTAACCC CATAGAGGTG

Name	Strand	Start	*p*-value	Sites
rat	+	265	3.44e-07	TGGCCAGGTC CATAAAGAAAA GGGCCCGCGG
4	+	28	1.05e-06	GGCTCTGAGC CACAAACAAAA CAAAATCTGA
3	+	225	8.11e-06	CCTAAATTCA TATAAAAAAAA ATACAACAAA
1	+	443	1.06e-05	CCGACACCAT CACCAATAATA CTTCAGTACG
2	+	28	1.87e-05	CCCACCCGCG CACCAGGAAAG ACTGAATCAT

Motif8

Name	Strand	Start	*p*-value	Sites
3	+	47	4.16e-08	AGTACATATA TATGAAGATGTATAT GAATGGCTGA
4	+	460	6.46e-08	GGAACCCTCA TATGATCACGAACAT CCACATCATC
1	-	278	1.15e-07	GAGCCACTGA CTTGATGACGTATAT GGTGACGTGG
2	-	407	3.65e-07	TGGCCCAAAC CATGAACGCGTTCAT CGGCCGTCTG
rat	-	66	3.20e-06	CTCGGGGAAA TAGGACGGCTGAAAT TCAAATGTCT

Motif9

Name	Strand	Start	*p*-value	Sites
4	+	437	1.79e-06	CATCGCCGTT GCCGTATAAA AATGGAACCC
1	+	153	2.83e-06	CTTAACTCGG GCCGTATCAG GGTCATCACC
2	-	203	6.50e-06	CTTCTCCGAC GCCGAATCAA AGTCAGGTTT
rat	+	466	1.51e-05	GTGTCCCGTG GCCGTAGCAG AGCGCGGGGT
3	+	26	2.61e-05	AAACCCAAGT GCTGTATATA CAGTACATAT

Name	Strand	Start	*p*-value	Sites
4	+	437	1.79e-06	CATCGCCGTT GCCGTATAAA AATGGAACCC
1	+	153	2.83e-06	CTTAACTCGG GCCGTATCAG GGTCATCACC
2	-	203	6.50e-06	CTTCTCCGAC GCCGAATCAA AGTCAGGTTT

Name	Strand	Start	*p*-value	Sites
4	-	15	2.42e-04	GTGGCTCAGA GCCTAA CCCACATTTT
3	+	214	2.42e-04	AGCTACTGAA GCCTAA ATTCATATAA
1	+	2	2.42e-04	CC GCCTAA TAATGCGGGT
rat	+	216	2.42e-04	CTGGGCGGGG GCCTAA GGACTAGGGT
2	+	428	4.84e-04	CATGGTTTGG GCCAAA GACGAGCGCA

CHAPTER 5.

- Appendix 5.1 - Experiment 1: ontogenetic changes in shell morphology.

N. lapillus was sampled from four shores in North Wales (Fig. S1). Two shores (Cable Bay and Ravens Point) represented habitat exposed to strong wave action generated by prevailing south-westerly winds blowing from the English Channel and across the Irish Sea, while the other two (Menai Bridge, Llanfairfechan) represented more sheltered habitat in the lee of the prevailing winds. Geographical coordinates and an index combining mean annual wind energy and fetch to quantify wave exposure (Thomas 1986) are given in the legend to Fig. S1.

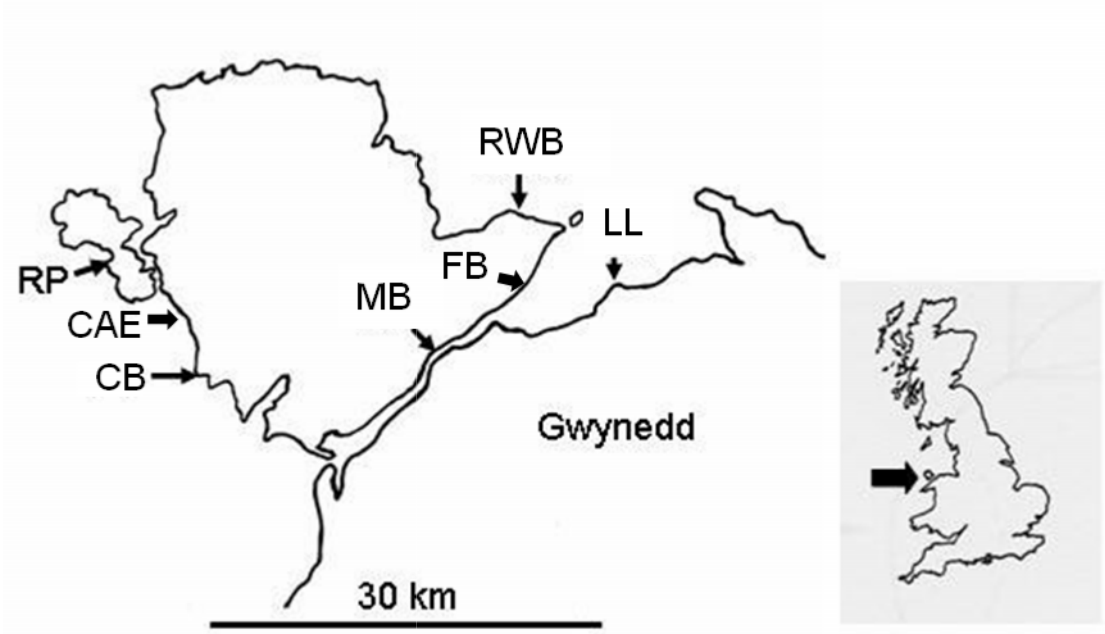


Figure S1. Location of study sites along the North Wales coastline. CB = Cable Bay (53° 12.410´ N, 04° 30.290´ W), Thomas Exposure Index (TEI) = 14; CAE = Caethle (53° 11.212´ N, 04° 30.249´ W), TEI = 15; FB = Friars Bay (53° 16.107´ N, 04° 05.113´ W), TEI = 3; LL = Llanfairfechan (53° 15.769´ N, 03° 55.142´ W), TEI = 2; MB = Menai Bridge (53° 13.272´ N, 04° 09.861´ W), TEI = 0; RP = Ravens Point (53° 16.161´ N, 04° 37.548´ W), TEI = 14; RWB = Redwharf Bay (53° 18.594´ N, 04° 08.495´ W), TEI = 8.

Aggregations of spawning adults were located at low-spring-tide level in February 1997. Samples of 50 adults and about 100 egg capsules were collected per shore. Care was taken to collect egg capsules from several different positions within a clump, in order to increase the likelihood of obtaining progeny from a number of parents. Adults were fixed in alcohol and egg capsules were transferred to laboratory aquaria measuring 60 x 40 x 30cm and dedicated one per population to avoid any possibility of mixing, as might occur using meshed cages within aquaria. Egg capsules were inaccessible at Ravens Point and so 100 adults were collected and maintained in an aquarium, where they spawned. To ensure absence of crab effluent, aquaria were provisioned with seawater obtained from Cable Bay North where crabs are absent intertidally. The seawater was renewed bimonthly, gently aerated and allowed to follow ambient temperature within a seasonal range of about 10-18°C. Evaporation was minimized by placing clear plastic sheeting over the aquaria. Small mussels, below 10mm in shell length, were provided as prey and renewed as needed to maintain unlimited availability to hatched snails. Aquaria were reshuffled at each water change to avoid position effects. Measurements were taken of shell length (truss 1-11 in Fig. 2B) and aperture external width (truss 7-8). Samples of 15 snails per population were fixed in alcohol and measured under a dissection microscope when they had reached a shell height of about 4mm, approximately 3 months after hatching. Independent, successive samples of 30 snails were fixed and measured with callipers to 0.05mm at 8, 12, 18 and 23 months after hatching. Ontogenetic changes in shell shape of native snails were quantified using samples of 50 individuals in successive size classes 1-5, 6-10, 11-15 and 16-20mm shell height collected in March 1998. Differentiation of shell morphology among snails from exposed and sheltered sites was discernible in the second size class corresponding to a shell length of 5-6mm and an age of 5-6mo, becoming more pronounced in successive size classes (Fig. S2).

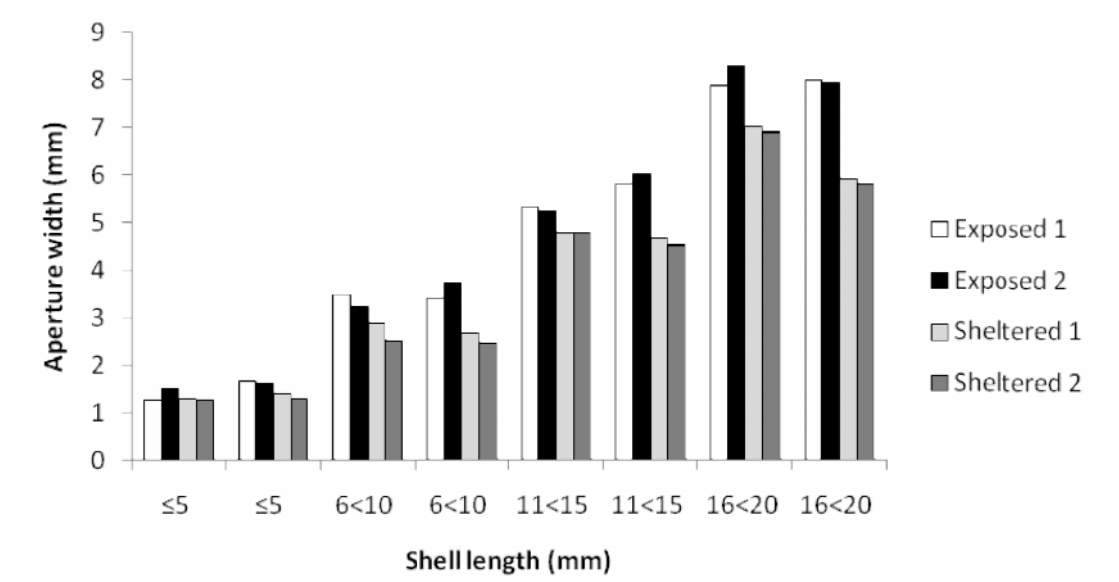


Figure S2. Experiment 1: ontogenetic changes in shell morphology. Mean aperture external width adjusted to shell length (ANCOVA). For each size class, the first group represents snails reared in the laboratory and the second group snails collected from the ancestral field-population. Sample size (N) and covariate value (CV) for adjusting mean aperture width were as follows. Group 1: N exposed 1 = 10, N exposed 2 = 15, N sheltered 1 = 11, N sheltered 2 = 12; CV = 2.264mm. Group 2: N exposed 1 = 11, N exposed 2 = 12, N sheltered 1 = 15, N sheltered 2 = 15; CV = 2.726mm. Group 3: N exposed 1 = 13, N exposed 2 = 14, N sheltered 1 = 12, N sheltered 2 = 6; CV = 7.915mm. Group 4: N exposed 1 = 15, N exposed 2 = 15, N sheltered 1 = 15, N sheltered 2 = 15; CV = 7.670mm. Group 5: N exposed 1 = 17, N exposed 2 = 17, N sheltered 1 = 15, N sheltered 2 = 18; CV = 12.331mm. Group 6: N exposed 1 = 15, N exposed 2 = 15, N sheltered 1 = 15, N sheltered 2 = 15; CV = 12.753mm. Group 7: N exposed 1 = 35, N exposed 2 = 30, N sheltered 1 = 28, N sheltered 2 = 22; CV = 18.066mm. Group 8: N exposed 1 = 39, N exposed 2 = 30, N sheltered 1 = 30, N sheltered 2 = 30; CV = 16.872mm.

Because of the relatively small range of shell length (5mm) per group, log transformation was not used.

Experiment 2: wave-exposure gradient

To represent a greater range of exposure to wave action, thirty adults were collected in March 2002 from the shores used in experiment 1 plus two extra shores, Friars Bay and Caethle (Fig. S1). Each population was housed in a separate aquarium and fed ad lib on mussels until sufficient numbers of egg capsules had been laid. Egg capsules were transferred to 2 l plastic bottles filled with gently aerated seawater held at 15°C.

Two bottles were assigned to each population. Mussels <10mm shell length were provided ad lib as food. Snails were measured at 12 months, when they had grown to a shell length ≥16mm and some had begun to lay eggs. Samples of 20-35 snails were available per population. Relative aperture external width of laboratory-reared snails was correlated with that of their ancestral field populations (Fig. S3; Pearson r = 0.893, P = 0.017), indicating a heritable component of variation in shell morphology. Relative aperture external width was also positively ranked with wave exposure at the study sites (Fig. S3; laboratory snails, Spearman rho = 1.000; field snails, Spearman rho = 0.943, P = 0.005).

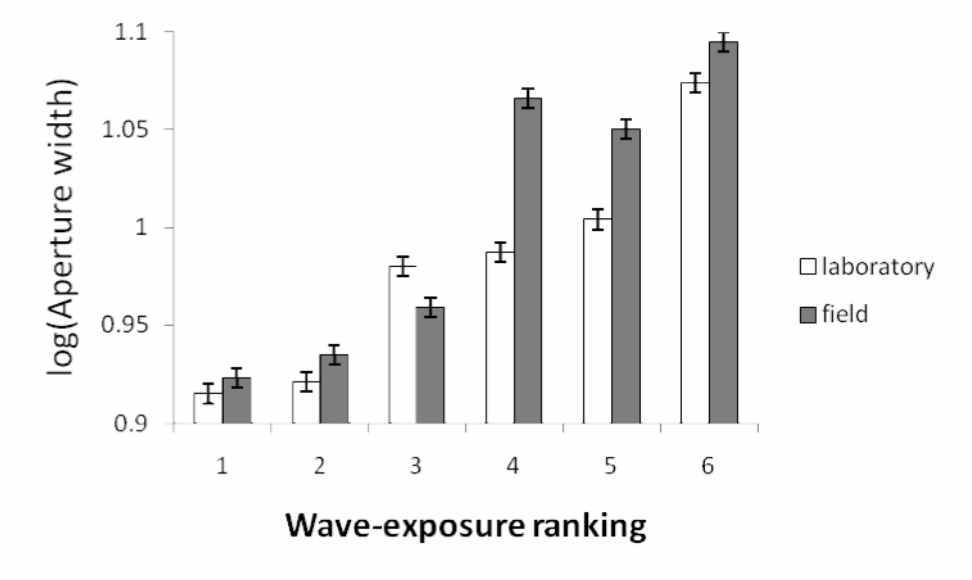


Figure S3. Comparison of relative aperture width of adult laboratory-reared and field-collected N. lapillus from shores differing in exposure to wave action. Mean aperture external width adjusted to shell length (ANCOVA of log-transformed data pooled from experiments 1 and 2). Shores are ranked in increasing order of wave exposure (see Fig. S1): Menai Bridge (Thomas exposure index (TEI) = 0; Llanfairfechan, TEI = 1; Friars Bay, TEI = 4; Redwharf Bay, TEI = 8; Ravens Point, TEI = 13; Cable Bay, TEI = 14; Caethle, TEI = 15. Standard errors are shown.

- Appendix 5.2 - Nucella lapillus karyotype

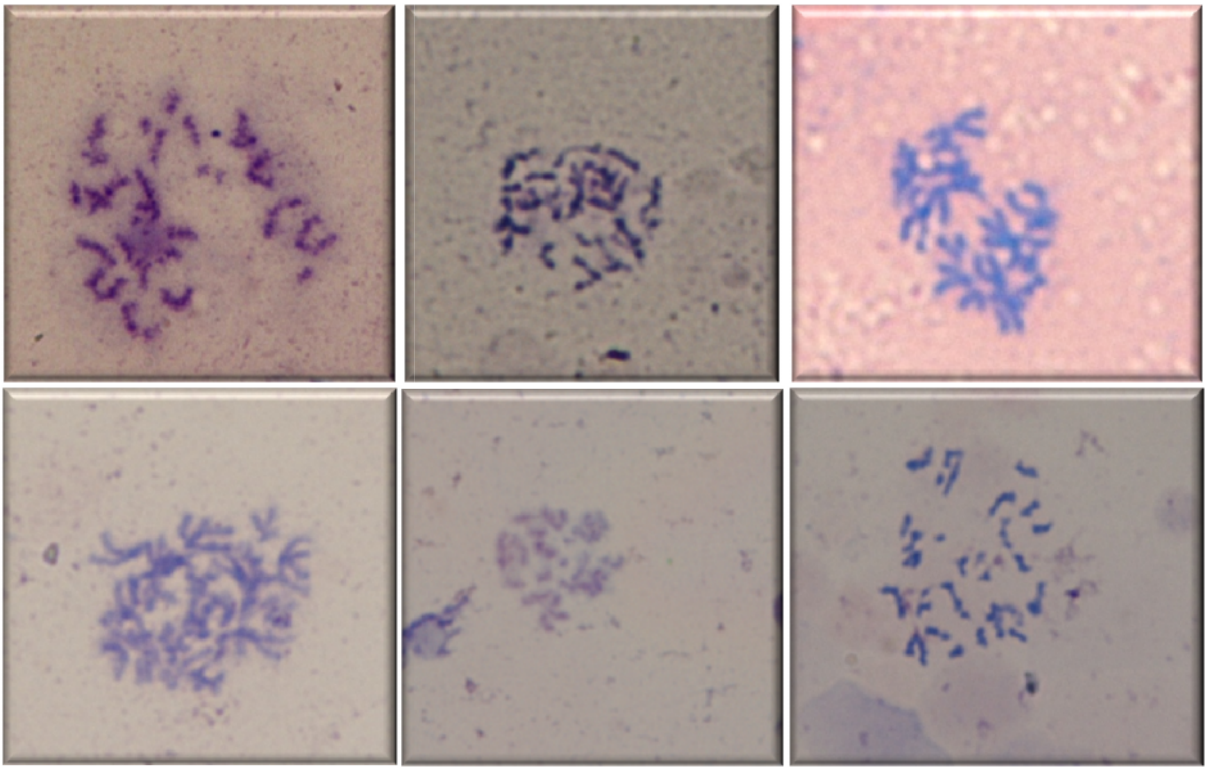


Figure S4. Nucella lapillus karyotype pictures; top: Llanfairfechan (sheltered); bottom: Cable Bay (exposed).