SARA DE JESUS DIAS CARDOSO

RNA-SEQ APPLIED TO THE PEACOCK BLENNY *Salaria pavo*: UNVEILING THE GENE NETWORKS AND SIGNALLING PATHWAYS BEHIND PHENOTYPIC PLASTICITY IN A LITTORAL FISH

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

Faro / Portugal

2019

SARA DE JESUS DIAS CARDOSO

RNA-SEQ APPLIED TO THE PEACOCK BLENNY *Salaria pavo*: UNVEILING THE GENE NETWORKS AND SIGNALLING PATHWAYS BEHIND PHENOTYPIC PLASTICITY IN A LITTORAL FISH

Doutoramento em Ciências Biológicas

(Especialidade em Ecologia Molecular)

Trabalho efetuado sob a orientação de:

Prof. Dr. Rui Filipe Oliveira

Prof. Dr. Adelino Vicente Canário

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

Faro / Portugal

2019

DECLARAÇÃO DE AUTORIA DE TRABALHO

RNA-SEQ APPLIED TO THE PEACOCK BLENNY *Salaria pavo*: UNVEILING THE GENE NETWORKS AND SIGNALLING PATHWAYS BEHIND PHENOTYPIC PLASTICITY IN A LITTORAL FISH

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

I declare to be the author of this work, which is original and unpublished. Authors and works consulted are duly cited in the text and are included in the list of references.

THIS THESIS WAS SUPPORTED BY

The work presented here was developed at Instituto Gulbenkian de Ciência (IGC) in Oeiras, with the support of both ISPA – Instituto Universitário in Lisbon, for the maintenance of live fish, and Centro de Ciências do Mar (CCMAR) at Universidade do Algarve, for logistics and support during fieldwork in Ria Formosa.

This thesis was supported by national funds through project research grants PTDC/MAR/69749/2006 and EXCL/BIA-ANM/0549/2012 from the Portuguese Foundation for Science and Technology (FCT) and by an individual Ph.D. fellowship (SFRH/BD/89072/2012) from FCT.

ACKNOWLEDGEMENTS

Throughout this work, I crossed paths with many people that had in some way an impact in my work and personal life that will be difficult to mention everyone at this point, but I will endeavour to mention them all.

First, I would like to start by acknowledging both my supervisors, Prof. Dr Rui Oliveira and Prof. Dr Adelino Vicente. They were my role models during the development of this work, and their patience and support during this period was essential to my growth as a researcher and the completion of this thesis. I would also like to include in this paragraph, Prof. Dr David Gonçalves, my nonofficial supervisor, which introduced me both to the amazing species used throughout this work, the peacock blenny, and to my present supervisors. To all of you thank so much for the knowledge and opportunities you presented to me.

Second, I would like to express my gratitude to the former and present members of the Integrative Behavioural Biology group (IBBG), for their support and always-helpful feedback on my work, specially to Ana Faustino, Ana Félix, André Monteiro, António Roleira, Diogo Ribeiro, Filipe Espigares, Gonçalo Oliveira, Ibukun Akrinidade, João Sollari, Miguel Simões, Olinda Almeida, Rodrigo Abreu, Sílvia Costa, and lastly my "big supporters" during this last year of my work in the group, Ana Rita Nunes, Júlia Pinho and Magda Teles. Beyond the professional accomplishment, this thesis gave me the opportunity to meet and work alongside Magda Teles, a true friend for life! Thank you for your advice, help and friendship when it mattered most.

To my companion during field and lab work, Pedro Vieira, each one of us pursuing different mechanisms underlying behavioural plasticity in the peacock blenny. The task was not easy, but we carried on!

To João Saraiva, the "rough guide" to everything I had to know about the wild populations sampled for the genomic analysis in this thesis. Thank you so much for taking some of your time to go with me to each location and help setting up the laboratories backing up the fieldwork, especially in Trieste with Professor Serena Fonda. Your work on the peacock blenny was one of the pillars in my thesis!

Third, I would like to acknowledge the PhD Programme in Biological Sciences of the University of Algarve, and my host institution, Instuto Gulbenkian the Ciência (IGC). During the four years of the PhD programme, the support given by Paula Caboz and from the programme itself was essential to my work, and the insights exchanged with my colleagues invaluable. On the other hand, IGC is a wonderful place to work in, the support the institute gives to their students is commendable, and I am grateful for it. Additionally, I would like to mention some of the people from the different units that made all the resources essential to my work available: the IT and Bioinformatics Unit, João Garcia and Daniel Sobral that allowed me to use the high-performance cluster; the Genomics unit, João Sobral, for the critical insight during quality check of my samples; and to Pedro Fernandes for his essential courses on Bioinformatics. I also cannot forget my "office mates", for their evolutionary perspectives and comments during my analyses, João

Proença, Nélson Frazão, Paulo Durão, Ricardo Ramiro and Roberto Soria. I am very fortunate to have had the opportunity to work in both places, alongside all of these people.

Fourth, I would like to mention Dr Alexander Goesman, from the Justus-Liebig-Universität, Germany, and Dr Alison Wright, from the Sheffield University, UK, whose labs I had the opportunity to visit and work in at different stages of my thesis and that were essential to give new insights to my work.

To my parents and my brother, words at this point are not enough to describe how much you mean to me. With all my love, this thesis is dedicated to you!

To my grandparents,

Maria de Lurdes and Eduardo

ABSTRACT

Phenotypic plasticity is the ability of an individual genome to produce different phenotypes depending on environmental cues. These plastic responses rely on diverse genomic mechanisms and allow an organism to maximize its fitness in a variety of social and physical settings. The development of next-generation sequencing (NGS) technologies, especially RNA Sequencing (RNA-Seq), has made it possible to investigate the distinct patterns of gene expression known to be underlying plastic phenotypes in species with ecological interest. In teleost fishes, changes in phenotypes is often observed during the reproductive season, with shifts and adjustments in dominance status that can lead to the co-existence of multiple reproductive morphs within the same population. One such example is the peacock blenny *Salaria pavo* (Risso, 1810), a species where the intensity of mating competition varies among populations due to nest-site availability, such that two different levels of plasticity arise: 1) intraspecific variation in reproductive behaviour for males that can follow either of two developmental pathways, grow directly into nest-holder males, or behave first as female mimics to sneak fertilizations (sneaker males) and later transition into nest-holder males, and 2) inter-population variation in courting roles of females and nest-holder males. This system provides the ideal basis to apply RNA-Seq methods to study plasticity since differences in reproductive traits within and among populations can reveal which genetic and genomic mechanisms underpin the observed variation in behavioural response to changes in the social environment. However, the genomic information available for this species was scarce, and hence multiple sequencing techniques were used and the methodologies applied optimized throughout the work. In this thesis, we start by first obtaining a *de novo* transcriptome assembly to develop the first genetic markers for this species (Chapter 2). These microsatellites were used to elucidate the reproductive success (i.e. consisting of mating success and fertilization success) of male ARTs, which can be used as a proxy of Darwinian fitness (Chapter 3). In this study, we detected a fertilization success for nest-holder males of 95%, and showed a stronger influence of the social environment rather than morphological variables in the proportion of lost fertilizations by nest-holder males of this species. Taking advantage of the developed transcriptome, we used high-throughput sequencing to obtain expression profiles for male morphs (i.e. intraspecific variation) and females in this species, and focus on the role of differential gene expression

in the evolution of sequential alternative reproductive tactics (ARTs) that involve the expression of both male and female traits (Chapter 4). Additionally, we show how the distinct behavioural repertoires are facilitated by distinct neurogenomic states, which discriminate not only sex but also male morphs. Lastly, using two different target tissues, gonads and forebrain, we focus on the genomic regulation of sex roles in courtship behaviour between females and males from two populations under different selective regimes (inter-population variation), the Portuguese coastal population with reversed sex roles and the rocky Italian population with 'conventional' sex roles (Chapter 5). Here we demonstrate that variation in gene expression at the brain level segregates individuals by population rather than by sex, indicating that plasticity in behaviour across populations drives variation in neurogenomic expression. On the other hand, at the gonad level, variation in gene expression segregates individuals by sex and then by population, indicating that sexual selection is also acting at the intrasexual level, particularly in nest-holder males by paralleling differences in gonadal investment. However, the genomic mechanisms underlying courtship behaviour were not fully elucidated, and more studies are necessary.

Keywords: RNA-Seq; phenotypic plasticity, alternative reproductive tactics, courtship roles; *Salaria pavo*

RESUMO ALARGADO

A plasticidade fenotípica consiste na capacidade de o mesmo genoma produzir diferentes fenótipos comportamentais dependendo das pistas ambientais recebidas. Estas respostas plásticas dependem de diversos mecanismos genómicos e permitem que o indivíduo maximize a sua fitness (aptidão) numa variedade de ambientes ecológicos. O desenvolvimento verificado nas tecnologias de sequenciação de alto desempenho ao longo da última década, globalmente denominadas de "Next Generation Sequencing" (NGS), permitiu o estabelecimento de métodos de análise e ferramentas genómicas que podem ser aplicadas em todos os sistemas ecológicos de interesse em biologia, sem a existência prévia de um genoma curado. Nomeadamente a tecnologia de sequenciação de ARN, conhecida globalmente como RNA-Seq, tornou possível a investigação dos perfis de expressão génica que se sabe serem determinantes na emergência de fenótipos plásticos, e consequentemente permitem determinar fenótipos em estados distintos de expressão genómica. Em peixes teleósteos, é possível observar com frequência modificações no fenótipo comportamental durante o período de reprodução, como por exemplo alterações e ajustes no estatuto de dominância que podem levar à coexistência de indivíduos que apresentam diferentes táticas de reprodução dentro da mesma população. Um desses exemplos é o peixe marachomba-pavão *Salaria pavo* (Risso, 1810), onde a intensidade na competição intra e intersexual varia entre populações sendo modulada pela disponibilidade de locais de nidificação, de forma a que dois níveis diferentes de plasticidade surgem: 1) variação intraespecífica no comportamento reprodutivo em machos que podem seguir uma de duas vias de desenvolvimento, investirem no seu crescimento e tornarem-se machos nidificantes na sua primeira época de reprodução, ou primeiro seguir uma tática de macho parasita onde investem em fertilizações furtivas, sendo que mais tarde no seu desenvolvimento fazem a transição para macho nidificante, e 2) variação interpopulacional nos papeis de corte de fêmeas e machos nidificantes. Os machos parasitas, conhecidos nesta espécie como "sneakers", possuem uma particularidade que os tornam singulares, para além de imitarem a morfologia das fêmeas também conseguem imitar o seu comportamento de corte direcionado ao macho nidificante, o que lhes permite aproximarem-se discretamente dos ninhos dos machos e fertilizar parte dos ovos que as fêmeas depositam. Este sistema constitui a base ideal para aplicar métodos de RNA-Seq e estudar esta plasticidade

fenotípica, uma vez que diferenças nas características reprodutivas dentro e entre populações podem revelar quais os mecanismos genéticos e genómicos subjacentes à variação observada em resposta a mudanças no ambiente ecológico. No entanto, a informação genómica disponível nesta espécie é reduzida e, por isso diferentes técnicas de sequenciação, assim como diferentes métodos de análise foram usados e otimizados ao longo deste trabalho. A presente tese é constituída por quatro trabalhos, sendo que no primeiro estudo se começa pela sequenciação de uma biblioteca de ARN proveniente de uma mistura de múltiplos indivíduos e de tecidos, de forma a captar a diversidade genética e desenvolver os primeiros marcadores genéticos nesta espécie (Capítulo 2). Com base nestes marcadores, microssatélites polimórficos, foi possível genotipar uma fração dos indivíduos da população existente na Ilha da Culatra (Ria Formosa, Portugal) bem como os ovos retirados de ninhos alvo, de forma a fazer análises de paternidade (Capítulo 3). Neste estudo, foi possível estimar o sucesso de fertilização de ovos de cada uma das táticas alternativas de reprodução, e usá-la como medida representativa de fitness de cada tática alternativa de reprodução nesta espécie. Os resultados indicam um sucesso de fertilização para os machos nidificantes de 95%, e mostramos que existe uma maior influência do ambiente social do que de variáveis morfológicas na proporção de fertilizações não obtidas pelos machos nidificantes, quando comparado com estudos anteriores. Usando o transcriptoma obtido no primeiro trabalho, avançámos com a caraterização genómica de cada um dos fenótipos presentas na população da ilha da Culatra, fêmeas, machos nidificantes, machos sneakers e machos de transição (machos que apenas investem no seu crescimento, não se reproduzindo, e consequente transição de sneaker para macho nidificante) (Capítulo 4). Para tal, foi sequenciado em profundidade o transcriptoma de cérebro de cada um deste fenótipos, e os perfis de expressão obtidos para machos e fêmeas desta espécie, onde o foco do estudo se centrava no papel da expressão génica diferencial na evolução de táticas reprodutivas alternativas sequenciais que envolvem a expressão de ambos os traços masculinos e femininos. Os resultados obtidos, mostram como repertórios comportamentais distintos são facilitados por estados neurogenómicos distintos, que discriminam não apenas o sexo, mas também as táticas alternativas de reprodução. Por fim, utilizando dois tecidos-alvo, gónadas e prosencéfalo, focámo-nos na regulação genómica dos papeis sexuais no comportamento de corte entre fêmeas e machos nidificantes de duas populações sob diferentes regimes seletivos, a população costeira portuguesa com papeis sexuais invertidos e a população rochosa italiana, com papeis sexuais ‘convencionais’ (Capítulo 5). Os resultados obtidos

mostram que ao nível do cérebro, a variação na expressão génica segrega os indivíduos por população e não por sexo, indicando que a plasticidade no comportamento entre as populações induz uma maior variação na expressão neurogenómica. Por outro lado, ao nível das gónadas, a variação na expressão génica segrega os indivíduos por sexo e também por população, indicando que a seleção sexual está a atuar ao nível intrasexual, particularmente nos machos nidificantes, indo de encontro a diferenças detetadas entre populações no investimento alocado às gónadas. No entanto, os mecanismos genómicos subjacentes ao comportamento de corte não foram totalmente elucidados, e mais estudos são necessários.

Palavras chave: RNA-Seq; plasticidade fenotípica; táticas alternativas de reprodução; comportamento de corte; *Salaria pavo*

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 | GENERAL INTRODUCTION

CHAPTER 3 | PATERNITY ASSESSMENT USING GENETIC MARKERS

CHAPTER 4 | NEUROGENOMIC PATTERNS OF ARTs

CHAPTER 5 | GENOMIC PATTERNS OF PLASTIC SEX ROLES

APPENDIX I | NEUROGENOMIC MECHANISMS OF SOCIAL PLASTICITY

APPENDIX IV | SUPPLEMENTARY INFORMATION CHAPTER 4

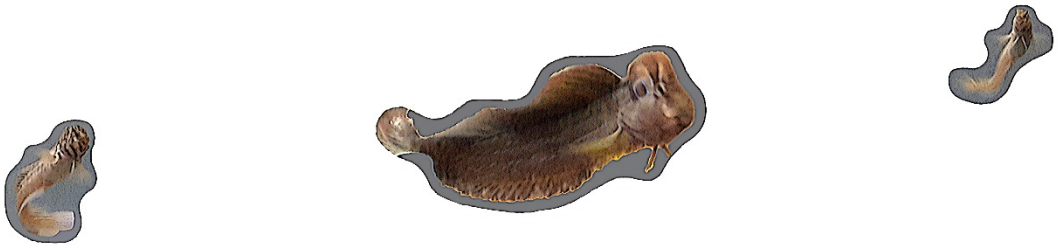
APPENDIX V | SUPPLEMENTARY INFORMATION CHAPTER 5

LIST OF TABLES

APPENDIX III | SUPPLEMENTARY INFORMATION CHAPTER 3

APPENDIX IV | SUPPLEMENTARY INFORMATION CHAPTER 4

APPENDIX V | SUPPLEMENTARY INFORMATION CHAPTER 5



CHAPTER 1 |

GENERAL INTRODUCTION

GENERAL INTRODUCTION

1.1. Chapter summary

Phenotypic plasticity is known to be generally characterized by distinct patterns of gene expression. For a long time microarrays were the primary technology to obtain gene expression levels at a large-scale, with the significant drawback of being limited to a few model species. Over the last decade, RNA-Sequencing (RNA-Seq) has become the tool of elation among molecular ecologists. Unlike microarrays, RNA-Seq expression is not limited to a few species but instead it allows the quantification of gene expression against a *de novo* transcriptome. Many different methods have been developed to identify differentially expressed genes in these RNA-Seq datasets, although no clear consensus exists on which perform better. In this thesis, I used two different high-throughput sequencing technologies and implemented customized workflows of analysis in order to explore genomic expression patterns underlying phenotypic plasticity in the non-model species peacock blenny *Salaria pavo*.

1.2. A new era in ecology – ecological genomics

The ability to adapt to environmental change is a ubiquitous characteristic of biological systems. Adaptation can be achieved broadly by two mechanisms: 1) natural selection acting on heritable phenotypic variation produced by genetic variation (e.g. mutation, recombination) (Darwin, 1859; Wallace, 1858), or 2) through adaptive change without genetic mutation occurring in situations where the rate of environmental change outpaces the rate of genetic evolutionary change (Pigliucci, 2001; West-Eberhard, 2003). In the second scenario, the evolution of phenotypic plasticity is favoured, according to which environmental cues sensed by the organism lead the same genotype to produce different phenotypes depending on environmental conditions cues (Pigliucci, 2001; West-Eberhard, 2003). Different traits show different evolutionary changes in plasticity in terms of the time lag to respond to the environmental cue and the magnitude of the response. Among animals, behavioural traits exhibit both more rapid and stronger plasticity than morphological traits, which makes behavioural plasticity a key adaptive response to changing environmental conditions (Pigliucci, 2001). [adapted from (Cardoso et al., 2015), Appendix I]

Behavioural plasticity depends on the development of a central nervous system that allows for rapid and integrated organismal responses in order to maintain homeostasis (or allostasis). An individual's ability to be behaviourally plastic relies on its aptitude to use information contained in both the social (i.e. conspecifics with which it shares space and/or resources) and physical environment (i.e. personal information) (Dall et al., 2005). The social domain is considered one the most ambiguous components of the environment as it is made of other behavioural agents with an inherent level of unpredictability of their actions, with whom the individual needs to interact. The latter is also crucial in animals pace of life, and there is ample evidence demonstrating that animals use information from the physical environment to make behavioural decisions (e.g. seasonal variation in photoperiod and the availability of resources have been shown to trigger mating behaviours (Fudickar et al., 2016)). Together both components of the environment act on individuals so that they may regulate the expression of their behaviour, to adapt their behavioural output to specific situations in a complex and variable social world, and these are expected to depend on the evolution of plastic responses (Fig. 1). These plastic responses rely on diverse genomic mechanisms and allow the same genotype to produce different behavioural phenotypes (i.e. phenotypic plasticity), and hence social plasticity should be viewed as a key ecological performance trait that impacts Darwinian fitness (Oliveira, 2009; Taborsky and Oliveira, 2012). [adapted from (Cardoso et al., 2015), Appendix I]

Genomic mechanisms of social plastic phenotypes can depend on several proximate pathways organized in at least six levels (Fig. 1). Specific behavioural states or social phenotypes (*sensu* (Zayed and Robinson, 2012)), are paralleled by specific neural states of a social decision-making network in the brain (i.e. connectome). This network is composed of two interconnected neural circuits, the social behaviour network (*sensu* (Newman, 1999); see also (Goodson, 2005)) and the mesolimbic reward circuit, that together regulate the expression of social behaviour (O'Connell and Hofmann, 2011, 2012). At the physiological level, most nodes of this social decision-making network express receptors for neuromodulators (i.e. neuropeptides and amines) and hormones (i.e. sex steroids and glucocorticoids) that modulate the state of the network (Oliveira, 2009). At the molecular level, the behavioural states correspond to specific neurogenomic states (*sensu* (Zayed and Robinson, 2012)) characterized by distinct patterns of gene expression across the social decision-making network in the brain, such that individuals in different behavioural states exhibit different brain transcriptomes. Differential gene expression in

the relevant neural network may change the weight of each node and/or the strength of the connectivity between them, therefore contributing to the generation of multiple network states. As a result, different neurogenomic states correspond to different behavioural states, and the switches between states are orchestrated by signalling pathways that interface the environment, social and ecological, and the genotype (Aubin-Horth and Renn, 2009; Taborsky and Oliveira, 2012). Furthermore, we now know that the link between genome and behaviour is more complex at the molecular level, since it includes other players such as epigenetic mechanisms establishing long-lasting and irreversible changes in behavioural states, different protein level thresholds and post-translational modifications, and the genome itself in its structure and content (for a review see (Landry and Aubin-Horth, 2014)). [adapted from (Cardoso et al., 2015), Appendix I]

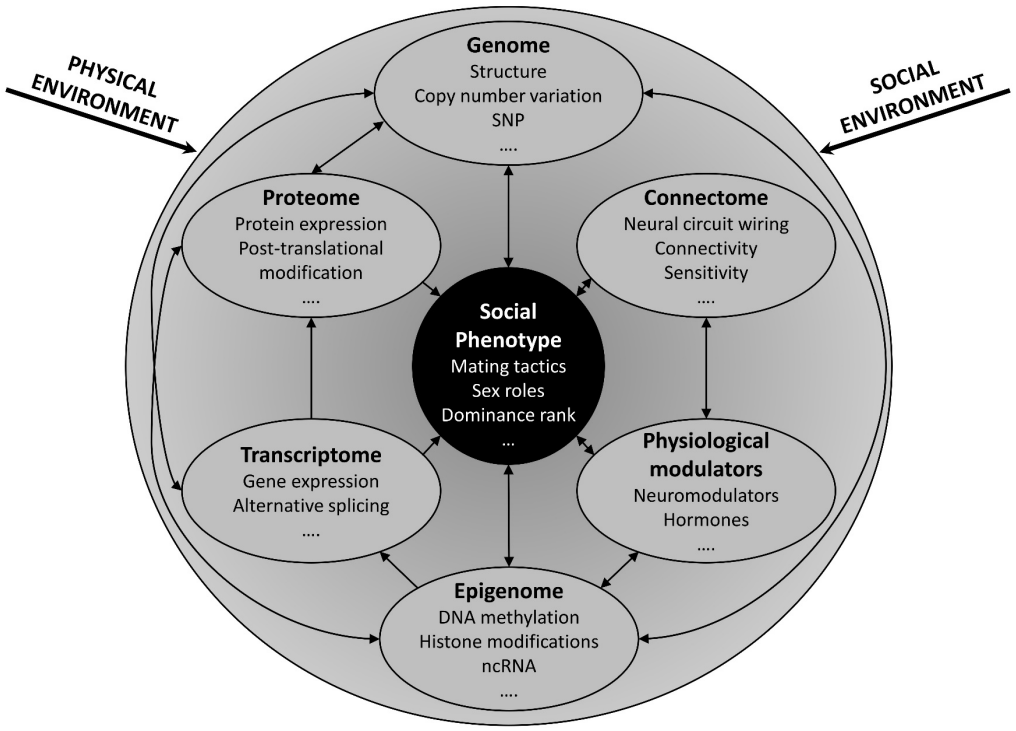


Fig. 1 – Diagram illustrating six levels where diverse proximate pathways may be underlying plastic social phenotypes, establishing the link between environment and genome.

Recent conceptual and technological developments allow dissecting the interaction between genomes and their environment over short and long time-scales. Over the last two decades, our understanding in molecular biology has grown exponentially with the advent of high-throughput genome profiling methods. Advances in DNA sequencing and the development of microarrays enabled genome-wide studies in comparative and

functional genomics. However, studies with DNA microarrays, while powerful, depend on prior knowledge of genome sequence and gene annotations. In some cases where a species-specific microarray was not available, microarrays developed for a related species have been used effectively (e.g. (Aubin-Horth et al., 2005b; Renn et al., 2004)). Although this approach enabled the molecular study in species lacking genomic information, heterologous hybridizations could introduce biases in the analysis and results that had to be controlled, since the hybridization ratio depends on the similarity of sequences between the platform species and of the heterologous species (Machado et al., 2009). This paradigm changed with the development of low cost, massively parallel DNA sequencing methods coined as next-generation sequencing (NGS) technologies (Metzker, 2010; Shendure and Ji, 2008; Wang et al., 2009). Diverse high-throughput applications have been developed so far that allow researchers to tackle the different proximate pathways mentioned above (Reuter et al., 2015). However, RNA sequencing (RNA-Seq) technology remains the preferred tool to detect broad patterns of gene expression without *a priori* genomic knowledge on the species of interest. This is especially true for non-model species whose genetic knowledge is fairly limited in most cases (for a review see (Ekblom and Galindo, 2011)), but that on the other hand offer additional flexibility and crucial ecological relevance for exploring the molecular basis of particular traits in greater detail (Parsons and Albertson, 2013). Several genomic (i.e. microarray and RNA-Seq) studies have demonstrated that different behavioural states, whether they occur or not within the same individual, are associated with different profiles of gene expression in the brain.

In this line of thought, species presenting alternative reproductive tactics (ARTs) have been investigated to explore the genomic mechanisms underlying plastic reproductive phenotypes. ARTs refers to discrete behavioural, morphological and physiological traits selected to maximize fitness in two or more alternative ways in the context of intrasexual competition. Individuals allocate resources to either one or the other (mutually exclusive) tactics, such as in the same population we can find a dominant male ('bourgeois' tactic), which invests in privileged access to mates (e.g. nest defense, secondary sexual characters, pheromones), and a sneaker male ('parasitic' tactic), which exploits the investments of bourgeois conspecifics (Oliveira et al., 2008). A descriptive classification of ARTs has been proposed by Oliveira (2006) that does not require the knowledge of their underlying evolutionary processes (i.e. genetic vs. conditional responses; (Gross, 1996)), as either fixed or plastic.

1.2.1. Genomic correlates of fixed alternative phenotypes

The Atlantic salmon (*Salmo salar*) males can develop into one of two alternative reproductive tactics that are fixed throughout their lifetimes, either mate as large males (i.e. mature male parr), whose substantial growth was obtained during their migration into seawater, or mate as small males that remained during their whole development in freshwater and opt for sneaking their conspecifics (Aubin-Horth et al., 2009; Fleming, 1998; Verspoor et al., 2007). Depending on the environment and internal conditions (i.e. body size), any male can develop into one of these two irreversible phenotypes characterized by specific behavioural states (Aubin-Horth and Dodson, 2004). In order to study the molecular basis of this plastic trait, Aubin-Horth and colleagues (Aubin-Horth et al., 2005a) compared males of the same age (sneaker and immature males that will eventually become large fighting males) in a genome-wide approach. The microarray analysis revealed that 15% of the genes examined vary in expression between the two male types. Many of these differentially expressed genes are involved in processes such as growth, reproduction and neural plasticity. Genes related to cognition (learning and memory) and reproduction were upregulated in sneaker males, while genes related to cellular growth were upregulated in immature males (Aubin-Horth et al., 2005a, 2005b). Interestingly, even within a life history, for instance, migrating males, differences were found between early and late migrants, indicating different genomic signatures at different life stages (Aubin-Horth and Renn, 2009). [adapted from (Cardoso et al., 2015), Appendix I]

1.2.2. Genomic correlates of sequential (developmental) plasticity

A well-characterized example of developmental plasticity is provided by the distinct life stages and different behavioural tasks displayed by honey bees (*Apis mellifera*). During their development, bees assume different roles in their colony: (1) soon after eclosion, bees assume brood care functions (nursing); (2) after a week, they assume new roles, such as storing and processing food (e.g. turning nectar into honey); and (3) around 3 weeks of age, most bees begin foraging for pollen and nectar (Ben-Shahar, 2005; Whitfield et al., 2006, 2003). These different behavioural states are characterized by different profiles of gene expression in the bee brain. More than 85% of ~5500 analysed genes showed differences in expression associated with the transition from nurse to forager that are largely independent of natural age-related changes (Whitfield et al.,

2006). Whitfield et al. (Whitfield et al., 2006, 2003) also showed that individual brain expression patterns are so dramatically different between life stages that they can be used to classify an individual honey bee as a nurse or as a forager with a very high accuracy rate. [adapted from (Cardoso et al., 2015), Appendix I]

Recently, the focus has turned to alternative reproductive tactics (ARTs), which have evolved in many fish species, where fixed and plastic tactics coexist within the same population. The first study employing large-scale RNA-Seq analysis to study ARTs in fish was performed in the black-faced blenny (*Tripterygion delaisi*) (Schunter et al., 2014). Although the ecology of this species is not so well known when compared with others, at least two males coexist in populations, a colourful territorial male, which defends a nest against predators and other males, and a sneaker male that parasitically fertilizes eggs in nests (De Jonge and Videler, 1989; Wirtz, 1978). Interestingly in this species, it was shown that in 20% of the cases when a territorial male is removed a sneaker male takes over and changes its colouration as well as its behaviour and becomes territorial (De Jonge and Videler, 1989). Schunter and colleagues (2014) showed that more genes were differentially expressed between the two male phenotypes than between males and females, suggesting that during the reproductive period phenotypic plasticity is a more important factor in differential gene expression than sexual dimorphism. Another example of plastic ARTs comes from the bluegill sunfish (*Lepomis macrochirus*) where males present two distinct life histories, parental and cuckolder, encompassing three reproductive tactics, parental, satellite, and sneaker (Gross, 1982). The parental life history is fixed, whereas individuals who enter the cuckolder life history transition from sneaker to satellite tactic as they grow (Gross, 1982; Gross and Charnov, 1980). Clustering of the differentially expressed transcripts showed that sneaker males grouped separately from the other male tactics, while satellite males tended to have expression profiles intermediate between sneakers and the other groups. Interestingly, females at the brain level were more similar to parental males (Partridge et al., 2016). When comparing life histories, the contrast between parental males and sneaker males was the one that returned a larger number of differentially expressed transcripts (Partridge et al., 2016). Differential expression within life histories (i.e. between satellite and sneaker males) showed an overlap of almost 96% of the genes present in the comparison between parental and sneaker males (Partridge et al., 2016), indicating common genetic pathways underlying parental and satellite males' traits. Lastly, two comprehensive studies studied ARTs looking into gene expression patterns in the brain (Nugent et al., 2016) and gonads

(Dean et al., 2017) in the ocellated wrasse (*Symphodus ocellatus*). In this species three distinct male phenotypes occur, large and colourful nesting males, satellite males that cooperate with nesting males by courting females and chasing off sneaker males and engage in sneaking behaviour due to their proximate to nests, and finally small sneaker males that only engage in parasitic spawning at the nest (Stiver and Alonzo, 2013; Taborsky et al., 1987; Warner and Lejeune, 1985). At the brain level, this study found a different pattern among male morphs, with nesting males showing the most differentiated phenotype, and sneaker males representing an intermediate gene expression between nesting and satellite males (Nugent et al., 2016), although a different pattern was detected in a microarray study in the same species (please see (Stiver et al., 2015)). Whereas at the gonad level, looking into sex-biased gene expression patterns and magnitude, satellite males, which among the male morphs experience more intense sperm competition, had a more masculinized expression profile when compared with nesting males, and nesting males in turn showed a feminization of their gonadal transcriptome, while sneaker males showed both a demasculinization and de-feminization of gene expression (Dean et al., 2017). Within a species, males and females share the majority of the genome, which indicates that sexual dimorphisms in morphological and behavioural traits are mainly the product of differences in regulation of loci present in both sexes (Dean and Mank, 2016; Ellegren and Parsch, 2007). These sex-biased genes are often viewed as having a role in resolving sexual conflict and that sex-specific regulatory mechanisms may be key in sexual selection (Mank et al., 2013; Pointer et al., 2013).

The mechanisms underlying the expression of ARTs can differ significantly across species, nonetheless, from these studies, general patterns of expression emerge with plastic phenotypes presenting broader and more pronounced changes at the brain level and at the gonad level, pointing to differentiated selective pressures in male morphs.

1.2.3. Genomic correlates of behavioural flexibility

In the African cichlid fish Astatotilapia burtoni, males have evolved two distinct phenotypes: dominant males, which are brightly coloured, defend territories and actively court and spawn with females, and subordinate males, which have dull colouration similar to females, do not hold territories and are reproductively suppressed (Fernald and Hirata, 1977). These behavioural and phenotypic differences are reversible, and males change social status many times during their life depending on social context. Renn et al. (2008)

examined whole-brain gene expression in dominant and subordinate males in stable hierarchies as well as in brooding females and identified 171 genes that were differentially expressed between the two male types. [adapted from (Cardoso et al., 2015), Appendix I]

Another example of the presence of ARTs is the sailfin mollies (*Poecillia latipinna*), where the same phenotypic transition is accomplished either by plasticity or by a Y-linked genetic polymorphism that determines body size (Travis, 1994), and both these strategies coexist in the same population. Large males typically exhibit courtship behaviour, while small males adopt a sneaker strategy and attempt to copulate without apparent female cooperation (Seda et al., 2012). Interestingly, intermediate-size males are plastic in behaviour, adopting either a courting strategy or a sneaking strategy whether in the absence or presence of other males, respectively (i.e. environmentally regulated). Exploring this social regulation of behaviour, Fraser et al. (2014) detected only a partial overlap in brain gene expression when comparing genetic and plastic variation in behaviour, with the latter showing broader and more robust changes in gene expression. Together these results support the role of plasticity in facilitating adaptive evolution, however a not without a physiological cost to plasticity.

1.3. RNA-Seq experimental design and methodology in ecological genomics

The transcriptomics field has seen an intense development of transcriptomic applications as well as the technology generating sequence data, supplanting microarrays as the technology of choice for gene expression analysis. When next-generation sequencing (NGS) technologies become widely available for research purposes, three mainstream technologies were favoured, and the choice for one of them depended on their characteristics (for a review see (Hudson, 2008; Mardis, 2008)). Roche 454 technology, whose sequencing method relies on pyrosequencing, had as the major advantage the relatively long reads (~300 bp) that enabled assembly of contigs even in the absence of a reference genome. However relatively fewer reads were obtained which resulted in shallower coverage of sequencing. Illumina/Solexa technology method relies on sequencing by synthesis and has as a major advantage very deep coverage because of the large number of reads generated which in turn gives accurate measurements of gene expression levels, and yet the read length (~ 32 − 40 bp) was a deterrent in the analysis when a genome was not available. Lastly, ABI SOLiD sequencing by ligation technology had a large output of reads with very deep coverage. Likewise, the read length (~ 25 − 35

bp) was also a limitation in the absence of a reference genome, and the colour space encoding of reads is problematic in downstream analysis and bioinformatic applications. In the 'early years' of RNA-Seq analysis the sequencing strategy when working with non-model species combined both Roche 454 long reads and Illumina's short reads to obtain a *de novo* assembly and proceed with differential expression analyses. However, with the continuous development on sequencing chemistry that resulted in Illumina's longer reads (e.g. 250 bp in the HiSeq 2500 sequencer) paired-end read[1] sequencing, taken together with the improvement of the assembly algorithms obtaining a *de novo* transcriptome nowadays is less laborious.

Over the years, several protocols have been published to assure common guidelines and best practices when setting-up an RNA-Seq experiment and determining which strategies to adopt based on the system under study and research questions (e.g. (Conesa et al., 2016; De Wit et al., 2012; Wolf, 2013)). Broadly, four different stages can be identified and considered to be crucial to obtain biological relevant results in ecological genomics (Fig. 2): 1) experimental and sequencing design; 2) quality control of sequenced data and *de novo* transcriptome assembly; 3) transcriptome profiling and quantification; and 4) expression analysis and interpretation.

1.3.1. Experimental and sequencing design

The experimental design of RNA-Seq remains an area of development and may have significant impacts on analysis strategy (van Dijk et al., 2014). The first step is to guarantee that RNA with high quality and without genomic contamination is isolated for sequencing. The integrity of an RNA sample (RNA integrity number – RIN; (Schroeder et al., 2006)) is evaluated using the ratio 28S:18S ribosomal RNA in a scale ranging from 0 to 10 (i.e. fully degraded RNA – intact RNA), and samples with RIN above 7 in eukaryotes are usually considered for sequencing with minimal effects on results (Gallego Romero et al., 2014). A critical aspect of the experimental design is whether to select only mRNA (i.e. poly(A) enrichment) from total RNA or perform a selective ribosomal RNA depletion with the purpose of obtaining more evenly representation of genes in the pool of RNA for sequencing (Griffith et al., 2015). Another consideration is whether to retain

[1] For single-end reads, the sequencer 'reads' a fragment from only one end to the other, generating only one sequence with a pre-determined length size. In paired-end reads, the sequencer starts one read, finishes this direction at the specified read length, and then starts another round of sequencing from the opposite end of the fragment, obtaining the second read in the pair.

the information pertaining to strand origin by modifying the standard RNA-Seq protocol (i.e. stranded RNA-Seq) or choose unstranded library construction methods (Zhao et al., 2015). Obtaining strand-specific RNA-Seq data enables the assembler algorithm to distinguish between sense and antisense transcripts and minimize erroneous fusions between neighbouring transcriptional units that are encoded on opposite strands (Haas et al., 2013). Furthermore, sequencing can involve single-end (SE), or paired-end (PE) reads, with the latter being preferred for transcript reconstruction (Garber et al., 2011).

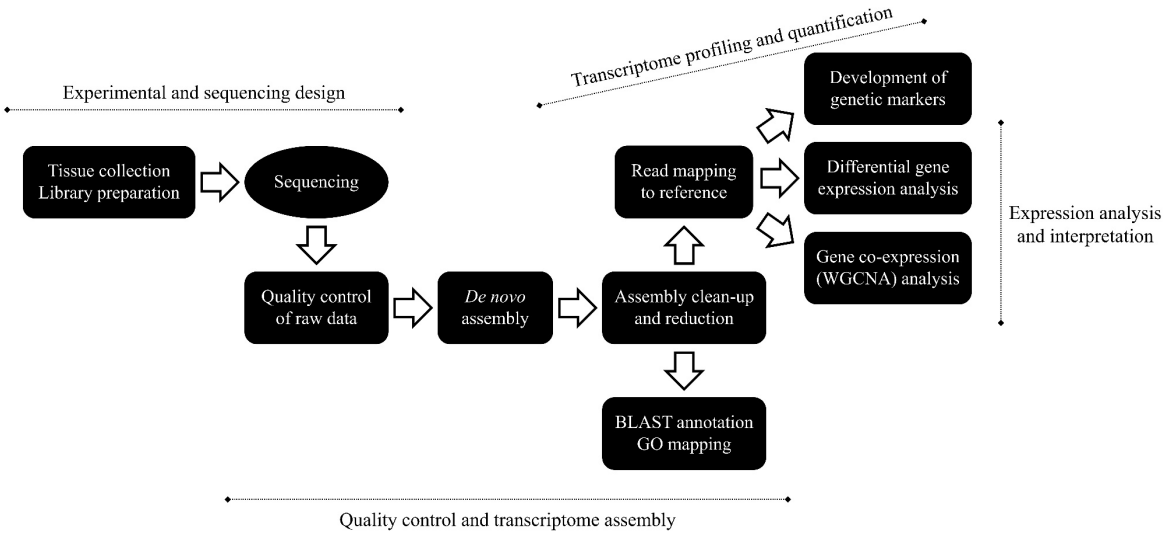


Fig. 2 – Basic RNA-Seq analysis workflow outline in ecological genomics.

Another critical factor is the sequencing depth or library size, which corresponds to the total number of sequenced reads per sample. As the sample is sequenced at a deeper level more transcripts will be detected and their quantification more precise (Mortazavi et al., 2008). However, a saturation in *novel* transcriptome discovery is usually achieved at a giving sequencing depth (Francis et al., 2013; Tarazona et al., 2011), and currently the recommendation is to sequence more biological replicates rather than sequencing samples at higher depth to obtain more robust results (Liu et al., 2014). The sequencing depth for studies whose main purpose is to analyse genes differentially expressed between conditions can be as low as 10 million reads when considering more biological replication to increase power and accuracy in large-scale differential expression RNA-Seq studies (Liu et al., 2014; Rapaport et al., 2013).

In summary, the adequate planning of an RNA-Seq experiment involves various decisions, many of them relying on the system under study (e.g. choice of long paired-end reads *versus* short single-end reads), some relying on the sequencing methodology to

avoid technical bias in the sequenced data (e.g. distribution of treatment and control samples in the sequencer) that will have an impact on the obtained sequenced data.

1.3.2. *De novo* assembly, reduction and functional annotation

Before proceeding with the assembly, quality control of the raw reads should be performed in order to detect sequencing errors, PCR artifacts or contaminations. Software tools such as Trimmomatic (Bolger et al., 2014) or CutAdapt (Martin, 2011), can be used to discard low-quality reads (i.e. defined by Phred quality scores[2]), trim adaptor sequences, and eliminate poor-quality baes. Libraries should be visualized using FastQC (Andrews, 2010) or other tools, before and after performing this step to assure that only high-quality reads are used in the study.

The process of reconstruction of full-length transcripts is challenging for diverse biological reasons, such as the coverage of transcriptome sequences varies in accordance to gene-expression levels, the presence of multiple splice-variants, allelic variants, paralogs and pseudogenes. Technically, transcriptome assembly algorithms have to accurately reconstruct transcripts using short-reads containing different quality scores, handle uneven coverage across the sequence length and ambiguities introduced due to conserved domains in closely related and duplicated genes (Martin and Wang, 2011). In the absence of genomic data for the system under study or existence of a closely related species with genomic information, the most common strategy is to use a reference-independent. The first reference-independent assemblers were used to assemble long reads and relied on overlap-layout-consensus strategy (e.g. Mira (Chevreux et al., 2004) and Newbler (Margulies et al., 2005) assemblers). Recent reference-independent assemblers assemble the short reads using a *de Bruijn* graph approach in which reads are broken down into sequences of length *k* (*k*-mers) that form nodes and are connected by edges based on *k-1* bp overlap to build the sequence of the contig (Martin and Wang, 2011). Several assemblers use this approach (for a review see (Zhao et al., 2011)), such as Trinity assembler that remains the preferred assembler for non-model species due to its excellent performance with relatively reasonable computer resources, positive assembly metrics and supported companion utilities for downstream analysis (Grabherr

[2] Phred quality score is a measure of the quality of the identification of the nucleobases generated by automated DNA sequencing (Ewing et al., 1998; Ewing and Green, 1998).

et al., 2011; Haas et al., 2013). Interestingly, Trinity groups transcripts into clusters, which can be loosely referred to as 'gene', based on shared sequence content.

Due to the challenges, either biological or technical, during the assembly of reads it is common for *de novo* transcriptomes being composed of hundreds of thousands of contigs (i.e. transcripts). Many of these transcripts represent noncoding (i.e. erroneously inferred isoforms), chimeric, or rare variants with low expression-level support that should be removed before downstream analysis. Foremost, sequence redundancy can be reduced in the original assembly by removing smaller sequences that have high homology with longer sequences. This step can be accomplished by using CD-HIT (Fu et al., 2012) with tailored parameters. Secondly, sequences with low expression support given by the sequenced samples should be removed. The measure FPKM (Fragments Per Kilobase of transcript per Million mapped reads; (Trapnell et al., 2010)), a derivate from the RPKM (Mortazavi et al., 2008) but applied for paired-end reads, is a within-sample normalization method that will remove the feature-length and library size effects. This measure is useful to report expression levels, and visualize samples' expression patterns, although, more recently has aided in the exclusion of low expression-level transcripts from the assembly (e.g. (Dean et al., 2017; Harrison et al., 2015, 2012)). Typically, an FPKM threshold for transcript expression is defined guarantying that the transcript is only retained in the circumstance of being expressed with high confidence in a fraction of the replicates of at least in one condition. The threshold for expression is chosen when similar normalization profiles across samples are achieved (see section 1.3.4). Finally, if the interest resides only in transcripts containing coding sequences (CDS), TransDecoder (Haas et al., 2013) can be used to identify candidate coding regions within transcript sequences.

Generally, functional annotation of a *de novo* transcriptome is inferred based on currently known proteins whose sequence data is available in public databases, such as the NCBI non-redundant protein database, UniProt Knowledgebase and Ensembl. BLASTx searches against each of these databases can be performed as a standalone or handled by Blast2GO (Conesa et al., 2005) software, which is dedicated to functional gene annotation, using different thresholds for the various annotation parameters. Additionally, by using the annotated accession numbers, Blast2GO retrieves Gene Ontology terms (Gene Ontology Consortium, 2004) as well as enzyme codes and KEGG pathway analysis (Kanehisa and Goto, 2000).

1.3.3. Read mapping and genetic marker development

Several sophisticated algorithms have been developed that quantify expression from transcriptome mappings (for a review see (Zhang et al., 2017)), namely RSEM (RNA-Seq by Expectation Maximization; (Li and Dewey, 2011)). This method handles the results from the aligning tool Bowtie2 (Langmead and Salzberg, 2012), by allocating multi-mapping reads among the mapped transcripts, and outputs the expression levels (i.e. raw count data) per transcript and per gene, supplemented with several other metrics. Raw counts by themselves are not sufficient to compare expression levels among samples, as these are affected by factors such as transcript length, the total number of reads, and sequencing biases, and hence more powerful normalization methods are applied before testing for differential gene expression (please see section 1.3.4).

RNA-Seq data, besides its primary purpose on gene expression quantification, contains valuable genetic information that can be applied in population genetics. Taking advantage of the alignment information obtained previously, two types of genetic markers can be evaluated *in silico* for their polymorphism. Microsatellite loci, which have been the marker of choice in classical population genetics (e.g. parentage and phylogenetic analysis; (Chistiakov et al., 2006)), are single sequence repeats often presenting highly polymorphic loci. Most microsatellite are present in non-coding regions, allowing mutation to occur on a more rapid pace and therefore neutral (type II markers; (O'Brien, 1991)). Equally, microsatellite can be found on the translated genome associated to genes of known functions (type I markers; (O'Brien, 1991)), making them more useful for comparative genetic mapping although of being less polymorphic due to functional constraints (Serapion et al., 2004). Microsatellites are easily identified in sequences from transcriptomic data, and yet some ambiguity to data analysis is introduced by the presence of null alleles and mutation patterns that are variable (Morin et al., 2004). SNP (single nucleotide polymorphism) markers, on the other hand, have the advantage of being both neutral or allow for the discovery of genes under selection (Morin et al., 2004). High-throughput SNP genotyping in non-model species has received more attention, and several methods are available that allow for the analysis of thousands of SNPs simultaneously (e.g. (Dean et al., 2017; Lopez-Maestre et al., 2016; Seeb et al., 2011). Hence, SNP markers have become the marker of choice for a variety of studies in population genetics, although microsatellites remain a reference marker to take into consideration.

1.3.4. Differential expression methods and gene co-expression network analysis

One of the most widely used applications of RNA-Seq is the estimation of gene abundance and comparison of these expression levels across biological conditions. Most of the methods available that handle count data matrices were developed for R statistical environment (R Core Team, 2018), and consequent downstream visualization and statistical analyses are performed in this environment. Differential gene expression analysis of RNA-Seq data generally consists of three components: normalization of the raw counts, statistical modelling of gene expression and testing for differential expression (Anders et al., 2013; Rapaport et al., 2013). The aim of normalization is to eliminate systematic technical effects that occur in the data and therefore are not associated with the biological differences of interest. The main source of variation are the significant differences in sequencing depth (i.e. total number of reads) that occur sometimes between samples. However, accurate estimation of the sequencing depth among samples is not trivial because RNA-Seq counts inherently represent relative abundances of the genes, and gene counts may be more or less evenly distributed in samples. To account for this, more complex normalization methods have been developed as part of standardized statistical packages developed with the sole purpose to handle RNA-Seq data (Costa-Silva et al., 2017; Schurch et al., 2016; Soneson and Delorenzi, 2013). The two most used packages are DESeq2 (Love et al., 2014) that was developed with substantial differences in its methodology relatively to its predecessor DESeq (Anders and Huber, 2010), and edgeR (Robinson et al., 2010) packages. Normalization methods, either in the TMM normalization (edgeR) or in the median-of-ratios normalization (DESeq, DESeq2), start from the assumption that most genes are not differentially expressed and that the set of differentially expressed genes is more or less equally divided between up-regulated and down-regulated genes. Common to all of the abovementioned packages is the statistical modelling of gene expression using a negative binomial (NB) distribution since it has been shown to be a good fit to RNA-Seq data (McCarthy et al., 2012) since it is flexible enough to account for biological variability. Lastly, once the statistical model has been defined, a test for differential expression to determine which genes show differences in their expression levels between experimental groups is applied taking into account technical and biological variation in these expression levels. In the case of DESeq an exact test is used, in DESeq2 the Wald test and in edgeR GLM the likelihood ratio test. The results are further screened to increase certainty when calling genes as differentially

expressed, by first performing a correction to the p-values for multiple testing by applying the Benjamini-Hochberg FDR procedure (Benjamini and Hochberg, 1995), as thousands of tests are performed (each gene corresponding to one test), and by defining a threshold of change in gene expression levels between experimental conditions (i.e. fold-change).

A different method used to explore the complex relationships between genes and phenotypes is the WGCNA algorithm (Weighted Gene Co-Expression Network Analysis; (Langfelder and Horvath, 2008)). Broadly, WGCNA analysis clusters together genes (modules) with highly correlated expression across all samples, which are summarized using the module eigengene or an intramodular hub gene used for relating modules to one another and to external sample traits (using eigengene network methodology), and for calculating module membership measures (Langfelder and Horvath, 2008). The analysis is accomplished in three phases, network construction, module detection, and relating modules to external traits, which involve choosing the appropriate parameters that better represent the biological data. Typically, for network construction, a signed weighted network is preferred to keep track of the sign of the co-expression information as well to reflect the continuous nature of the underlying co-expression information (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). Weighting a network involves choosing a soft-thresholding power β, which weights each connection between genes, in accordance to the criterion of obtaining a network satisfying scale-free topology with a signed $R^2 > 0.8$ (Zhang and Horvath, 2005). Once the network has been constructed, module detection is accomplished using a similarity measure, topological overlap measure (TOM), that clusters densely interconnected genes in different modules identified with colour names. Lastly, associations between modules and external traits are quantified, usually by correlating module eigengenes with traits of interest. Afterwards, *in silico* validation of the results can be performed by assessing the relationship between gene significance for each trait and module membership (i.e. correlation of the module eigengene with the gene expression profile).

Finally, enrichment analysis of the set of genes found to be differentially expressed or present in a particular module found to be correlated with a specific condition is essential to aid in the interpretation of the extensive gene lists. This can be done for the gene function by using the gene ontology (GO) framework (Gene Ontology Consortium, 2004), which uses GO terms to classify gene function along three aspects, molecular activity of the gene product (i.e. molecular function), where gene products are active (i.e. cellular component) and identify pathways and larger processes made up of the activities

of multiple gene products (i.e. biological process). These terms can be found enriched using several statistical tests, such as the hypergeometric test or Fisher's exact test.

1.4. The peacock blenny *Salaria pavo*, a model system in plasticity

Salaria pavo (Risso, 1810; Teleostei: Blenniidae) is a small intertidal fish, usually found in rocky shores of the Mediterranean and adjacent Atlantic areas (Zander, 1986). This species presents a strong sexual dimorphism, with nest-holder males being larger than females and having well-developed secondary sexual characters that consist of a conspicuous head crest and a pheromone and antibiotic-producing anal gland, which are used to attract females to their nests for spawning (Fig. 3) (Barata et al., 2008; Gonçalves et al., 2002a; Patzner et al., 1986). In this blenny, the mating system is usually promiscuous with males spawning with several females throughout the breeding season and females laying their eggs with more than one male. Nest-holder (bourgeois) males acquire nests that mainly consist of holes or crevices in the rock, where they provide sole paternal care to eggs until hatching (i.e. nest cleaning and egg fanning (Fishelson, 1963; Patzner et al., 1986)), and defend courting territories around the entrance of the nest from which they attract females during the breeding season (Zander, 1986).

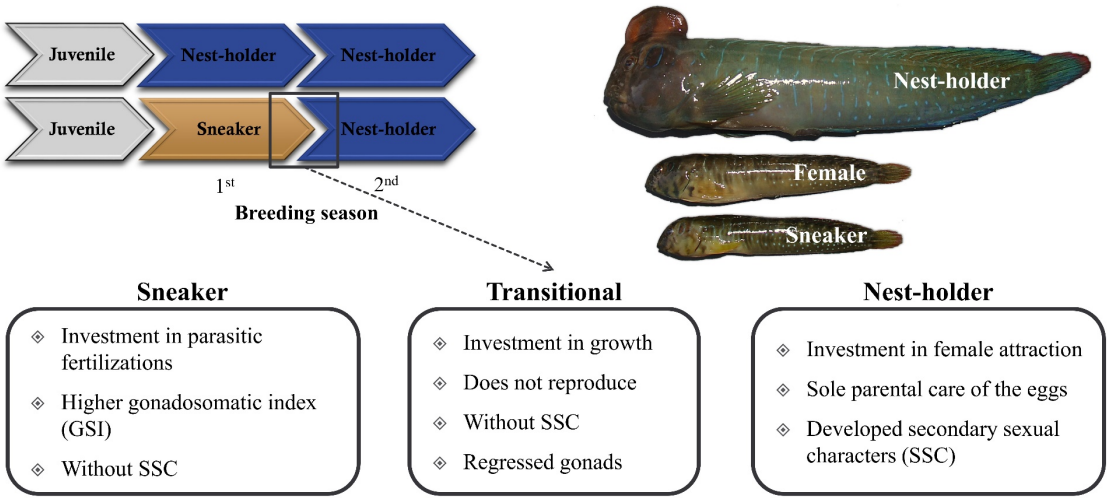


Fig. 3 – Schema representing *Salaria pavo* phenotypes and developmental pathways followed by males in this species consisting of alternative reproductive tactics (ARTs). Sexual dimorphism is evident between the nest-holder male (top) and the female (middle). Also, the intra-sexual polymorphism can be observed between the nest-holder male and the sneaker male (bottom).

In the peacock blenny, the intensity of mating competition varies among populations due to nest-site availability, such that two different levels of plasticity arise: 1)

intraspecific variation in reproductive behaviour for males that can follow either of two developmental pathways, grow directly into nest-holder males, or behave first as female mimics to sneak fertilizations (sneaker males) and afterward transition into nest-holder males, and 2) inter-population variation in courting roles of females and nest-holder males. In the Mediterranean rocky shore populations, as in Gulf of Trieste (e.g. Adriatic Sea, Italy), nests are available in abundance. Nest-holder males establish nests in rock crevices or holes, aggressively defend a territory around the nest and actively court females, while females assume a passive role in courtship behaviour, responding with changes in colouration and a few displays before they enter the nest to spawn (i.e. 'conventional' sex roles (Patzner et al., 1986; Saraiva et al., 2012)). In contrast, coastal populations such as the one at Culatra island (Ria Formosa, Portugal) the habitat is mainly composed of an extensive intertidal mudflat area, where the only hard substrates available are artificial reefs (i.e. agglomerated bricks and tiles) used by clam culturists to delimit their fields (Almada et al., 1994). Because of the scarcity of nest sites, strong intra-sexual competition between males is present and only large competitive males can acquire and defend a nest. After the breeding season starts, males rarely leave their nests and do not defend any area around the nest (the breeding territories are restricted to the nest itself), and it is common to observe males nesting in adjacent holes of the same brick (Almada et al., 1994). At the peak of the breeding season, most nests are filled with eggs and nest space may become a limiting factor for female reproduction (Almada et al., 1994). The environmental constraints promote two peculiarities in the mating system of this population. Primarily, a sex role reversal in courtship behaviour, where females have the most active role in courtship (Almada et al., 1995). Sex role reversal is considered to evolve under female-biased operational sex ratio[3] (OSR) in populations, which is the case of the peacock blenny coastal lagoon populations. Secondly, the presence with a high frequency of sneaker males in the Culatra population (Almada et al., 1995), adopting a parasitic tactic that is considered an example of a plastic sequential alternative reproductive tactic (*sensu* (Oliveira, 2006)). Although both sexes mate with multiple mates, males are selective with respect to the females they accept as mates. Females, in

[3] The OSR of an organism describes the relative abundance of males and females ready to mate at any point of time (Emlen and Oring, 1977; Kvarnemo and Ahnesjö, 1996), and predicts that the sex facing a shortage of potential mates (i.e. toward which the OSR is biased) should show stronger mating competition, which can be manifested in an increased intrasexual agonistic encounters in the effort to attract the other sex (e.g. courtship behaviour).

order to compete for the access to nesting males, approach them with a typical nuptial coloration, consisting of a pattern of light brown and dark vertical bars in the anterior portion of the body and head, that can be turned on/off within seconds, along with elaborated courtship behaviours, involving flickering the pectoral fins and opening-and-closing the mouth in synchrony (Almada et al., 1995). Small males are unable to acquire nests and adopt alternative reproductive tactics, acting as sneakers (parasitic tactic). They approach nest-holder males mimicking the female's morphology and courtship displays in order to come close to the nest-holder male's nest and parasitically fertilize part of the eggs (Fig. 3) (Gonçalves et al., 2005, 1996). The female-mimicry seems to be efficient as nesting males court and attack small sneakers and females with equal frequency (Gonçalves et al., 2005). These alternative reproductive tactics are sequential since sneaker males at a later age switch to nest-holder males (Fagundes et al., 2015), hence adopting a bourgeois tactic. During the transition from the sneaker to the nest-holder male phenotype, the male (aka transitional male) does not reproduce and mainly invests in growth. These alternative reproductive tactics pose an interesting case study since the same male expresses both male and female reproductive behaviour during his lifetime.

1.5. Aims and structure of the thesis

In this thesis, I report the results of a series of studies done in the peacock blenny *Salaria pavo*, where RNA sequencing (RNA-Seq) was used to develop genetic markers and uncover patterns of gene expression underlying the observed plasticity in natural populations. Four main questions were posed:

1) How successful is each male alternative reproductive tactic (ART) in fertilizing eggs?
2) How is sexual dimorphism between females and nest-holder males paralleled by gene expression differences?
3) Which is the role of temporal differential gene expression in the evolution of sequential male ARTs?
4) Which gene expression profiles are associated with shifts in sex roles in courtship behaviour?

In order to understand how male ARTs evolve and are maintained in populations, it is necessary to measure their reproductive success (i.e. consisting of mating success and

fertilization success), which can be used as a proxy to Darwinian fitness. In chapter 2, I used pyrosequencing data generated from pooled individuals and tissues, to mine for microsatellites (simple sequence repeats (SSRs) which are spread across the genome) and characterize them in regards to their polymorphism *in silico*. I provide evidence that combining the knowledge of the number of repeats with other predictors of variability (i.e. *in silico* microsatellite polymorphism) improves the rates of polymorphism when the microsatellites were applied in natural populations. In chapter 3, selected microsatellites from the previous work were used to assess paternity at nest-holder male nests, and hence determine the fertilization success of each male tactic. Combining the genetic information with social dynamics observed during the breeding season (i.e. social network analysis), I integrate the results of fertilization success, with previous information on mating success available for this species (Gonçalves et al., 2002b; Oliveira et al., 1999), and make final considerations for reproductive success of each male ART.

In chapter 4, I used the pyrosequencing data together with Illumina HiSeq reads to assemble the first transcriptome in this species and characterize the neurogenomic states associated with male ARTs. In this study, I integrate differential gene expression with co-expression analysis to characterize general patterns of expression underlying each phenotype present in Culatra population. Additionally, I look into sex-biased genes to further characterize sneaker males' magnitude of expression when compared with females and nest-holder males. The results are discussed in the context of ARTs and integrated with published work in other species.

In chapter 5, I focus on the second level of plasticity observed in the peacock blenny, plastic sex roles in courtship behaviour. I sampled females and nest-holder from the two well-characterized populations exhibiting contrasting behaviours and sequenced two target tissues, forebrain and gonads. General expression profiles were obtained for each tissue and inferences on the observed patterns made to further the research with this data.

Finally, in Chapter 6, I summarise the most important results of the thesis and discuss the implications of my findings in relation to previous literature, and pose final considerations for future research.

1.6. References

Almada V.C., Gonçalves E.J., Oliveira R.F. & Santos A.J., 1995. Courting females: ecological constraints affect sex roles in a natural population of the blenniid fish Salaria pavo. *Animal Behaviour*, *49*, 1125–1127.

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Anders S. & Huber W., 2010. Differential expression analysis for sequence count data. *Genome Biology*, *11*, R106.

Anders S., McCarthy D.J., Chen Y., Okoniewski M., Smyth G.K., Huber W. & Robinson M.D., 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, *8*, 1765–1786.

Andrews S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aubin-Horth N. & Dodson J.J., 2004. Influence of individual body size and variable thresholds on the incidence of a sneaker male reproductive tactic in Atlantic salmon. *Evolution*, *58*, 136–144.

Aubin-Horth N., Landry C.R., Letcher B.H. & Hofmann H.A., 2005a. Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 1655–1662.

Aubin-Horth N., Letcher B.H. & Hofmann H.A., 2009. Gene-expression signatures of Atlantic salmon's plastic life cycle. *General and Comparative Endocrinology*, *163*, 278–284.

Aubin-Horth N., Letcher B.H. & Hofmann H.A., 2005b. Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon. *Journal of Heredity*, *96*, 261–278.

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

Barata E.N., Serrano R.M., Miranda A., Nogueira R., Hubbard P.C. & Canário A.V.M., 2008. Putative pheromones from the anal glands of male blennies attract females and enhance male reproductive success. *Animal Behaviour*, *75*, 379–389.

Ben-Shahar Y., 2005. The foraging gene, behavioral plasticity, and honeybee division of labor. *Journal of Comparative Physiology A*, *191*, 987–994.

Benjamini Y. & Hochberg Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.

Bolger A.M., Lohse M. & Usadel B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Cardoso S.D., Teles M.C. & Oliveira R.F., 2015. Neurogenomic mechanisms of social plasticity. *Journal of Experimental Biology*, *218*, 140–149.

Chevreux B., Pfisterer T., Drescher B., Driesel A.J., Müller W.E.G., Wetter T. & Suhai S., 2004. Using the miraEST assembler for reliable and automated mRNA transcript

assembly and SNP detection in sequenced ESTs. *Genome Research*, *14*, 1147–1159.

Chistiakov D.A., Hellemans B. & Volckaert F.A.M., 2006. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, *255*, 1–29.

Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M. & Robles M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676.

Conesa A., Madrigal P., Tarazona S., Gomez-Cabrero D., Cervera A., McPherson A., Szcześniak M.W., Gaffney D.J., Elo L.L., Zhang X., Mortazavi A. & Mortazavi A., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13.

Costa-Silva J., Domingues D. & Lopes F.M., 2017. RNA-Seq differential expression analysis: An extended review and a software tool. *PloS ONE*, *12*, e0190152.

Dall S.R.X., Giraldeau L.-A., Olsson O., McNamara J.M. & Stephens D.W., 2005. Information and its use by animals in evolutionary ecology. *Trends in Ecology & Evolution*, *20*, 187–193.

Darwin C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. Murray, John, London.

De Jonge J. & Videler J.J., 1989. Differences between the reproductive biologies of Tripterygion tripteronotus and T. delaisi (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour. *Marine Biology*, *100*, 431–437.

De Wit P., Pespeni M.H., Ladner J.T., Barshis D.J., Seneca F., Jaris H., Therkildsen N.O., Morikawa M. & Palumbi S.R., 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, *12*, 1058–1067.

Dean R. & Mank J.E., 2016. Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression. *The American Naturalist*, *188*, E74–E84.

Dean R., Wright A.E., Marsh-Rollo S.E., Nugent B.M., Alonzo S.H. & Mank J.E., 2017. Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. *Molecular Ecology*, *26*, 505–518.

Ekblom R. & Galindo J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*, 1–15.

Ellegren H. & Parsch J., 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*, 689–698.

Emlen S.T. & Oring L.W., 1977. Ecology, sexual selection, and the evolution of mating systems. *Science*, *197*, 215–223.

Ewing B. & Green P., 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, *8*, 186–194.

Ewing B., Hillier L., Wendl M.C. & Green P., 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, *8*, 175–185.

Fagundes T., Simões M.G., Saraiva J.L., Ros A.F.H., Gonçalves D. & Oliveira R.F., 2015. Birth date predicts alternative life history pathways in a fish with sequential reproductive tactics. *Functional Ecology*, *29*, 1533–1542.

Fernald R.D. & Hirata N.R., 1977. Field study of Haplochromis burtoni: habitats and co-habitant. *Environmental Biology of Fishes*, *2*, 299–308.

Fishelson L., 1963. Observations on littoral fishes of Israel. I. Behaviour of Blennius pavo Risso (Teleostei: Blenniidae). *Israel Journal of Zoology*, *12*, 67–80.

Fleming I.A., 1998. Pattern and variability in the breeding system of Atlantic salmon (Salmo salar), with comparisons to other salmonids. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*, 59–76.

Francis W.R., Christianson L.M., Kiko, R., Powers M.L., Shaner N.C. & Haddock S.H., 2013. A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics*, *14*, 167.

Fraser B.A., Janowitz I., Thairu M., Travis J. & Hughes K.A., 2014. Phenotypic and genomic plasticity of alternative male reproductive tactics in sailfin mollies. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132310.

Fu L., Niu B., Zhu Z., Wu S. & Li W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*, 3150–3152.

Fudickar A.M., Greives T.J., Atwell J.W., Stricker C.A. & Ketterson E.D., 2016. Reproductive Allochrony in Seasonally Sympatric Populations Maintained by Differential Response to Photoperiod: Implications for Population Divergence and Response to Climate Change. *The American Naturalist*, *187*, 436–446.

Gallego Romero I., Pai A.A., Tung J. & Gilad Y., 2014. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology*, *12*, 42.

Garber M., Grabherr M.G., Guttman M. & Trapnell C., 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, *8*, 469–477.

Gene Ontology Consortium, 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*, 258–261.

Gonçalves D., Matos R., Fagundes T. & Oliveira R., 2005. Bourgeois males of the peacock blenny, Salaria pavo, discriminate female mimics from females? *Ethology*, *111*, 559–572.

Gonçalves D.M., Barata E.N., Oliveira R.F. & Canário A.V.M., 2002a. The role of male visual and chemical cues on the activation of female courtship behaviour in the sex-role reversed peacock blenny. *Journal of Fish Biology*, *61*, 96–105.

Gonçalves D.M., Simoes P.C., Chumbinho A.C., Correia M.J., Fagundes T. & Oliveira R.F., 2002b. Fluctuating asymmetries and reproductive success in the peacock

blenny. *Journal of Fish Biology*, *60*, 810–820.

Gonçalves E.J., Almada V.C., Oliveira R.F. & Santos A.J., 1996. Female mimicry as a mating tactic in males of the blenniid fish Salaria pavo. *Journal of the Marine Biological Association of the United Kingdom*, *76*, 529–538.

Goodson J.L., 2005. The vertebrate social behavior network: evolutionary themes and variations. *Hormones and Behavior*, *48*, 11–22.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652.

Griffith M., Walker J.R., Spies N.C., Ainscough B.J. & Griffith O.L., 2015. Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. *PLoS Computational Biology*, *11*, e1004393.

Gross B.M.R., 1982. Sneakers, Satellites and Parentals: Polymorphic Mating Strategies in North American Sunfishes. *Zeitschrift für Tierpsychologie*, *60*, 1–26.

Gross M.R., 1996. Alternative reproductive strategies and tactics diversity within sexes. *Trends in Ecology & Evolution*, *11*, 92–98.

Gross M.R. & Charnov E.L., 1980. Alternative male life histories in bluegill sunfish. *Proceedings of the National Academy of Sciences of the United States of America*, *77*, 6937–6940.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N. & Regev A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512.

Harrison P.W., Mank J.E. & Wedell N., 2012. Incomplete sex chromosome dosage compensation in the Indian meal moth, Plodia interpunctella, based on de novo transcriptome assembly. *Genome Biology and Evolution*, *4*, 1118–1126.

Harrison P.W., Wright A.E., Zimmer F., Dean R., Montgomery S.H., Pointer M.A. & Mank J.E., 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4393–4398.

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

Kanehisa M. & Goto S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*, 27–30.

Kvarnemo C. & Ahnesjö I., 1996. The dynamics of operational sex ratios and competition for mates. *Trends in Ecology & Evolution*, *11*, 404–408.

Landry C.R. & Aubin-Horth N., 2014. Ecological Genomics - Ecology and the Evolution of Genes and Genomes. Springer Netherlands.

Langfelder P. & Horvath S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

Langmead B. & Salzberg S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Li B. & Dewey C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Liu Y., Zhou J. & White K.P., 2014. RNA-seq differential expression studies: more sequence, or more replication? *Bioinformatics*, *30*, 301–304.

Lopez-Maestre H., Brinza L., Marchet C., Kielbassa J., Bastien S., Boutigny M., Monnin D., Filali A. El, Carareto C.M., Vieira C., Picard F., Kremer N., Vavre F., Sagot M.-F. & Lacroix V., 2016. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Research*, *44*, e148.

Love M.I., Huber W. & Anders S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550.

Machado H.E., Pollen A.A., Hofmann H.A. & Renn S.C.P., 2009. Interspecific profiling of gene expression informed by comparative genomic hybridization: A review and a novel approach in African cichlid fishes. *Integrative and Comparative Biology*, *49*, 644–59.

Mank J.E., Wedell N. & Hosken D.J., 2013. Polyandry and sex-specific gene expression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*, 20120047.

Mardis E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*, 133–141.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.-J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X. V., Godwin B.C., He W., Rothberg J.M. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*, 376–380.

Martin J.A. & Wang Z., 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*, 671–682.

Martin M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10.

McCarthy D.J., Chen Y. & Smyth G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*, 4288–4297.

Metzker M.L., 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*, *11*, 31–46.

Morin P.A., Luikart G. & Wayne R.K., 2004. SNPs in ecology, evolution and

conservation. *Trends in Ecology & Evolution*, *19*, 208–216.

Mortazavi A., Williams B.A., McCue K., Schaeffer L. & Wold B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*, 621–628.

Newman S.W., 1999. The medial extended amygdala in male reproductive behavior. A node in the mammalian social behavior network. *Annals of the New York Academy of Sciences*, *877*, 242–257.

Nugent B.M., Stiver K.A., Alonzo S.H. & Hofmann H.A., 2016. Neuroendocrine profiles associated with discrete behavioural variation in Symphodus ocellatus, a species with male alternative reproductive tactics. *Molecular Ecology*, *25*, 5212–5227.

O'Brien S.J., 1991. Molecular genome mapping lessons and prospects. *Current Opinion in Genetic Development*, *1*, 105–111.

O'Connell L.A. & Hofmann H.A., 2012. Evolution of a vertebrate social decision-making network. *Science*, *336*, 1154–1157.

O'Connell L.A. & Hofmann H.A., 2011. The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. *The Journal of Comparative Neurology*, *519*, 3599–3639.

Oliveira R.F., 2009. Social behavior in context: Hormonal modulation of behavioral plasticity and social competence. *Integrative and Comparative Biology*, *49*, 423–440.

Oliveira R.F., 2006. Neuroendocrine mechanisms of alternative reproductive tactics in fish, in: Sloman, K.A., Wilson, R.W., Balshine, S. (Eds.), Fish Physiology: Behaviour and Physiology of Fish. Elsevier, New York, pp. 297–357.

Oliveira R.F., Almada V.C., Forsgren E. & Gonçalves E.J., 1999. Temporal variation in male traits, nesting aggregations and mating success in the peacock blenny. *Journal of Fish Biology*, *54*, 499–512.

Oliveira R.F., Taborsky M. & Brockmann H.J., 2008. Alternative Reproductive Tactics: an Integrative Approach. Cambridge University Press, Cambridge.

Parsons K.J. & Albertson R.C., 2013. Unifying and generalizing the two strands of evo-devo. *Trends in Ecology & Evolution*, *28*, 584–91.

Partridge C.G., MacManes M.D., Knapp R. & Neff B.D., 2016. Brain Transcriptional Profiles of Male Alternative Reproductive Tactics and Females in Bluegill Sunfish. *PLoS ONE*, *11*, e0167509.

Patzner R.A., Seiwald M., Adlgasser M. & Kaurin G., 1986. The reproduction of Blennius pavo (Teleostei, Blenniidae) V. Reproductive behavior in natural environment. *Zoologischer Anzeiger*, *216*, 338–350.

Pigliucci M., 2001. Phenotypic Plasticity: Beyond Nature and Nurture. John Hopkins University Press, Baltimore, MD.

Pointer M.A., Harrison P.W., Wright A.E. & Mank J.E., 2013. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genetics*, *9*, e1003697.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rapaport F., Khanin R., Liang Y., Pirun M., Krek A., Zumbo P., Mason C.E., Socci N.D. & Betel D., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, *14*, R95.

Renn S.C., Aubin-Horth N. & Hofmann H.A., 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics*, *5*, 42.

Renn S.C.P., Aubin-Horth N. & Hofmann H.A., 2008. Fish and chips: functional genomics of social plasticity in an African cichlid fish. *The Journal of Experimental Biology*, *211*, 3041–3056.

Reuter J.A., Spacek D. V & Snyder M.P., 2015. High-throughput sequencing technologies. *Molecular Cell*, *58*, 586–597.

Robinson M.D., McCarthy D.J. & Smyth G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.

Saraiva J.L., Pignolo G., Gonçalves D. & Oliveira R.F., 2012. Interpopulational variation of the mating system in the peacock blenny Salaria pavo. *Acta Ethologica*, *15*, 25–31.

Schroeder A., Mueller O., Stocker S., Salowsky R., Leiber M., Gassmann M., Lightfoot S., Menzel W., Granzow M. & Ragg T., 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, *7*, 3.

Schunter C., Vollmer S. V, Macpherson E. & Pascual M., 2014. Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics*, *15*, 167.

Schurch N.J., Schofield P., Gierliński M., Cole C., Sherstnev A., Singh V., Wrobel N., Gharbi K., Simpson G.G., Owen-Hughes T., Blaxter M. & Barton G.J., 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, *22*, 839–851.

Seda J.B., Childress M.J. & Ptacek M.B., 2012. Individual Variation in Male Size and Behavioral Repertoire in the Sailfin Molly Poecilia latipinna. *Ethology*, *118*, 411–421.

Seeb J.E., Carvalho G., Hauser L., Naish K., Roberts S. & Seeb L.W., 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, *11 Suppl 1*, 1–8.

Serapion J., Kucuktas H., Feng J. & Liu Z., 2004. Bioinformatic Mining of Type I Microsatellites from Expressed Sequence Tags of Channel Catfish (Ictalurus punctatus). *Marine Biotechnology*, *6*, 364–377.

Shendure J. & Ji H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, *26*,

1135–1145.

Soneson C. & Delorenzi M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14*, 91.

Stiver K.A. & Alonzo S.H., 2013. Does the Risk of Sperm Competition Help Explain Cooperation between Reproductive Competitors? A Study in the Ocellated Wrasse (Symphodus ocellatus). *The American Naturalist*, *181*, 357–368.

Stiver K.A., Harris R.M., Townsend J.P., Hofmann H.A. & Alonzo S.H., 2015. Neural Gene Expression Profiles and Androgen Levels Underlie Alternative Reproductive Tactics in the Ocellated Wrasse, Symphodus ocellatus. *Ethology*, *121*, 152–167.

Taborsky B. & Oliveira R.F., 2012. Social competence: an evolutionary approach. *Trends in Ecology & Evolution*, *27*, 679–688.

Taborsky M., Hudde B. & Wirtz P., 1987. Reproductive behaviour and ecology of symphodus (crenilabrus) ocellatus, a European wrasse with four types of male behaviour. *Behaviour*, *102*, 82–117.

Tarazona S., García-Alcalde F., Dopazo J., Ferrer A. & Conesa A., 2011. Differential expression in RNA-seq: a matter of depth. *Genome Research*, *21*, 2213–2223.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J. & Pachter L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*, 511–515.

Travis J., 1994. Size dependent behavioral variation and its genetic control within and among populations, in: Boake, B.R.C. (Ed.), Quantitative Genetic Studies of Behavioral Evolution. University of Chicago Press, Chicago, IL, pp. 165–187.

van Dijk E.L., Jaszczyszyn Y. & Thermes C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, *322*, 12–20.

Verspoor E., Stradmeyer L. & Nielsen J.L., 2007. The Atlantic Salmon: Genetics, Conservation and Management. Blackwell Publishing Ltd, Oxford.

Wallace A.R., 1858. On the tendency of varieties to depart indefinitely from the original type. *Journal of the Proceedings of the Linnean Society. Zoology.*, *3*, 53–62.

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

Warner R.R. & Lejeune P., 1985. Sex change limited by paternal care: a test using four Mediterranean labrid fishes, genus Symphodus. *Marine Biology*, *87*, 89–99.

West-Eberhard M.J., 2003. Developmental Plasticity and Evolution. Oxford University Press, New York.

Whitfield C.W., Ben-Shahar Y., Brillet C., Leoncini I., Crauser D., Leconte Y., Rodriguez-Zas S. & Robinson G.E., 2006. Genomic dissection of behavioral maturation in the honey bee. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 16068–16075.

Whitfield C.W., Cziko A.-M. & Robinson G.E., 2003. Gene expression profiles in the brain predict behavior in individual honey bees. *Science*, *302*, 296–299.

Wirtz P., 1978. The Behaviour of the Mediterranean Tripterygion Species (Pisces, Blennioidei). *Zeitschrift für Tierpsychologie*, *48*, 142–174.

Wolf J.B.W., 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources*, *13*, 559–572.

Zander C.D., 1986. Blenniidae, in: Whitehead, P.J.P., Bauchot, M.L., Hureau, J.C., Nielsen, J., Tortonese, E. (Eds.), Fishes of the North-Eastern Atlantic and the Mediterranean. UNESCO, Paris, pp. 1096–1112.

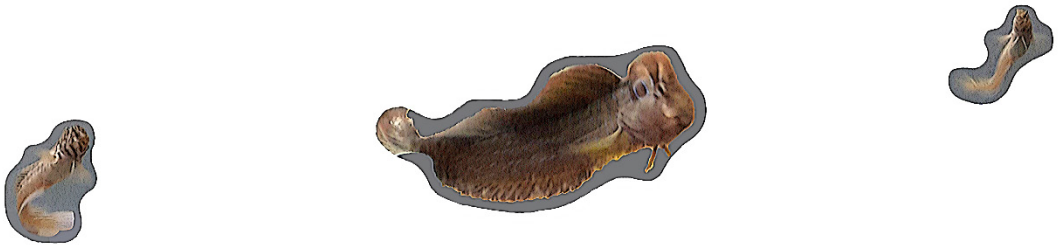
Zayed A. & Robinson G.E., 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annual Review of Genetics*, *46*, 591–615.

Zhang B. & Horvath S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article17.

Zhang C., Zhang B., Lin L.-L. & Zhao S., 2017. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics*, *18*, 583.

Zhao Q.-Y., Wang Y., Kong Y.-M., Luo D., Li X. & Hao P., 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*, *12*, S2.

Zhao S., Zhang Y., Gordon W., Quan J., Xi H., Du S., von Schack D. & Zhang B., 2015. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics*, *16*, 675.



CHAPTER 2 |

DEVELOPMENT OF GENETIC MARKERS

Cardoso S.D., Gonçalves D., Robalo J.I., Almada V.C., Canário A.V.M. & Oliveira R.F., 2013. Efficient isolation of polymorphic microsatellites from high-throughput sequence data based on number of repeats. *Marine Genomics*, *11*, 11–16. doi:10.1016/j.margen.2013.04.002

Efficient isolation of polymorphic microsatellites from high-throughput sequence data based on number of repeats

Sara D. Cardoso [1], David Gonçalves [1,2,3], Joana I. Robalo [1], Vitor C. Almada [1], Adelino V.M. Canário [2], Rui F. Oliveira [1,4]

[1] Unidade de Investigação em Eco-Etologia, Instituto Superior de Psicologia Aplicada – Instituto Universitário, Rua Jardim do Tabaco, 34, 1149-041 Lisboa, Portugal; [2] Centro de Ciências do Mar (CCMAR-CIMAR), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal; [3] University of Saint Joseph, Rua de Londres, 16, Macau SAR, China; [4] Champalimaud Neuroscience Programme, Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-901 Oeiras, Portugal

2.1. Abstract

Transcriptome data are a good resource to develop microsatellites due to their potential in targeting candidate genes. However, developing microsatellites can be a time-consuming enterprise due to the numerous primer pairs to be tested. Therefore, the use of methodologies that make it efficient to identify polymorphic microsatellites is desirable. Here we used a 62,038 contigs transcriptome assembly, obtained from pyrosequencing a peacock blenny (*Salaria pavo*) multi-tissue cDNA library, to mine for microsatellites and *in silico* evaluation of their polymorphism. A total of 4190 microsatellites were identified in 3670 unique unigenes, and from these microsatellites, *in silico* polymorphism was detected in 733. We selected microsatellites based either on their *in silico* polymorphism and annotation results or based only on their number of repeats. Using these two approaches, 28 microsatellites were successfully amplified in twenty-six individuals, and all but 2 were found to be polymorphic, being the first genetic markers for this species. Our results showed that the strategy of selection based on number of repeats is more efficient in obtaining polymorphic microsatellites than the strategy of *in silico* polymorphism (allelic richness was 8.2 ± 3.85 and 4.56 ± 2.45 respectively). This study demonstrates that combining the knowledge of number of repeats with other predictors of variability, for example *in silico* microsatellite polymorphism, improves the rates of polymorphism, yielding microsatellites with higher allelic richness, and decreases the number of monomorphic microsatellites obtained.

Keywords: Microsatellite development; *In silico* polymorphism Pyrosequencing; Number of repeats; *Salaria pavo*

2.2. Introduction

Microsatellites, or simple sequence repeats (SSRs), are among the widely used genetic markers in biology. Because of their high mutation rates, Mendelian inheritance and high reproducibility they can be used for genome mapping and to answer a wide range of biological questions, from the level of the individual (identity, sex, parentage) to the level of the species (phylogenetics, conservation) (Chistiakov et al., 2006).

Until recently, the advantages of microsatellite markers were partially offset by the difficulties inherent in marker development that is required for each species. The most commonly used approaches rely on laborious procedures from preparation and screening of genomic libraries to sequencing of isolated clones and primer design and validation or testing microsatellite primers already developed for closely related species (cross-species microsatellites) (Selkoe and Toonen, 2006; Zane et al., 2002a, 2002b). For species with genome sequences available, bioinformatic tools for *in silico* mining can be used to identify microsatellites and to design primers targeting these regions (Tóth et al., 2000). And while sequencing entire genomes of non-model organisms is still out of reach for most researchers, sequencing smaller subsets of the genome or of the transcriptome, presents an attractive alternative. This can now be achieved at affordable prices through next-generation sequencing platforms, that offer the possibility of sequencing long reads (up to 1000 bp), and make possible *de novo* transcriptome assembly without a reference genome (Abdelkrim et al., 2009; Csencsics et al., 2010; Hoffman and Nichols, 2011; Vera et al., 2008; Vogiatzi et al., 2011). Microsatellites developed from expressed sequence tags (ESTs) represent a potential source of type I markers, which are loci situated in transcribed regions associated to genes of known functions (O'Brien, 1991), making them more useful for comparative genetic mapping, linkage and quantitative trait loci association studies (Scaglione et al., 2009). These microsatellites are less polymorphic, due to functional constraints (Serapion et al., 2004), compared to those derived from non-coding genomic sequences, but their flanking regions are expected to be more conserved across closely related species (Slate et al., 2007; Vogiatzi et al., 2011), decreasing the appearance of null alleles.

Sequence assemblies have been extensively used for finding single nucleotide polymorphisms (SNPs) (Grattapaglia et al., 2011; Louro et al., 2010; Seeb et al., 2011), but much less to find polymorphic microsatellites *in silico*. The first steps in this direction were given by developing PolySSR (Tang et al., 2008), a database that stores information

about polymorphic SSRs using sequences from public EST databases (limited to seven organisms), and by Slate et al. (2007) in zebra finch and Shirasawa et al. (2012) in two cultivated peanut lines, which assembled sequences containing only microsatellites and inspected the alignments for contigs comprising sequences with different lengths of the same repeat motif. Recently, Hoffman and Nichols (2011) in Antarctic fur seal manually mined a transcriptome assembly for microsatellite polymorphism and obtained a positive relationship between the inferred number of alleles *in silico* and observed allele number. Furthermore, Neff and Gross (2001) by analyzing 592 AC microsatellite loci from 98 species obtained a positive relationship between microsatellite repeat length and the number of observed alleles across five vertebrate classes (fish, reptiles, amphibians, birds and mammals) and within each class.

We have therefore taken two different approaches for pre-screening microsatellites from next generation sequence data obtained from a normalized multi-tissue cDNA library in order to improve the level of polymorphism detected. In one approach microsatellites were mined for their polymorphism *in silico*, by screening the assembled contigs for variation in the number of repeats, and in the other approach microsatellites were selected based only on their number of repeats (repeat units comprising the microsatellite) which defines the alleles at each loci. Our species of choice was the peacock blenny (*Salaria pavo*) and its choice resulted from the lack of genetic markers for parentage assignment, an essential tool to understand the evolutionary advantage of the different reproductive tactics in this species (Gonçalves et al., 2005, 1996). The microsatellites selected using the two approaches were evaluated on individuals from three peacock blenny populations and the efficiency of the two approaches compared.

2.3. Materials and methods

2.3.1. Fish samples

Fish used for collecting the tissue samples for the normalized library were euthanized by rapid severance of the spinal cord with a scalpel. The fin samples for the genotyping procedures from individuals at Culatra Island (36°59′N, 7°51′W, Algarve, Portugal) were collected by light anesthetizing the fish with MS222 (Sigma) followed by recovery in a container with abundant aeration. These fish were released into the same place where they had been captured. At Formentera (38°41′N, 1°27′E, Spain) and Borovac (43°9′N, 16°24′E, Croatia), samples were collected from fish killed for other research purposes by

immersion in a lethal dosage of MS222. Animal protocols were performed in accordance with accepted veterinary practice under a "Group-1" license issued by the Directorate General for Veterinary of the Ministry for Agriculture, Rural Development and Fisheries of Portugal.

2.3.2. Sequence data and bioinformatic analysis

The peacock blenny transcriptome was sequenced on a GS-FLX System at Max Planck Institute (Berlin, Germany). Peacock blenny tissue samples were taken from 13 individuals (3 females, 3 bourgeois males, 3 sneakers and 4 transitional males (transition from sneaker to bourgeois male) sampled at Culatra Island. Total RNA was separately isolated with TRI Reagent® (Sigma-Aldrich) following standard procedures from 12 tissues (skin, muscle, bone, brain, olfactory epithelium, eyes, heart, kidneys, spleen, intestine, gonads and anal gland). Equal mass of total RNA from these tissues was pooled and used to construct one normalized multi-tissue cDNA library. Sample preparation and analytical processing such as base calling, were performed at Max Planck Institute using the manufacturer's protocol. After vector and quality trimming (≥q20), over 640,000 reads longer than 100 bp were assembled *de novo* using the MIRA3 assembler (Chevreux et al., 2004), in a total of 62,038 transcribed contigs with an average length of 452 bp. The mean coverage of these contigs was of 4.87 ± 17.3 reads (maxCoverage = 1054.5 and minCoverage = 1; mean ± standard deviation) and the mean nucleotide quality score was 35.71 ± 9.36. These contigs putatively correspond to different transcripts and henceforth were designated unigenes. The basic local alignment search tool (BLASTX) algorithm (Gish and States, 1993) was used to query for sequence similarities on all transcripts against the NCBI non-redundant (nr) protein sequence database (e-value b 1e−5, release of May 2010) using Blast2GO suite (Götz et al., 2008). Gene Ontology (GO) terms (Ashburner et al., 2000) were also obtained using Blast2GO with default parameters.

2.3.3. Microsatellite mining and selection

The identification and localization of perfect microsatellites in the assembled unigenes were accomplished using MSATCOMMANDER version 0.8.2 (Faircloth, 2008). The parameters were set for detection of di-, tri-, tetra-, penta-, and hexanucleotide motifs with a minimum of six repeats, and the option "Design Primers" was also chosen.

Tab-delimited files were generated from the searches using this software, and converted to spread-sheet files for subsequent data manipulation as described in Santana et al. (2009). Unigenes harbouring microsatellites were manually curated with the aid of Tablet (Milne et al., 2010), which allows graphical visualization of polymorphisms in unigene reads. Information collected with the software included number of reads covering completely the microsatellite (read coverage), completeness of the microsatellite at the 5′ end or 3′ end of the unigene and number of repeat unit variants found for the microsatellite (alleles). Microsatellites also received the annotation of its unigene.

In order to maximize the selection of polymorphic microsatellites for genotyping, two different strategies were pursued. The first strategy required the microsatellite to (i) display polymorphism in the reads forming the unigene; (ii) its unigene to have BLAST hits (e-value b 1e−5), and to (iii) have at least a pair of primers. In the second strategy, the microsatellites were only selected based on the number of repeats and the existence of a pair of primers, irrespective of BLAST hits and *in silico* polymorphism.

2.3.4. PCR amplification and fragment analysis

A set of 63 microsatellites developed from *S. pavo* unigenes were selected for amplification test using one peacock blenny DNA sample. PCR amplifications were set up in 50 µl volume composed of ~ 100 ng DNA, 0.25 pmol of each primer (MWG), 1.5 mM MgCl2, 120 µM of each dNTP, 5× Green GoTaq® Flexi Buffer 1×, and 1.5 u Taq DNA polymerase (Promega). PCRs were performed in a thermal cycler (Stratagene RoboCycler® Gradient 96) programmed as: 3 min at 94 °C for initial denaturation, followed by 35 cycles of 94 °C for 1 min, primer specific annealing temperature for 1 min, 72 °C for 45 s, and a final extension at 72 °C for 7 min. The success of the PCRs was determined by running 10 µL of each PCR product and co-running 6 µl of a mixture of DNA loading dye with a 50 bp DNA ladder (GeneRuler™ 50 bp DNA Ladder − 0.5 µg/µl; Fermentas) on a 1× Tris-acetate-EDTA buffer and 2% agarose gel stained with GelRed 3×, visualized under UV light and photographically documented.

For peacock blenny's loci that seemed to amplify well in the agarose gels, the respective forward primers were 5' fluorescently labelled with 6-carboxyfluorescein (6-FAM) or with hexachloro-fluorescein (HEX) dyes. A total of 26 adult peacock blenny individuals sampled from Culatra (20 samples), Formentera (3 samples) and Borovac Islands (3 samples) were employed for polymorphism assessment. Individuals from the

populations of Formentera and Borovac were used in order to verify if the microsatellite primers worked in all DNA samples and not only on those of Culatra where the primers were designed. DNA was extracted from the dorsal fin using Extract-N-Amp™ Tissue PCR Kit (Sigma-Aldrich). Microsatellite amplification reactions were performed in 25 µl volume containing ~ 100 ng DNA, 0.25 pmol of each primer (MWG), 1.5 mM MgCl2 (for exceptions see Table 1 and S2), 60 µM of each dNTP, 5× Green GoTaq® Flexi Buffer 1×, and 0.75 u Taq DNA polymerase (Promega). PCR thermal program was run as previously described (for annealing temperatures see Table 1).

DNA fragments were separated on a commercial ABI 3730XL DNA analyzer and sized by co-running a GeneScan HD400 (Applied Biosystems) size standard. DNA fragments were scored manually with the aid of GeneMarker® version 1.95 (SoftGenetics). For each working loci the type of microsatellite and the number of repeat variants were confirmed by commercial sequencing.

2.3.5. Microsatellite loci evaluation

Tests for Hardy–Weinberg equilibrium and genotypic linkage disequilibrium were performed using GENEPOP version 4.0.11 (Rousset, 2008) with the default setting (10,000 dememorization steps, 100 batches, and 5000 iterations per batch). Genetic diversity estimates, including expected (He) and observed (Ho) heterozygosities, were also calculated using GENEPOP. The test for the presence of null alleles was conducted using MICRO-CHECKER version 2.2.3 (Van Oosterhout et al., 2004).

In order to evaluate whether the two strategies used in this work potentially influence or not the polymorphism obtained, a Generalized Linear Model (GLM) was constructed within R (R Core Team, 2011). The number of different alleles observed among all individuals genotyped from the three populations for each microsatellite locus was modelled as a response variable using a Poisson error structure. The microsatellite number of repeats (minimum number of repeat units observed *in silico*) and the number of alleles observed *in silico* were used as predictor variables and fitted as continuous variables. Each variable was dropped from the model and the change in deviance between full and reduced model was distributed as χ^2 with degrees of freedom equal to the difference in degrees of freedom between the models with or without the variable in question. The residual deviance (difference between the deviance of the current model and the maximum deviance of the ideal model where the predicted values are identical to

the observed) was used to perform a goodness of fit for the overall model. Allelic richness (mean number of observed alleles per locus) between strategies was examined by using a two-tailed Student's *t* test or a Welch Two Sample *t* test, after testing for variance homogeneity.

2.4. Results and discussion

2.4.1. Microsatellite mining and *in silico* assessment of polymorphism

A complete search of the peacock blenny assembly of 62,038 unigenes for 5 types of microsatellites with a minimum number of repeats of 6 units identified 4190 microsatellite loci in 3670 unique unigenes, representing 5.9% of the sequenced transcriptome. Dinucleotide repeats accounted for 79.0% of all microsatellite loci, followed by 14.5% for trinucleotide, 4.4% for tetranucleotide, 1.4% for pentanucleotide and 0.7% for hexanucleotide repeats, values in the range observed in other fish species (Ju et al., 2005). It was not possible to determine the *in silico* polymorphism in 1428 microsatellites either because they were incomplete (28.2%) or because of single read coverage for the microsatellite region (71.8%). Polymorphic microsatellites were 733, of which 727 were dinucleotides and only 6 were trinucleotides. Two dinucleotide microsatellites had the maximum of 4 alleles each, while the majority of the microsatellites had two alleles (91.5%).

2.4.2. Microsatellite application and evaluation

Applying the first strategy criteria, 108 microsatellites were available comprising only dinucleotide repeats. From these, 33 microsatellites were selected based on the quality of the microsatellite flanking regions for primer design (Table S1, Appendix II). These microsatellites had a mean read coverage of 21.72 ± 60.59 reads, of which two unigenes accounted for 98 and 346 reads, and a mean number of repeats of 7.24 ± 1.94 units. When following the second strategy, 1340 microsatellites were available, of which 30 microsatellites were selected, comprising ten dinucleotides, six trinucleotides, eleven tetranucleotides, one pentanucleotide and two hexanucleotides. They had a mean read coverage of 3.43 ± 2.75 reads and a mean number of repeats of 12.03 ± 3.26 units. With the exception of locus *Spavo14*, none of these microsatellites appeared to be polymorphic *in silico* or originated BLAST hits (Table S1, Appendix II).

When the 63 primer pairs selected by the two strategies were PCR checked on one peacock blenny DNA sample and reaction conditions optimized, 38.1% (n = 24) led to different or multiple PCR products and 61.9% (n = 39) resulted in PCR products of the expected size. Amplification of DNA samples from three different peacock blenny populations of the Islands of Culatra, Formentera and Borovac, showed that three microsatellites had mononucleotide variation (variation of only one nucleotide between different alleles) and six microsatellites had multiple peaks or lacked a clear peak in the target region and were discarded. Fragment variation in two other microsatellites was not in accordance to the corresponding type of microsatellite, possibly because of insertion–deletion (indel) polymorphisms (Väli et al., 2008) combined with the polymorphism of the microsatellite. Twenty-eight microsatellites, 18 from using the first strategy and 10 from using the second strategy, were successfully characterized in all individuals used from the three locations (Table 1 and S2), and their sequences submitted to GenBank with the accession numbers: JQ619676–JQ619703. In 20 individuals genotyped from the population of Culatra, from which the cDNA library was originated, all but five dinucleotide microsatellite loci were found to be polymorphic (*Spavo15–Spavo19*; Table 1). The number of alleles ranged from 2 to 12 (4.83 ± 2.59) per locus and the observed and expected heterozygosities ranged from 0.05 to 0.85 and from 0.05 to 0.79, respectively. The mean number of alleles per locus and the expected heterozygosity were highest in microsatellite loci isolated using the second strategy (6.5 ± 2.88 and 0.62 ± 0.18) compared to the first strategy (3.54 ± 1.39 and 0.4 ± 0.22). Variation in allele number (Welch's unpaired *t* test, $t_{(12.233)} = 3.0$, $P = 0.01$) and expected heterozygosity (unpaired *t* test, $t_{(21)} = 2.52$, $P = 0.02$) were statistically significant. In this population, only *Spavo14* ($P < 0.001$) and *Spavo25* ($P < 0.05$) loci departed from Hardy–Weinberg (HW) equilibrium expectations, most probably because of heterozygote deficit (homozygote excess). The deviation of HW expectation in the first loci is significant possibly due to the presence of null alleles or stuttering leading to scoring errors. The presence of SNPs in *Spavo14* primer-binding sites cannot be excluded considering the low depth read coverage (2 reads) of the unigene. No other loci were detected with null alleles. Two of the possible pairwise comparisons between loci were in linkage disequilibrium ($P < 0.01$: *Spavo05–Spavo08* and *Spavo08–Spavo25*). For the Formentera and Borovac samples, all but eight and twelve microsatellite loci, respectively, were found to be polymorphic (Table S2, Appendix II) in the 3 individuals genotyped from

each population. The number of alleles ranged from 2 to 5 (mean 2.5 ± 0.83 and 2.69 ± 1.01 respectively) and expected heterozygosities from 0.33 to 0.93.

Table 1 – Locus primer sequences and microsatellite polymorphism characteristics. Microsatellites were identified *in silico* and developed for 28 loci from *Salaria pavo* unigenes, applied in twenty individuals from Culatra population. For each locus, the repeated motif and GenBank accession number are given.

Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	T_a (ºC)	CULATRA Size (bp)	*k*	Ho	He
Spavo01§ JQ619676	$(GT)_6$	F-CACCTCGAACAGTTGGCTTC GCTGCATTAGCCCAGATCC	58	387−397	3	0.30	0.27
Spavo02§ JQ619677	$(GA)_8C(GA)_4$	F-CCCTGGCTGATGTGACTCC ACTCTCCAGGTGTAAGGCAC	61	250−258	5	0.25	0.28
Spavo03§ JQ619678	$(AC)_6$-$(GT)_6$	F-GCACAAGTCGGCACTCAAG GCCAAGCCGAGTATGAAGC	60	229−237	4	0.50	0.58
Spavo04§ JQ619679	$(AC)_6$	F-CCCACGTCTGTTCAGTTGAC GGAGTTGGCACATTCCGTG	58	259−266	3	0.40	0.45
Spavo05§ JQ619680	$(AC)_9$	F-ATCAGCGCGAAACACATCG ACTGCACTCAAGTCAAAGCC	56	185−189	3	0.55	0.52
Spavo06§ JQ619681	$(TG)_8$	F-GCTGGTCGATGGCAGAATG GCGTCGGAAATACCGTTCC	58	295−297	2	0.05	0.05
Spavo07§ JQ619682	$(AC)_4G(AC)_{10}$	F-CACGACAGCTGGTCTCAAC GGGCTCACCAGTCCCATTC	58	331−337	3	0.35	0.42
Spavo08§ JQ619683	$(CA)_9$	F-CGTGACTTCATGGCAAGGG TGTGTGGAAACGATATGTGC	58	221−235	7	0.75	0.79
Spavo09 [c] JQ619684	$(AC)_8$	F-CGCTAAAAGGAGGCAACATC ACAGCGACGAGCTTCATCTT	61	196−200	3	0.10	0.10
Spavo10§ JQ619685	$(AC)_9$	F-AGAGTAGGGGTCCGTCGATT TGGCAGTGAGAAAGTGCAAG	61	137−141	3	0.10	0.19
Spavo11§ JQ619686	$(CT)_9$	F-GGTAGCGAGAGACGCAGAAG GGTAGACCAGCGGTCTGAAG	62	232−234	2	0.60	0.43
Spavo12§ JQ619687	$(AC)_7G(AC)_{12}$	F-GCTGTAAAACTGCGTGGACA GGACGTGAACCTGGAGAAGA	61	179−204	5	0.60	0.56
Spavo13§ JQ619688	$(AC)_{10}$	F-CCTCGCAGCAGTAACTCAGA TCCGTCTATGGAGGCTAACG	61 [b]	136−146	3	0.60	0.59
Spavo14 JQ619689	$(AC)_{17}$	F-GGGGATCGAAATGTTTCACA CCACATGGAACCAACTTCCT	59	246−260	5	0.40	0.75**

(Continued)

Table 1 *(Continuation)*

Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	T_a (ºC)	CULATRA Size (bp)	*k*	Ho	He
Spavo15§ JQ619690	$(AC)_6T(AC)_4$	F-CATGGCCTATCTGTTCCGC AGACCAACATCCCAGTCGC	58	240	1	–	–
Spavo16§ JQ619691	$(AC)_5T(AC)_5$	F-GTTCAGGATGACCCGGTGG TGTGTATGAGTTCCTGCCC	56	168	1	–	–
Spavo17§ JQ619692	$(TC)_7$	F-TGTCAAGCTCACAGCGAC ATGGCACCCATGCTTCAGG	56 [a]	216	1	–	–
Spavo18§ JQ619693	$(GA)_7$	F-CCATGACCAACTACGACGAG GGAGCTTAGGTCGCTCACC	62	175	1	–	–
Spavo19§ JQ619694	$(CA)_3T(CA)_7$	F-ACCTTCCAGCCTACGAGAGC TGTGTCAGGAGTAGGCAGACC	62	170	1	–	–
Spavo20 JQ619695	$(AGC)_{10}$	F-TGCTCGGCTCTACGGTTC CCCTCACAGAGTTCACGGG	60	209−239	8	0.60	0.50
Spavo21 JQ619696	$(AATG)_{15}$	F-TGTGTTGGTTTGAGACGGC CCTCAAAGACATTGGATGCG	60	298−330	8	0.85	0.79
Spavo22 JQ619697	$(ATCC)_{14}$	H-GGCAGAAGGAAACCTGGAC GGCCCTTGAAACTCCACTCT	61	139−187	9	0.85	0.77
Spavo23 JQ619698	$(CATT)_8$	H-CGACCCATTTCGGTTACAAG GAACGAGTAACGTGATGCTGA	61	245−269	6	0.75	0.72
Spavo24 JQ619699	$(CTGT)_9$	F-GCTCCAACAGAGATAAAACGCTCT TCACTGTAGGAACACGGGAAT	62	170−182	4	0.30	0.27
Spavo25 JQ619700	$(CTGT)_{10}$	H-GAGTGAGCCGGAGTGTTCTG GGCTAAACTGTGGCTGCCTA	62	232−244	3	0.30	0.55*
Spavo26 JQ619701	$(GTTT)_9$	H-CACGTTGCCAATTCCAGTAG GAAGACGACAACCACTCTCAG	59	212−220	3	0.40	0.38
Spavo27 JQ619702	$(AAAC)_{13}$	F-GAGCTGGCGTTTCCCAAATA ACGGCGTAGTGAGCATGTTG	59	169−232	12	0.80	0.76
Spavo28 JQ619703	$(CTATT)_{10}$	H-GCAGAGTGACAATAAAGGACGA CCACAAGGCTCAGTTTGACA	59	292−328	7	0.75	0.68

Ta (ºC) – annealing temperature; Ho – observed heterozygosity; He – expected heterozygosity; k – number of alleles; "F-" or "H-" at the 5′ end of the primer indicates FAM- or HEX-labelled primer; Hardy-Weinberg expectation deviations, *P<0.05, **P<0.001.

a – Mg = 1.0 mM.

b – Mg = 1.75 mM.

§ – Strategy 1

All but five microsatellites (*Spavo15–Spavo19*) were polymorphic in the Culatra population. Since the apparently monomorphic microsatellites were isolated using the

first strategy and were therefore expected to be polymorphic, it is possible by increasing the number of individuals the polymorphism could be detected. In the other two populations, only two microsatellites (*Spavo15* and *Spavo18*) were not confirmed as polymorphic.

2.4.3. Relationship of *in silico* variability and microsatellite number of repeats with PCR polymorphism

The strategies described here were developed in order to achieve higher rates of polymorphism. The novelty in the approach relies on a pre-screening of microsatellites, based on their polymorphism *in silico* or based on their number of repeats. The relatively low success rate of nearly 45% of functional microsatellites obtained is in the range reported in other studies developing microsatellites from unigenes (mean = 65%; range 45%–76% (Csencsics et al., 2010; Hoffman and Nichols, 2011; Kim et al., 2008; Li et al., 2009, 2004; Vogiatzi et al., 2011). The proportion of polymorphic microsatellites was also comparable to those reported by Li et al. (2009) in oyster (15/29 microsatellite loci), Hoffman and Nichols (2011) in Antarctic fur seal (23/38 microsatellite loci) and Csencsics et al. (2010) in dwarf bulrush (17/22 microsatellite loci).

The success of the rate of microsatellite PCR amplification was higher using the first strategy (54.5%) compared to the second (33.3%) (Table 2). The difference may have resulted from the lower read coverage of the flanking regions of the microsatellites isolated using the second strategy, affecting the base call confidence of these regions where the primers were designed. However, the second strategy was more effective in yielding more highly polymorphic microsatellites (8.2 ± 3.85 alleles per locus), considering all different alleles observed in the three populations, than the first strategy (4.56 ± 2.45 alleles per locus) (unpaired *t* test, $t_{(24)} = 2.96$, P = 0.0069). DNA slippage may increase in proportion to the number of repeats so that microsatellite loci with more repeats generally show higher mutation rates, which could explain these results (Petit et al., 2005). Furthermore, Li et al. (2004) reported that microsatellites present in protein-coding regions (strategy 1), could lead to gain or loss of gene function via frameshift mutations, which could explain the lower allele richness found in these loci. However, long stretches of repeats may also accumulate imperfections that persist because they favour slippage reduction and consequently improve microsatellite stability (Bhargava and Fuentes, 2010; Zhu et al., 2000), which is important for microsatellites harboured in genes. Examples of this are *Spavo02*, *Spavo07*, *Spavo12*, *Spavo15*, *Spavo16* and *Spavo19*

loci, where one nucleotide was inserted or substituted interrupting the stretch of perfect repeats. Only in Spavo12 locus the smaller stretch of repeats was confirmed as polymorphic.

Table 2 – Summary of the microsatellite results obtained for each strategy. Microsatellite results are based on the 26 peacock blenny individuals genotyped from Culatra, Formentera and Borovac populations.

	Strategy 1	Strategy 2
Selected for application	33	30
Successfully genotyped	18	10
Polymorphic	16	10
Repeat length *in silico* [a]	7.17±1.69	11.8±4.16
Allelic richness [b]	4.56±2.45	8.2±3.85

Strategy 1 – *in silico* polymorphism with GO terms; strategy 2 – number of repeats.

[a] Welch's unpaired *t* test, $t_{(10.678)}$= 3.37, P=0.0065.

[b] unpaired *t* test, $t_{(24)}$=2.96, P=0.0069.

To our knowledge, only a recent study on Antarctic fur seal used an approach similar to our first strategy. Hoffman and Nichols (2011) selected microsatellites either on the basis of GO codes or high variability *in silico*, and obtained a positive relationship between the number of alleles *in silico* and the observed allele number. However, a lower number of polymorphic microsatellites were obtained (61% of the microsatellite loci compared to the 93% in the present study). A Generalized Linear Model (GLM) to evaluate the impact of the two strategies on the rates of polymorphism indicates that between the two predictor variables considered, number of alleles observed *in silico* and microsatellite number of repeats, only the latter was retained as a significant predictor variable (estimate = 0.75, $\chi 2 = 6.67$, P = 0.0098) in the reduced model. One explanation could be that the microsatellites were not as polymorphic *in silico* (1 to 3 alleles; Table S1, Appendix II) as with Hoffman and Nichols (2011), where the microsatellites had between 1 and 6 alleles. Some variation may have been lost during normalization of the cDNA library, although to some extent this may have been compensated by a larger number of unigenes sequenced as a result. No conclusions can be drawn in relation to which type of microsatellite is more prone to be polymorphic. Although tetranucleotides appear to be candidates, this may be because they were the type of microsatellites successfully applied (7/11 microsatellite loci) in the second strategy.

2.5. Data accessibility

DNA sequences − GenBank accessions JQ619676–JQ619703.

2.6. References

Abdelkrim J., Robertson B.C., Stanton J.-A.L. & Gemmell N.J., 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques, 46, 185–192.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M. & Sherlock G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics, 25, 25–29.

Bhargava A. & Fuentes F.F., 2010. Mutational Dynamics of Microsatellites. Molecular Biotechnology, 44, 250–266.

Chevreux B., Pfisterer T., Drescher B., Driesel A.J., Müller W.E.G., Wetter T. & Suhai S., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome research, 14, 1147–59.

Chistiakov D.A., Hellemans B. & Volckaert F.A.M., 2006. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. Aquaculture, 255, 1–29.

Csencsics D., Brodbeck S. & Holderegger R., 2010. Cost-Effective, Species-Specific Microsatellite Development for the Endangered Dwarf Bulrush (Typha minima) Using Next-Generation Sequencing Technology. Journal of Heredity, 101, 789–793.

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

Gish W. & States D.J., 1993. Identification of protein coding regions by database similarity search. Nature Genetics, 3, 266–272.

Gonçalves D., Matos R., Fagundes T. & Oliveira R., 2005. Bourgeois males of the peacock blenny, Salaria pavo, discriminate female mimics from females? Ethology, 111, 559–572.

Gonçalves E.J., Almada V.C., Oliveira R.F. & Santos A.J., 1996. Female mimicry as a mating tactic in males of the blenniid fish Salaria pavo. Journal of the Marine Biological Association of the United Kingdom, 76, 529–538.

Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J. & Conesa A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research, 36, 3420–3435.

Grattapaglia D., Silva-Junior O.B., Kirst M., de Lima B.M., Faria D.A. & Pappas G.J.,

2011. High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species. BMC Plant Biology, 11, 65.

Hoffman J.I. & Nichols H.J., 2011. A novel approach for mining polymorphic microsatellite markers in silico. PloS ONE, 6, e23283.

Ju Z., Wells M.C., Martinez A., Hazlewood L. & Walter R.B., 2005. An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, Fundulus, and Xiphophorus. In silico biology, 5, 439–463.

Kim K.S., Ratcliffe S.T., French B.W., Liu L. & Sappington T.W., 2008. Utility of EST-derived SSRs as population genetics markers in a beetle. Journal of Heredity, 99, 112–124.

Li Q., Liu S. & Kong L., 2009. Microsatellites within genes and ESTs of the Pacific oyster Crassostrea gigas and their transferability in five other Crassostrea species. Electronic Journal of Biotechnology, 12, 15–16.

Li Y.-C., Korol A.B., Fahima T. & Nevo E., 2004. Microsatellites Within Genes: Structure, Function, and Evolution. Molecular Biology and Evolution, 21, 991–1007.

Louro B., Passos A.L.S., Souche E.L., Tsigenopoulos C., Beck A., Lagnel J., Bonhomme F., Cancela L., Cerdà J., Clark M.S., Lubzens E., Magoulas A., Planas J. V, Volckaert F.A.M., Reinhardt R. & Canario A.V.M., 2010. Gilthead sea bream (Sparus auratus) and European sea bass (Dicentrarchus labrax) expressed sequence tags: Characterization, tissue-specific expression and gene markers. Marine Genomics, 3, 179–191.

Milne I., Bayer M., Cardle L., Shaw P., Stephen G., Wright F. & Marshall D., 2010. Tablet-next generation sequence assembly visualization. Bioinformatics, 26, 401–402.

Neff B.D. & Gross M.R., 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. Evolution, 55, 1717–1733.

O'Brien S.J., 1991. Molecular genome mapping lessons and prospects. Current Opinion in Genetic Development, 1, 105–111.

Petit R.J., Deguilloux M.-F., Chat J., Grivet D., Garnier-Géré P. & Vendramin G.G., 2005. Standardizing for microsatellite length in comparisons of genetic diversity. Molecular Ecology, 14, 885–890.

R Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rousset F., 2008. GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources, 8, 103–106.

Santana Q.C., Coetzee M.P.A., Steenkamp E.T., Mlonyeni O.X., Hammond G.N.A., Wingfield M.J. & Wingfield B.D., 2009. Microsatellite discovery by deep

sequencing of enriched genomic libraries. BioTechniques, 46, 217–223.

Scaglione D., Acquadro A., Portis E., Taylor C.A., Lanteri S. & Knapp S.J., 2009. Ontology and diversity of transcript-associated microsatellites mined from a globe artichoke EST database. BMC Genomics, 10, 454.

Seeb J.E., Carvalho G., Hauser L., Naish K., Roberts S. & Seeb L.W., 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular ecology resources, 11 Suppl 1, 1–8.

Selkoe K.A. & Toonen R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology letters, 9, 615–629.

Serapion J., Kucuktas H., Feng J. & Liu Z., 2004. Bioinformatic Mining of Type I Microsatellites from Expressed Sequence Tags of Channel Catfish (Ictalurus punctatus). Marine Biotechnology, 6, 364–377.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T. & Isobe S., 2012. In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biology, 12, 80.

Slate J., Hale M.C. & Birkhead T.R., 2007. Simple sequence repeats in zebra finch (Taeniopygia guttata) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. BMC Genomics, 8, 52.

Tang J., Baldwin S.J., Jacobs J.M., van der Linden C.G., Voorrips R.E., Leunissen J.A., van Eck H. & Vosman B., 2008. Large-scale identification of polymorphic microsatellites using an in silico approach. BMC Bioinformatics, 9, 374.

Tóth G., Gáspári Z. & Jurka J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research, 10, 967–981.

Väli Ü., Brandström M., Johansson M. & Ellegren H., 2008. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genetics, 9, 8.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4, 535–538.

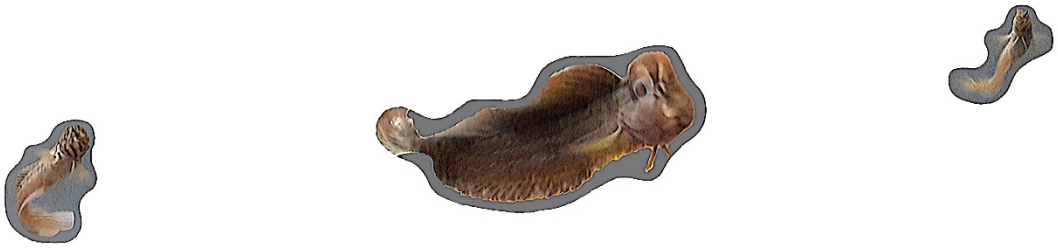
Vera J.C., Wheat C.W., Fescemyer H.W., Frilander M.J., Crawford D.L., Hanski I. & Marden J.H., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology, 17, 1636–1647.

Vogiatzi E., Lagnel J., Pakaki V., Louro B., V.M. Canario A., Reinhardt R., Kotoulas G., Magoulas A. & Tsigenopoulos C.S., 2011. In silico mining and characterization of simple sequence repeats from gilthead sea bream (Sparus aurata) expressed sequence tags (EST-SSRs); PCR amplification, polymorphism evaluation and multiplexing and cross-species assays. Marine Genomics, 4, 83–91.

Zane L., Bargelloni L. & Patarnello T., 2002a. Strategies for microsatellite isolation: a review. Molecular Ecology, 11, 1–16.

Zane L., Patarnello T., Ludwig A., Fontana F. & Congiu L., 2002b. Isolation and characterization of microsatellites in the Adriatic sturgeon (Acipenser naccarii). Molecular Ecology Notes, 2, 586–588.

Zhu Y., Strassmann J.E. & Queller D.C., 2000. Insertions, substitutions, and the origin of microsatellites. Genetical research, 76, 227–236.



CHAPTER 3 |

PATERNITY ASSESSMENT USING GENETIC MARKERS

Cardoso S.D., Faustino A.I., Costa S.S., Valério F., Gonçalves D. & Oliveira R.F., 2017. Social network predicts loss of fertilizations in nesting males of a fish with alternative reproductive tactics. *acta ethologica*, *20*, 59–68.
doi: 10.1007/s10211-016-0249-9

Social network predicts loss of fertilizations in nesting males of a fish with alternative reproductive tactics

Sara D. Cardoso 1,2,3, Ana I. Faustino 1,2,3, Silvia S. Costa2, Fábio Valério2, David Gonçalves 4, Rui F. Oliveira 1,2,3

[1] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal; [2] ISPA, Instituto Universitário, Rua Jardim do Tabaco 34, 1149-041 Lisbon, Portugal; [3] Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Avenida Brasília, 1400-038 Lisbon, Portugal; [4] Institute of Science and Environment, University of Saint Joseph, Rua de Londres 16, Macau, SAR, China.

3.1. Abstract

Abstract Alternative reproductive tactics (ARTs) evolve when there is strong intra-sexual competition between conspecifics for access to mates. Typically, larger bourgeois males reproduce by securing the access to reproductive resources while smaller parasitic males reproduce by stealing fertilizations from larger males. A number of factors can influence the reproductive success of each tactic, including intrinsic (e.g. size) and extrinsic (e.g. tactic relative frequency) variables. An example where plastic ARTs occur is the peacock blenny *Salaria pavo*, with large males reproducing by defending nests and attracting females (bourgeois tactic) and small males reproducing by achieving sneaked fertilizations (parasitic tactic). In this study, we conducted field observations on individually tagged animals to determine their social network and collected eggs from 11 nests to determine the fertilization success of each male tactic. Paternity estimates for 550 offspring indicated an average fertilization success for nest-holder males of 95%. Nest-holder male morphological traits and social network parameters were tested as predictors of fertilization success, but only the number of sneakers present in the nest-holder's social networks was found to be a predictor of paternity loss. Although male morphological traits had been previously found to be strongly correlated with reproductive success of nest-holder males, as measured by the number of eggs collected in the male's nest, no correlation was found between any of the measured morphological traits and fertilization success for these males. The results suggest a stronger influence of the social environment than of morphological variables in the proportion of lost fertilizations by nest-holder males of this species.

Keywords: *Salaria pavo*; Social network analysis; Paternity estimates; Blenniidae; Fertilization success

3.2. Introduction

In some species, individuals of the same sex adopt alternative reproductive tactics (ARTs) to get access to mates in response to strong sexual competition (Taborsky et al., 2008). ARTs have been well characterized in teleost fishes (Taborsky, 1994), where typically two consistent and discrete male reproductive phenotypes occur within the same population, each one trying to maximize their fitness by allocating resources to one tactic. One male morph makes the investment of getting privileged access to females by defending a nest and presenting attractive traits for females, such as courtship displays and secondary sexual characters, while the alternative male morph exploits the conspecifics' investment by stealing fertilizations in sneaky or coercive ways. The former males are usually called bourgeois or parental, the later parasitic or sneakers (Taborsky, 1997). These tactics can be fixed for an entire lifetime, or individuals may switch tactics (plastic ARTs) in a reversible or irreversible manner (Brockmann, 2001; Moore, 1991; Taborsky et al., 2008). Switching from the parasitic to the bourgeois tactic is characteristic of sequential tactics, which are viewed as being conditionally triggered by developmental thresholds (i.e. status-dependent) instead of being genetically determined (Gross, 1996). In these cases, fitness is not expected to be equal among alternative morphotypes. Moreover, the fitness of each male type may also be influenced by the frequency of both male tactics in the population (i.e. frequency-dependent), so that the fitness gain of sneakers will be dependent on the proportion of bourgeois males that they can exploit in the population, this balance being closely related to the bourgeois males' loss in fitness (Gross, 1996). Fitness can be measured as reproductive success, and in the case of bourgeois males, it consists of the number of eggs spawned in the nest (i.e. mating success) that are fertilized by them (i.e. fertilization success). With the development of molecular techniques to quantify the fertilization success of each male reproductive tactic, it is possible to better understand how these tactics evolve and are maintained in populations.

The peacock blenny, *Salaria pavo*, is an emerging model in the integrative study of proximate and ultimate causes of ARTs and plasticity of mating systems (Oliveira, 2006; Saraiva et al., 2013). This blenny is a small intertidal fish inhabiting coastal lagoons and

rocky shores along the Mediterranean and adjacent Atlantic coasts (Zander, 1986). Bourgeois males are larger than females and have well-developed secondary sexual characters (viz. a head crest and an anal gland), which are used to attract females to their nests for spawning (Barata et al., 2008; Gonçalves et al., 2002a), after which these males provide sole parental care to eggs until hatching (Fishelson, 1963; Patzner et al., 1986). Moreover, due to the promiscuous mating system, where males and females may spawn with multiple mates throughout the breeding season, it is possible to find in each nest several batches of eggs in different developmental stages (Patzner et al., 1986).

Peacock blenny males can follow one of two developmental pathways, with some males growing directly into the bourgeois tactic while others first behave as female mimics to sneak fertilizations before switching to the bourgeois tactic at a later stage (Almada et al., 1994; Fagundes et al., 2015; Gonçalves et al., 1996). These alternative life histories are sequential and consistent with body size (i.e. condition) that each male has at sexual maturity in their first breeding season, as a consequence to their hatching time (Fagundes et al., 2015). Early born individuals arrive at their first breeding season with a body size that is too large for them to successfully mimic females, and therefore, they continue growing to become bourgeois males. In contrast, late born individuals arrive at their first breeding season with a small body size that allows them to mimic females' morphology and courtship displays (Gonçalves et al., 2005), switching to the nesting tactic in the following breeding season and going through a phase in which they are reproductively inactive (i.e. transitional males) (Fagundes et al., 2015).

To understand the dynamics of ARTs in *S. pavo*, an estimate of the success of nesting and sneaker tactics is required. Previous work has repeatedly established a strong positive correlation between the nesting male's body size and its mating success, as measured by the number of eggs in the nest, even when controlling for nest size (Gonçalves et al., 2002b; Oliveira et al., 1999). The recent development of genetic markers for *S. pavo* (Cardoso et al., 2013) now allows for estimation of the fertilization success of the male tactics in this species by doing paternity tests in nest-holder male's nests. In this study, we aimed to first confirm that nest-holders are subject to cuckoldry, showing that sneakers fertilize eggs, and then assess how biological and social factors affect the fertilization success of nesting males.

3.3. Materials and methods

3.3.1. Study site and field observations

This study was conducted during the peak of the breeding season (i.e. beginning of June until the end of the first week of July) in 2009 at Culatra Island (36° 59′ N, 7° 51′ W, Algarve, Southern Portugal). In this population, *S. pavo* nesting sites are very scarce and males mainly nest in artificial reefs made of bricks used by clam culturists to delimit their fields (for a detailed description of the habitat, see (Almada et al., 1994)). An area of approximately 422 m^2 containing four experimental transects, totalling 104 aligned bricks (T1 to T4; Fig. 1), and four sets of tiles distributed in the periphery (sets of tiles A to D; Fig. 1) were monitored throughout the study. A total of 152 fish (58 females, 64 bourgeois males, 24 sneakers and 6 transitional males) were captured during low tide along these transects and adjacent regions. These fish were individually tagged with a combination of coloured beads (diameter = 0.3 cm) attached to a fishing line inserted at the base of the dorsal fin, following the procedure described by Patzner (1984), and previously proved successful in this species (Fagundes et al., 2007; Gonçalves et al., 2005, 2003; Saraiva et al., 2009). To minimize stress, fish were lightly anaesthetized with MS-222 (tricaine methanesulfonate, dilution 1:10,000, Sigma-Aldrich) before manipulation. The male morphotypes were discriminated as follows: bourgeois, males with fully developed secondary sexual characters (SSCs) that are actively defending a nest (i.e. nest-holders) or found wandering the transects (i.e. floaters); sneakers, males lacking male secondary sexual characters and from which sperm could be easily extruded from their vas deferens by gently pressing the abdomen; and transitional, sneakers with poorly developed SSCs undergoing the transition into bourgeois males and so not reproductively active. For each individual, measures of standard length, head height, body height, total length and width of the anal gland (modification of the first two rays of the anal fin) were taken using a Vernier calliper to the nearest 0.01 cm. A small section of the dorsal fin tissue was also collected and preserved in absolute ethanol at 4 ºC for genetic analyses. Fish recovered from anaesthesia in a container with abundant water and aeration and were returned to the brick/shelter from where they were captured. No mortality due to manipulation or anaesthesia occurred. Field observations consisted of inspections (n = 9) during low tide, when all bricks and tiles were exposed, and scans (n = 10) of all bricks during high tide while snorkelling, always recording fish positions (i.e. brick or tile). At the end of the first week of July, a sample of eggs accessible from the

nest entrance was collected from 11 nests randomly chosen and preserved in absolute ethanol at 4 ºC.

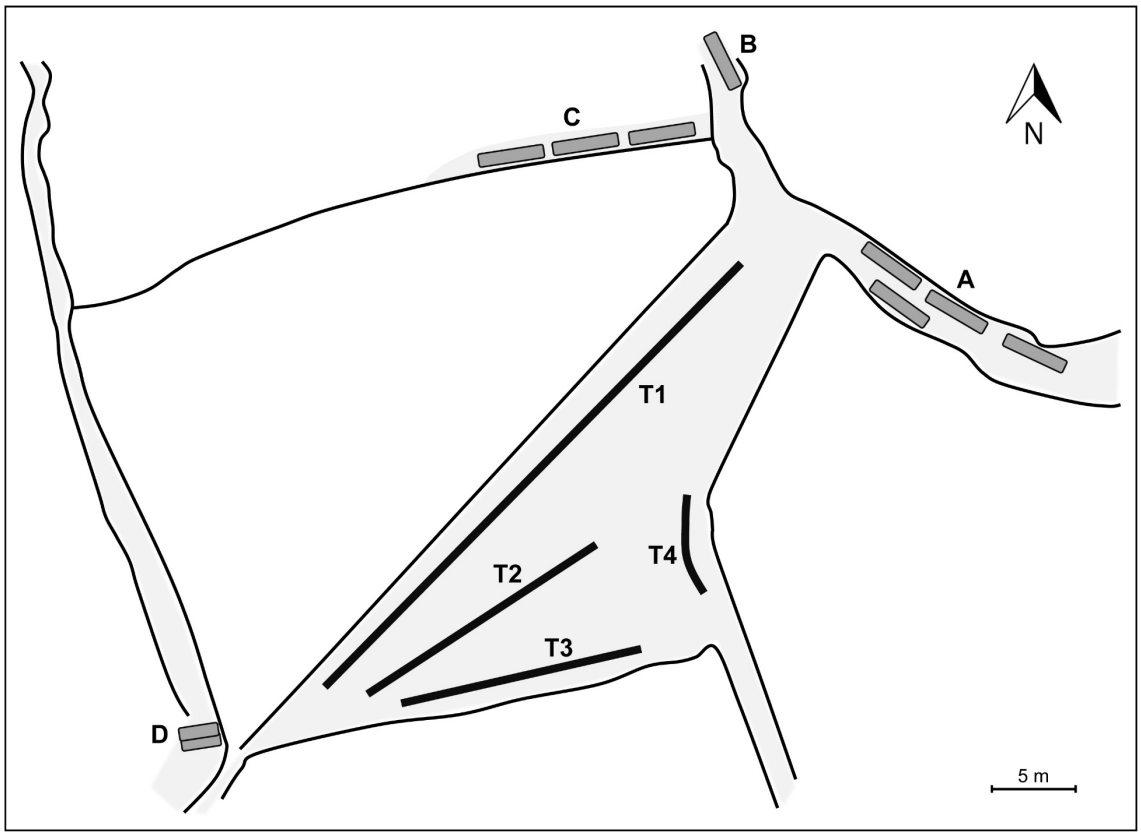


Fig. 1 – Schema of the field site located in the intertidal zone at Culatra Island, representing an area of approximately 422 m^2 containing four experimental transects (*T1* to *T4*) and four sets of tiles distributed in the periphery (sets of tiles *A* to *D*). Shaded areas represent artificial channels that run through this part of the field and become shallow or submerged during low tide and high tide, respectively.

3.3.2. Genetic analyses

Five microsatellite markers previously isolated from the peacock blenny (*Spavo08*, *Spavo21*, *Spavo22*, *Spavo23*; (Cardoso et al., 2013)) and shanny *Lipophrys pholis* (*6_6*; (Guillemaud et al., 2000)) were used to assess paternity at 11 nests. For each of these nests, 50 eggs were randomly selected for genotyping. Larvae were extracted from their eggs using a stereo microscope (VWR SZB 200) to avoid possible contaminations with maternal DNA. In order to determine the allele frequencies of the breeding population, 144 adults (two samples were lost), including the parental male at each nest, were also used for genotyping. Genomic DNA was extracted from larvae and fin samples using Extract-N-Amp™ Tissue PCR Kit (Sigma-Aldrich) following standard procedures.

Polymerase chain reaction (PCR) amplifications were carried out in 25 µl volume reactions on a thermal cycler (Stratagene RoboCycler® Gradient 96) using fluorescent dyed forward primers (Eurofins MWG Operon, Ebersberg, Germany), following the protocol outlined in Cardoso et al. (2013). Locus *6_6* was optimized in this study with an annealing temperature of 59 ºC and 1.75 mM of $MgCl_2$. For larvae genotyping, only the $MgCl_2$ concentrations were optimized as follows: *Spavo08* and *Spavo21* at 2.25 mM, *Spavo22* at 2.5 mM and *Spavo23* at 2 mM. Since these microsatellites had different distributions of allele sizes and fluorescent labels, PCR products from each individual were pooled and sent for analysis. DNA fragments were separated on a commercial ABI 3730XL DNA Analyser (Macrogen Inc., Korea) and sized by co-running a GeneScan™ 400HD (Applied Biosystems, Foster City, CA, USA) size standard. DNA fragments were scored manually with the aid of GeneMarker® version 1.95 (SoftGenetics, State College, PA, USA).

Tests for Hardy-Weinberg equilibrium and genotypic linkage disequilibrium were performed using GENEPOP version 4.2.1 (Rousset, 2008) with the default setting and the significance levels adjusted for multiple comparisons using the sequential Bonferroni correction with $\alpha = 0.05$ (Rice, 1989). Genetic diversity estimates, including expected (He) and observed (Ho) heterozygosities, and exclusion probabilities, for one candidate parent (P_{E1}) and for one candidate parent given the genotype of a known parent of the opposite sex (P_{E2}), were obtained using CERVUS version 3.0.3 (Kalinowski et al., 2007).

3.3.3. Paternity analyses

Estimations of paternity for the parental males were obtained using the *two-sex paternity* model developed by Neff et al. (2000a, 2000b) and Neff (2001). This model was selected since it does not require the sampling of all potential parents when the mating system is promiscuous, as in the peacock blenny, where both males and females may spawn with multiple mates and nest-holder males may be the target of sneaked fertilizations. This model requires the genotype of the parental male, the genotypes of the brood in his nest and allele frequencies in the breeding population and gives as an output a paternity measure (*Pat*) expressed as a proportion of NG_{dad} and ng_{dad} (Neff et al., 2000b, 2000a). NG_{dad} is the expected proportion of the offspring that can be genetically compatible with the putative father by chance, calculated independently for each nest

using information from all loci; ng_{dad} is the proportion of the offspring that is genetically compatible with the putative father (Neff et al., 2000a). Additionally, paternity estimates were also obtained for tagged sneakers observed near the nests of nest-holder males assessed in this study. In these analyses, we considered all the eggs genotyped in each nest (n = 50 per nest) and also tested if eggs that were not genetically compatible with the parental male could have been fathered by these sneakers.

3.3.4. Social network analysis

Of the initial 152 individuals marked, only 105 were subsequently observed (35 females, 53 bourgeois males, 15 sneakers and 2 transitional males). A total of 520 observations were registered, 371 from individuals identified at the beginning of this study (mean ± standard error (SE) = 3.53 ± 0.18 observations per individual) and 149 from unknown individuals, which were not used for further analyses. In order to obtain the social network for this population, a two-dimensional matrix was first created, consisting of associations between individuals and the respective brick or tile in which they were seen (two-mode data). As the main focus of this work was to assess the fertilization success of nest-holder males in their nests, we choose to use the location of each individual relative to a brick (where nests are located) or tile (where non-breeding individuals usually seek shelter) and, from there, extract the relationships between individuals. Therefore, a relationship between two individuals is drawn only if they had visited the same brick or were sheltered under the same tile together at the same time or at different times. This type of social network is called an affiliation network and was obtained from the two-mode data matrix by extracting relationships between individuals based on common visits to bricks and tiles (one-mode data) using the function *data* > *affiliations* tool present in UCInet version 6.488 (Borgatti et al., 2002). This matrix was converted in order to be unweighted, and various measures were obtained using UCInet. As a measure of prominence for each individual in the network, the *Eigenvalue* centrality was selected since it takes into account how well an individual is connected (i.e. degree centrality), weighted by the number of connections his relationships also have (Makagon et al., 2012). To detect different social groups present in the network (i.e. network structure), the Girvan-Newman algorithm (Girvan and Newman, 2002) was used. Network graphs of the relational matrix were drawn using the spring embedding based on distance in Netdraw version 2.118 (Borgatti, 2002).

3.3.5. Statistical analyses

In this study, the effects of two types of variables (morphometric and social) on nest-holder males' fertilization success were tested. Morphometric variables comprised measurements of morphological traits related to territoriality and mate attraction (i.e. secondary sexual characters), namely the standard length (SL), head crest size (ratio between head height and body height taken at the insertion of the pectoral fins) and relative anal gland area (determined by assuming an ellipsoid shape for this structure and dividing this value by SL). We did not examine body mass because length and mass are highly correlated in the peacock blenny (Oliveira et al., 1999). Social variables comprised measures obtained from the social network analysis, namely the nest-holder male's *Eigenvector* centrality, egonetwork heterogeneity (EgoNet heterogeneity) and proportion of the egonetwork comprised by sneakers (EgoNet sneaker).

Statistical analyses were performed using R (R Core Team, 2014). Data were square root transformed to comply with the normality assumption and tested with Jarque-Bera normality test. Correlations between morphometric and social variables were examined with Pearson correlation coefficient (r_P) with the significance level adjusted for multiple comparisons using the sequential Bonferroni correction with $\alpha = 0.05$. A multiple regression model was used to predict nest-holder males' fertilization success using as predictors morphometric and social variables previously mentioned. Univariate tests of significance, effect size and observed power ($\alpha = 0.05$) for each predictor variable in the model were also obtained. Descriptive analyses are reported as mean ± SE.

3.4. Results

3.4.1. Paternity and social network analyses

The number of alleles found in the peacock blenny population of Culatra varied between 6 and 14 alleles per locus (mean of 9.6 alleles; Table 1). All loci were in Hardy-Weinberg equilibrium and were considered statistically independent since no linkage disequilibrium between loci pairs was observed after Bonferroni correction.

Paternity estimates were obtained for 11 different nests, from which 50 eggs were randomly selected for genotyping. Using all paternity estimates calculated, the mean paternity for nest-holder males was 95.34% ± 1.97 (range 83.36–100%; Table 2). Of the 11 nests analysed, we detected cuckoldry in five, with two of these nests having a paternity loss greater than 10%. The mean NG_{dad} value obtained for each nest was 0.11 ±

0.01 (range 0.07–0.19), indicating that on average, 89% of all individuals in this population could be excluded as a potential father by chance alone.

Table 1 – Summary statistics for five microsatellite loci (Cardoso et al., 2013; Guillemaud et al., 2000) used to characterize the peacock blenny (*Salaria pavo*) breeding population (n = 144) at Culatra Island.

Locus	N_A	Size (bp)	Heterozygosity		Excl. Prob.	
			Obs.	Exp.	P_{E1}	P_{E2}
Spavo08	10	216-242	0.78	0.79	0.43	0.61
Spavo21	11	292-336	0.81	0.77	0.38	0.56
Spavo22	14	129-189	0.84	0.80	0.43	0.61
Spavo23	7	245-269	0.65	0.70	0.28	0.45
6_6	6	295-311	0.63	0.60	0.18	0.31
Combined P_E					0.88	0.97

N_A, number of alleles; Size, range of allele lengths in base pairs (bp); P_{E1}, exclusion probability when neither parent is known; P_{E2}, exclusion probability when one parent is known.

The social network analysis allowed us to examine the social dynamics relative to this subpopulation of *S. pavo* at Culatra Island. The network obtained in this study had 1314 ties between 105 individuals (mean number of ties per individual 12.51 ± 0.93; Fig. 2). The proportion of individuals observed in the consecutive days after being tagged was approximately 69%; transitional males were less re-observed, followed by females and sneakers (see the "Materials and methods" section). The social network presented a structured pattern similar to the spatial structure of experimental transects (Fig. 1), although this was not supported by structural analyses used to identify different social groups. Only nest-holder males defending nests allocated to the largest transect (T1) were distributed in three different groups across the network.

From the social network, individual networks (i.e. egonetworks) for each nest-holder male assessed for paternity were obtained. There was great variability in the composition and number of associations across all egonetworks (Fig. 3). For the six nests in which no sneaked fertilizations were detected, only in two cases were sneakers associated (nests 3 and 11), whereas for the five nests with sneaked fertilizations, only in one case (nest 2) was there no association with a sneaker male (Fig. 3). For these sneakers, additional paternity estimates were obtained to confirm whether or not their success in fertilizing eggs in nests with which they were associated (Table 3). When all genotyped eggs in each nest were considered, the paternity estimates obtained for the nine cases of possible

sneaking showed that only in four cases sneakers could have fathered some of the eggs, including a nest where all the eggs were compatible with the parental male (nest 11). When analysing only eggs not fathered by the parental male, in two nests, these eggs remained without a known father (nests 2 and 9), while for the other cases, the sneakers associated with the nest could have fathered some of these eggs.

Table 2 – Summary statistics of paternity (Pat) analyses obtained from 50 eggs sampled at each nest in peacock blenny population at Culatra Island, calculated according to Neff et al. (2000b, 2000a), with a 95% confidence interval (CI).

Nest	NG_{dad}[a]	ng_{dad}[b]	Pat (%)	95% CI
1	0.08	1.00 (50)	100	93−100
2	0.12	0.86 (43)	84.12	74−93
3	0.11	1.00 (50)	100	93−100
4	0.19	0.92 (46)	90.14	81−97
5	0.07	0.96 (48)	95.72	87−99
6	0.12	0.96 (48)	95.44	86−99
7	0.11	1.00 (50)	100	94−100
8	0.08	1.00 (50)	100	93−100
9	0.16	0.86 (43)	83.36	74−93
10	0.11	1.00 (50)	100	93−100
11	0.12	1.00 (50)	100	94−100
Mean	0.11	0.96	95.34	−
Range	0.07-0.19	0.86-1.00	83.36-100	−

[a] The expected proportion of offspring compatible with the parental male by chance alone;

[b] Proportion of offspring compatible with the parental male and corresponding number of eggs within parentheses.

3.4.2. Fertilization success of nest-holder males

Correlations between the variables used as predictors of fertilization success for the nest-holder males under study were inspected (Supplementary Table 1, Appendix III). For the morphometric variables, a negative correlation between standard length and crest size ($r_P = -0.68$, $n = 11$, $P = 0.021$) was found. For the social variables, only a positive correlation between the Eigenvalue centrality and the EgoNet heterogeneity ($r_P = 0.78$, $n = 11$, $P = 0.005$) was obtained, indicating that males who were more central in the network also had more heterogeneous egonetworks. After controlling for multiple comparisons, both correlations became non-significant.

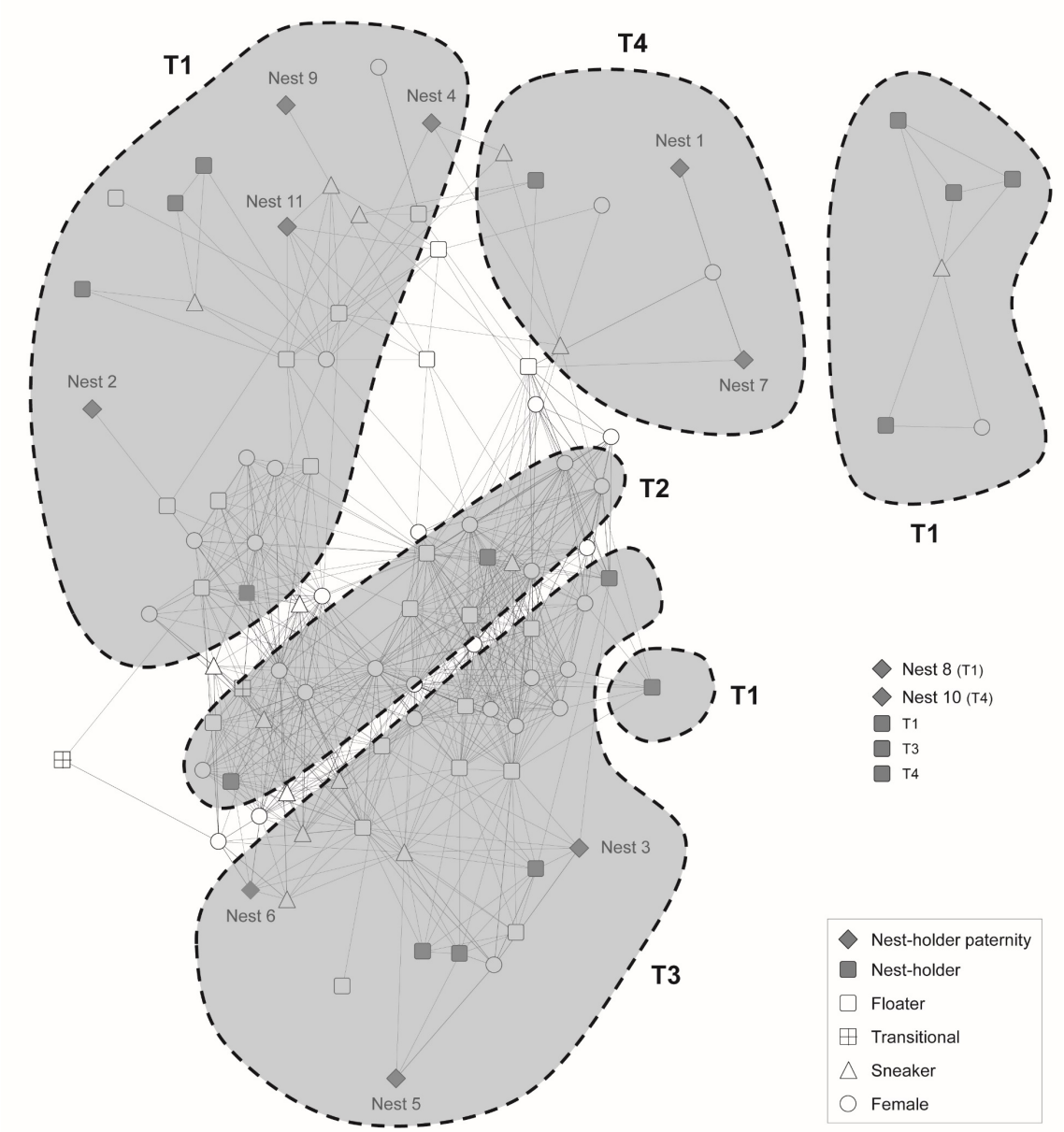


Fig. 2 – Affiliation network depicting social relationships between individuals marked at the beginning of the breeding season. Social ties between individuals are unweighted and present when the individuals visited the same brick or tile. Symbol colour designates behaviour: white symbols represent wandering and black symbols represent site attachment. Symbol shape indicates sex and morphotype: ♦ nest-holder male assessed for paternity; □ nest-holder not assessed for paternity or floater; ⊞ transitional male; △ sneaker and ○ female. Shaded areas correspond to the location of nests in the experimental transects (*T1* to *T4*).

To examine the influence of morphometric and social traits on nest-holder male's fertilization success, we conducted a multiple regression analysis, incorporating standard length, crest size, relative anal gland area, EgoNet sneaker and EgoNet heterogeneity and centrality. The results of this regression indicated the six predictors explained 73.8% of the variance in fertilization success, although not statistically significant ($R^2 = 0.74$, $F_{(6,4)} = 1.88$, $P = 0.282$; Table 4). Only the proportion of sneakers present in the egonetworks was found to be a predictor of the nest-holder male's fertilization success ($\beta = -1.33$, $P = 0.045$; Table 4).

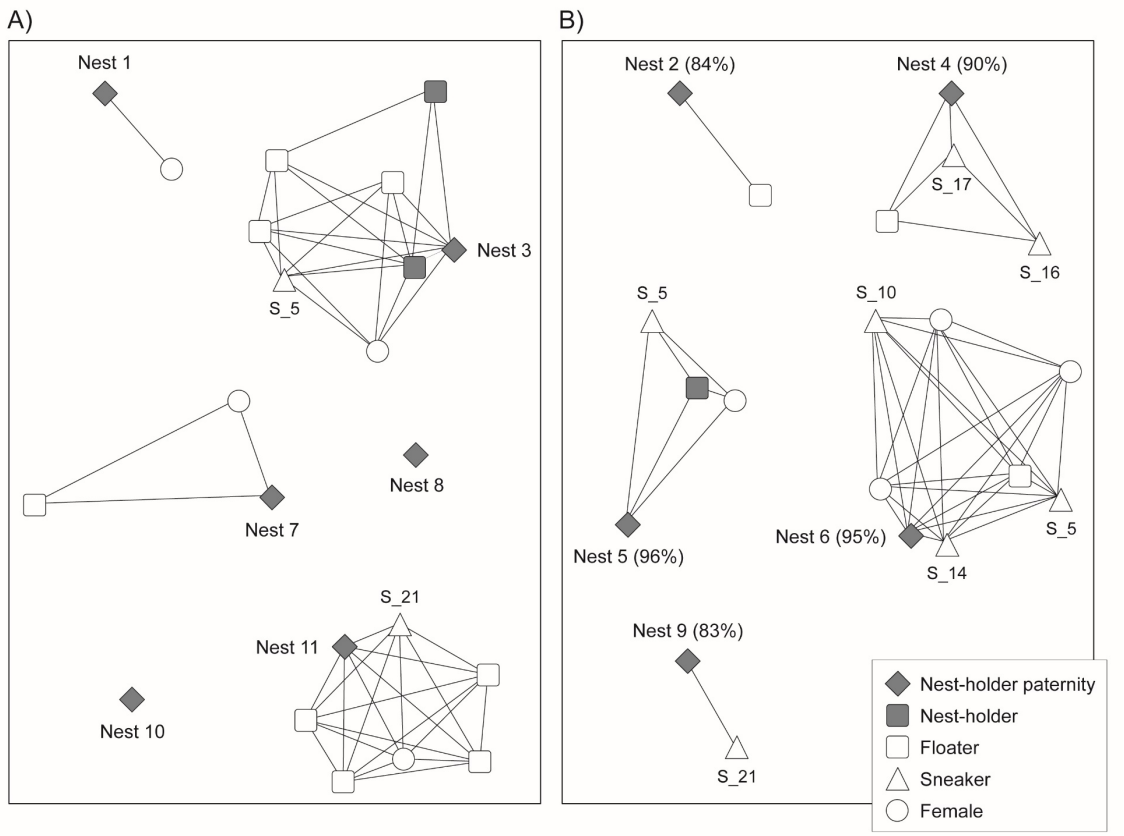


Fig. 3 – Egonetworks depicting social relationships for a) nest-holder males with no sneaked fertilizations, and b) nest-holder males with sneaked fertilizations, including the respective paternity estimates. Social ties between individuals are unweighted and present when individuals visited the brick where the nest of the focal male was located. Symbol colour designates behaviour: white symbols represent wandering and black symbols represent site attachment. Symbol shape indicates sex and morphotype: ♦ nest-holder male assessed for paternity; □ nest-holder not assessed for paternity or floater; △ sneaker and ○ female.

Table 3 – Summary statistics of paternity (Pat) analyses obtained for sneakers present in the social networks of the nest-holder males assessed in this study.

Nest	Sneaker ID	All eggs				Sneaked eggs				
		NG_{dad}[a]	ng_{dad}[b]	Pat (%)	95% CI	*n*	NG_{dad}[a]	ng_{dad}[b]	Pat (%)	95% CI
1	–	–	–	–	–	0	–	–	–	–
2	–	–	–	–	–	7	–	–	–	–
3	S−05	0.17	0.12 (6)	0	0.06−0.24	0	–	–	–	–
4	S−16	0.28	0.34 (17)	8.12	0.22−0.48	4	0.28	0.50 (2)	30.39	0.15−0.85
	S−17	0.28	0.44 (22)	22.52	0.31−0.58		0.28	0.50 (2)	30.83	0.15−0.85
5	S−05	0.17	0.08 (4)	0	0.03−0.19	2	0.17	0.50 (1)	40.13	0.09−0.91
6	S−05	0.17	0.28 (14)	13.79	0.18−0.42	2	0.17	0.50 (1)	40.13	0.09−0.91
	S−10	0.02	0	0	0.00−0.06		0.02	0	0	0.00−0.71
	S−14	0.08	0.06 (3)	0	0.02−0.16		0.08	0.50 (1)	45.41	0.10−0.91
7	–	–	–	–	–	0	–	–	–	–
8	–	–	–	–	–	0	–	–	–	–
9	S−21	0.15	0.02 (1)	0	0.01−010	7	0.15	0.14 (1)	0	0.03−0.53
10	–	–	–	–	–	0	–	–	–	–
11	S−21	0.15	0.20 (10)	6.29	0.11−0.33	0	–	–	–	–
Range		0.02−0.28	0−0.44	0−22.52	–	0−7	0.02−0.28	0.14−0.50	0−45.41	–

Estimates were obtained considering either all 50 eggs genotyped in each nest or only the eggs that were genetically incompatible with the parental male (*n*, sneaked eggs).

[a] The expected proportion of offspring compatible with the sneaker male by chance alone;

[b] Proportion of offspring compatible with the sneaker male and corresponding number of eggs within parentheses.

Table 4 – Multiple regression model of nest-holder male's fertilization success.

Variable	Coefficient (β)	*SE*	*t*	*P* value	eta-squared	Observed power ($\alpha = 0.05$)
Standard length	−0.552	0.515	−1.071	0.344	0.058	0.133
Crest size	−0.789	0.533	−1.478	0.214	0.111	0.209
Relative anal gland area	0.509	0.389	1.309	0.261	0.087	0.174
EgoNet sneaker	−1.334	0.462	−2.884	0.045	0.423	0.587
EgoNet heterogeneity	0.652	0.432	1.508	0.206	0.116	0.215
Centrality	−0.080	0.555	−0.144	0.892	0.001	0.051

The model has the following statistics, $R^2 = 0.74$, $F_{(6,4)} = 1{,}88$, $P = 0.282$

3.5. Discussion

In this study, we report the first estimates of paternity for ARTs in the peacock blenny *S. pavo*; nest-holder males had an average fertilization success of 95%, leaving the remaining offspring assumed to be sired by sneaker males. The incidence of sneaked fertilizations obtained in this work is one of the lowest values published in the literature for species with paternal care of eggs (for a review see (Avise et al., 2002; Coleman and Jones, 2011)), only surpassing the reported estimates in the Centrarchid family (reviewed in (Mackiewicz et al., 2005)). It is still debated whether the high investment made by males in species with extensive parental care to eggs, which, in blennies, includes nest defence and cleaning, egg fanning (Almada and Santos, 1995) and antimicrobial protection (Giacomello et al., 2006), strongly selects against sneaked fertilizations (e.g. (Alonzo, 2010; Kokko and Jennions, 2008; Trivers, 1972; Westneat and Sherman, 1993)). Generally, the proportion of eggs fertilized by sneakers is much smaller than the proportion fertilized by parental males. However, in a recent study, Cogliati et al. (2013) found for the plainfin midshipman fish (*Porichthys notatus*) the lowest levels of fertilization success for the nesting tactic described so far in fish, which was only 63% success when excluding from the analysis potential events of nest takeover.

The fertilization success detected for peacock blenny sneakers in this study could be considered a low payoff due to their high frequency and investment in female mimicry. Two main constraints imposed on sneakers may explain their low success. First, in this population, sneakers depend on female mimicry to successfully approach nest-holders, lacking male secondary sexual characters and adopting female nuptial colouration and courtship behaviours towards nest-holder males (Gonçalves et al., 1996; Saraiva et al.,

2013). The efficiency of this tactic is size dependent, as nest-holder males court and attack small sneakers and females with equal frequency, but as sneakers grow, they receive proportionally more attacks and less courts (Gonçalves et al., 2005). Furthermore, due to the existence of a female-biased operational sex ratio in this population (Oliveira et al., 1999), nest-holder males can be 'choosy' regarding potential mates. Consequently, it can be expected that a preference for larger females, as seen in a previous laboratory study (Gonçalves and Oliveira, 2003), may play an important role in avoiding sneaking events.

Second, nest characteristics used by nest-holder males as spawning sites, holes in bricks with only one partially obstructed entrance (Almada et al., 1994), may also play a role. The nest characteristics requires sneakers to deceive the nest-holder males in letting them enter inside the nest following a mating event; otherwise, they are left with only the option of releasing sperm from outside the nest during the narrow window of the spawning event. This hypothesis was recently tested in the cichlid *Lamprologus lemairii*, where territorial males use rock holes as nesting sites (Ota et al., 2014). This study showed that when considering both pre- (body size of territorial males and size of nest opening) and post-mating (milt quality traits) factors, larger territorial males spawning in nests with narrower openings had greater siring success (Ota et al., 2014). In contrast, species that reproduce using natural substrates more exposed to conspecifics are expected to be subject to a higher rate of parasitic fertilizations. For example, in the ocellated wrasse (*Symphodus ocellatus*), where territorial males build nests from algae, parasitic fertilizations were detected on 100% of the analysed nests and, on average, 28% of all offspring were not sired by the parental male (Alonzo and Heckman, 2010). Therefore, nest characteristics may also be a factor influencing the rate of success of sneakers in the peacock blenny.

However, when these fertilization rates take into account that: (1) each nest-holder male has in his nest, on average, approximately 15,000 eggs in different developmental stages (Gonçalves et al., 2002b), and (2) depending on the incubation temperature, eggs can hatch within 8 to 16 days (Westernhagen, 1983); sneakers may be able to fertilize thousands of eggs during the whole breeding season.

Reproductive success can be seen as being composed of two components: mating success (i.e. how successful an individual is in obtaining mating events) and fertilization success (i.e. how successful an individual is in fertilizing gametes on each mating event). The distinction between these two components of reproductive success is important when studying factors that affect Darwinian fitness in behavioural ecology, since different sets

of factors may be influencing pre- and post-mating selection. In the peacock blenny, mating success, measured as the number of eggs present in the nest defended by nest-holder male, has been shown to correlate with male body size, suggesting that peacock blenny females are using this character as a mate choice criterion (Gonçalves et al., 2002b; Oliveira et al., 1999). However, fertilization success is not related either to male body size or to any other morphometric variable analysed in this study. In contrast, the proportion of sneaker males present in the nest-holder males' networks was the only significant predictor of fertilization success of nest-holder males, with nests in which sneakers were more commonly observed having a higher probability of having sneaked fertilizations. This result is consistent with previous work, where nests that received more eggs during the breeding season had a higher number of different sneakers associated with them (Gonçalves et al., 2003). Together, these results suggest a stronger influence of the social environment than of morphological variables in the proportion of lost fertilizations by nest-holder males. Contrasting this result, Alonzo and Heckman (2010) showed in the ocellated wrasse a positive relationship between the number of sneakers near nests and the proportion of offspring sired by the parental male. However, in this species, it was previously shown that ocellated females preferred nests with high mating success because nesting males at these sites are more likely to provide parental care (Alonzo, 2008; Alonzo and Heckman, 2010).

In summary, in the peacock blenny, the two components of reproductive success of nest-holder males depend on two different types of factors. Whereas mating success depends on male characteristics used in female mate choice, fertilization success is influenced by the characteristics of the social network of the nest-holder male. This result highlights the importance that social factors may have on reproductive success and ultimately individual fitness. Finally, it should be stressed that reproductive success should not be viewed as a static trait, but rather as a dynamic characteristic of each of the tactics. The strength of sexual selection may vary between populations (e.g. due to different environment *regimes*; (Bessert et al., 2007; Ota et al., 2012) and within populations over time (e.g. seasonal changes in the operational sex ratio; (Wacker et al., 2014)), thus affecting both spatially and temporally the frequency and relative reproductive success of ARTs. Therefore, future studies in this species should consider both the temporal and spatial dynamics of sexual selection when estimating reproductive success for each male tactic.

3.6. Acknowledgements

We would like to thank António José dos Santos and João Daniel from the Developmental Psychology Unit at ISPA-Instituto Universitário for their help with the social network analysis; the late Vitor Almada and Joana Robalo for discussions on paternity estimate approaches in this species; and the editor, Peter K. McGregor, and the two anonymous reviewers for their comments that helped to improve the manuscript. This study was supported by the research grant EXCL/BIA-ANM/0549/2012 from the Portuguese Foundation for Science and Technology (FCT) and by grant no. 012/2012/A1 from the Macao Science and Technology Development Fund (FDCT). During the writing of this manuscript, S.D.C. was being supported by an FCT grant (SFRH/BD/89072/2012).

3.7. Compliance with ethical standards

Ethical approval All animal protocols were performed in accordance with accepted veterinary practice under a BGroup-1^ licence issued by the BDirecção-Geral de Veterinária, Ministério da Agricultura, do Desenvolvimento Rural e das Pescas^, Portugal.

3.8. Conflict of interest

The authors declare that they have no conflict of interest.

3.9. References

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Almada V.C. & Santos R.S., 1995. Parental care in the rocky intertidal: a case study of adaptation and exaptation in Mediterranean and Atlantic blennies. *Reviews in Fish Biology and Fisheries*, *5*, 23–37.

Alonzo S.H., 2010. Social and coevolutionary feedbacks between mating and parental investment. *Trends in Ecology & Evolution*, *25*, 99–108.

Alonzo S.H., 2008. Female mate choice copying affects sexual selection in wild populations of the ocellated wrasse. *Animal Behaviour*, *75*, 1715–1723.

Alonzo S.H. & Heckman K.L., 2010. The unexpected but understandable dynamics of mating, paternity and paternal care in the ocellated wrasse. *Proceedings of the Royal Society of London B: Biological Sciences*, *277*, 115–122.

Avise J.C., Jones A.G., Walker D. & DeWoody J.A., 2002. Genetic mating systems and

reproductive natural histories of fishes: lessons for ecology and evolution. *Annual Review of Genetics*, *36*, 19–45.

Barata E.N., Serrano R.M., Miranda A., Nogueira R., Hubbard P.C. & Canário A.V.M., 2008. Putative pheromones from the anal glands of male blennies attract females and enhance male reproductive success. *Animal Behaviour*, *75*, 379–389.

Bessert M.L., Brozek J. & Ortí G., 2007. Impact of nest substrate limitations on patterns of illegitimacy in the fathead minnow, Pimephales promelas (Cypriniformes: Cyprinidae). *Journal of Heredity*, *98*, 716–722.

Borgatti S.P., 2002. NetDraw: Graph visualization software. *Analytic Technologies, Harvard*, .

Borgatti S.P., Everett M.G. & Freeman L.C., 2002. Ucinet 6 for Windows: Software for Social Network Analysis. *Analytic Technologies, Harvard*, .

Brockmann H.J., 2001. The evolution of alternative strategies and tactics. *Advances in the Study of Behavior*, *30*, 1–51.

Cardoso S.D., Gonçalves D., Robalo J.I., Almada V.C., Canário A.V.M. & Oliveira R.F., 2013. Efficient isolation of polymorphic microsatellites from high-throughput sequence data based on number of repeats. *Marine Genomics*, *11*, 11–16.

Cogliati K.M., Neff B.D. & Balshine S., 2013. High degree of paternity loss in a species with alternative reproductive tactics. *Behavioral Ecology and Sociobiology*, *67*, 399–408.

Coleman S.W. & Jones A.G., 2011. Patterns of multiple paternity and maternity in fishes. *Biological Journal of the Linnean Society*, *103*, 735–760.

Fagundes T., Gonçalves D.M. & Oliveira R.F., 2007. Female mate choice and mate search tactics in a sex role reversed population of the peacock blenny Salaria pavo (Risso, 1810). *Journal of Fish Biology*, *71*, 77–89.

Fagundes T., Simões M.G., Saraiva J.L., Ros A.F.H., Gonçalves D. & Oliveira R.F., 2015. Birth date predicts alternative life history pathways in a fish with sequential reproductive tactics. *Functional Ecology*, *29*, 1533–1542.

Fishelson L., 1963. Observations on littoral fishes of Israel. I. Behaviour of Blennius pavo Risso (Teleostei: Blenniidae). *Israel Journal of Zoology*, *12*, 67–80.

Giacomello E., Marchini D. & Rasotto M.B., 2006. A male sexually dimorphic trait provides antimicrobials to eggs in blenny fish. *Biology Letters*, *2*, 330–333.

Girvan M. & Newman M.E.J., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 7821–7826.

Gonçalves D., Fagundes T. & Oliveira R., 2003. Reproductive behaviour of sneaker males of the peacock blenny. *Journal of Fish Biology*, *63*, 528–532.

Gonçalves D., Matos R., Fagundes T. & Oliveira R., 2005. Bourgeois males of the peacock blenny, Salaria pavo, discriminate female mimics from females? *Ethology*, *111*, 559–572.

Gonçalves D.M., Barata E.N., Oliveira R.F. & Canário A.V.M., 2002a. The role of male visual and chemical cues on the activation of female courtship behaviour in the sex-role reversed peacock blenny. *Journal of Fish Biology*, *61*, 96–105.

Gonçalves D.M. & Oliveira R.F., 2003. Time spent close to a sexual partner as a measure of female mate preference in a sex-role-reversed population of the blenny Salaria pavo (Risso) (Pisces: Blenniidae). *acta ethologica*, *6*, 1–5.

Gonçalves D.M., Simoes P.C., Chumbinho A.C., Correia M.J., Fagundes T. & Oliveira R.F., 2002b. Fluctuating asymmetries and reproductive success in the peacock blenny. *Journal of Fish Biology*, *60*, 810–820.

Gonçalves E.J., Almada V.C., Oliveira R.F. & Santos A.J., 1996. Female mimicry as a mating tactic in males of the blenniid fish Salaria pavo. *Journal of the Marine Biological Association of the United Kingdom*, *76*, 529–538.

Gross M.R., 1996. Alternative reproductive strategies and tactics diversity within sexes. *Trends in Ecology & Evolution*, *11*, 92–98.

Guillemaud T., Almada F., Santos R.S. & Cancela M.L., 2000. Interspecific utility of microsatellites in fish: a case study of (CT)n and (GT)n markers in the shanny Lipophrys pholis (Pisces: Blenniidae) and their use in other Blennioidei. *Marine Biotechnology*, *2*, 248–253.

Kalinowski S.T., Taper M.L. & Marshall T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*, 1099–1106.

Kokko H. & Jennions M.D., 2008. Parental investment, sexual selection and sex ratios. *Journal of Evolutionary Biology*, *21*, 919–948.

Mackiewicz M., Porter B. a., Dakin E.E. & Avise J.C., 2005. Cuckoldry rates in the Molly Miller (scartella cristata; blenniidae), a hole-nesting marine fish with alternative reproductive tactics. *Marine Biology*, *148*, 213–221.

Makagon M.M., McCowan B. & Mench J.A., 2012. How can social network analysis contribute to social behavior research in applied ethology? *Applied Animal Behaviour Science*, *138*, 152–161.

Moore M.C., 1991. Application of organization-cctivation theory to alternative male reproductive strategies : a review. *Hormones and Behavior*, *25*, 154–179.

Neff B.D., 2001. Genetic paternity analysis and breeding success in bluegill sunfish (Lepomis macrochirus). *Journal of Heredity*, *92*, 111–119.

Neff B.D., Repka J. & Gross M.R., 2000a. Parentage analysis with incomplete sampling of candidate parents and offspring. *Molecular Ecology*, *9*, 515–528.

Neff B.D., Repka J. & Gross M.R., 2000b. Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? *Molecular Ecology*, *9*, 529–539.

Oliveira R.F., 2006. Neuroendocrine mechanisms of alternative reproductive tactics in fish, in: Sloman, K.A., Wilson, R.W., Balshine, S. (Eds.), Fish Physiology:

Behaviour and Physiology of Fish. Elsevier, New York, pp. 297–357.

Oliveira R.F., Almada V.C., Forsgren E. & Gonçalves E.J., 1999. Temporal variation in male traits, nesting aggregations and mating success in the peacock blenny. *Journal of Fish Biology*, *54*, 499–512.

Ota K., Awata S., Morita M., Yokoyama R. & Kohda M., 2014. Territorial males can sire more offspring in nests with smaller doors in the cichlid Lamprologus lemairii. *Journal of Heredity*, *105*, 416–422.

Ota K., Hori M. & Kohda M., 2012. Changes in reproductive life-history strategies in response to nest density in a shell-brooding cichlid, Telmatochromis vittatus. *Die Naturwissenschaften*, *99*, 23–31.

Patzner R.A., 1984. Individual tagging of small fish. *Aquaculture*, *40*, 251–253.

Patzner R.A., Seiwald M., Adlgasser M. & Kaurin G., 1986. The reproduction of Blennius pavo (Teleostei, Blenniidae) V. Reproductive behavior in natural environment. *Zoologischer Anzeiger*, *216*, 338–350.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rice W.R., 1989. Analyzing tables of statistical tests. *Evolution*, *43*, 223–225.

Rousset F., 2008. GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106.

Saraiva J.L., Barata E.N., Canário A.V.M. & Oliveira R.F., 2009. The effect of nest aggregation on the reproductive behaviour of the peacock blenny Salaria pavo. *Journal of Fish Biology*, *74*, 754–762.

Saraiva J.L., Gonçalves D. & Oliveira R.F., 2013. Ecological modulation of reproductive behaviour in the peacock blenny: a mini-review. *Fish Physiology and Biochemistry*, *39*, 85–89.

Taborsky M., 1997. Bourgeois and parasitic tactics: do we need collective, functional terms for alternative reproductive behaviours? *Behavioral Ecology and Sociobiology*, *41*, 361–362.

Taborsky M., 1994. Sneakers, Satellites, and Helpers: Parasitic and Cooperative Behavior in Fish Reproduction. *Advances in the Study of Behavior*, *23*, 1–100.

Taborsky M., Oliveira R.F. & Brockmann H.J., 2008. The evolution of alternative reproductive tactics: concepts and questions, in: Oliveira, R.F., Taborsky, M., Brockmann, H.J. (Eds.), Alternative Reproductive Tactics: An Integrative Approach. Cambridge University Press, Cambridge, pp. 1–21.

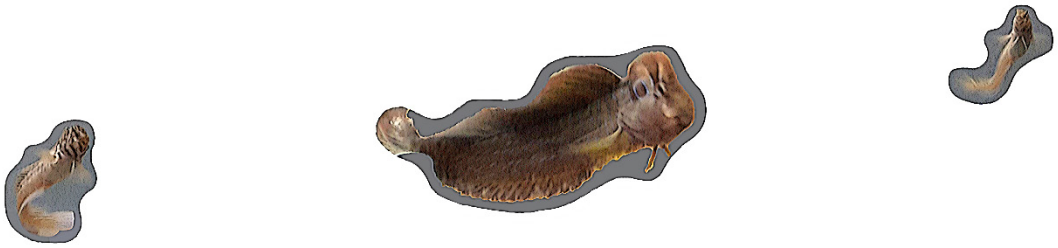
Trivers R.L., 1972. Parental Investment and Sexual Selection, in: Campbell, B.G. (Ed.), Sexual Selection and the Descent of Man. Aldine Press, Chicago, pp. 136–179.

Wacker S., Amundsen T., Forsgren E. & Mobley K.B., 2014. Within-season variation in sexual selection in a fish with dynamic sex roles. *Molecular Ecology*, *23*, 3587–3599.

Westernhagen H. v., 1983. Observations on the reproductive and larval biology of Blennius pavo (Pisces: Teleostei). *Helgoländer Meeresuntersuchungen*, *36*, 323–335.

Westneat D.F. & Sherman P.W., 1993. Parentage and the evolution of parental behavior. *Behavioral Ecology*, *4*, 66–77.

Zander C.D., 1986. Blenniidae, in: Whitehead, P.J.P., Bauchot, M.L., Hureau, J.C., Nielsen, J., Tortonese, E. (Eds.), Fishes of the North-Eastern Atlantic and the Mediterranean. UNESCO, Paris, pp. 1096–1112.



CHAPTER 4 |

NEUROGENOMIC PATTERNS OF ARTs

Cardoso S.D., Gonçalves D., Goesmann A., Canário A.V.M. & Oliveira R.F., 2018. Temporal variation in brain transcriptome is associated with the expression of female mimicry as a sequential male alternative reproductive tactic in fish.
Molecular Ecology, *27*, 789–803.
doi: 10.1111/mec.14408

Temporal variation in brain transcriptome is associated with the expression of female mimicry as a sequential male alternative reproductive tactic in fish

Sara D. Cardoso [1,2,3], David Gonçalves [4], Alexander Goesmann [5#], Adelino V.M. Canário [6], Rui F. Oliveira [1,2,3]

[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal; [2] ISPA – Instituto Universitário, Lisbon, Portugal; [3] Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon, Portugal; [4] Institute of Science and Environment, University of Saint Joseph, Macau, China; [5] Center for Biotechnology, CeBiTec, Bielefeld University, Bielefeld, Germany; [6] CCMAR – Centro de Ciências do Mar, Universidade do Algarve, Faro, Portugal; [#] Present address: Bioinformatics and Systems Biology, Justus-Liebig-University, Giessen, Germany.

4.1. Abstract

Distinct patterns of gene expression often underlie intra- and inter-sexual differences, and the study of this set of co-regulated genes is essential to understand the emergence of complex behavioural phenotypes. Here, we describe the development of a *de novo* transcriptome and brain gene expression profiles of wild-caught peacock blenny, *Salaria pavo*, an intertidal fish with sex-role reversal in courtship behaviour (i.e. females are the courting sex) and sequential alternative reproductive tactics in males (i.e. larger and older nest-holder males and smaller and younger sneaker males occur). Sneakers mimic both female's courtship behaviour and nuptial colouration to get access to nests and sneak fertilizations, and later in life transition into nest-holder males. Thus, this species offers the unique opportunity to study how the regulation of gene expression can contribute to intersex phenotypes and to the sequential expression of male and female behavioural phenotypes by the same individual. We found that at the whole brain level, expression of the sneaker tactic was paralleled by broader and divergent gene expression when compared to either females or nest-holder males, which were more similar between themselves. When looking at sex-biased transcripts, sneaker males are intersex rather than being either nest-holder or female-like, and their transcriptome is simultaneously demasculinized for nest-holder-biased transcripts and feminized for female-biased transcripts. These results indicate that evolutionary changes in reproductive plasticity can be achieved through regulation of gene expression, and in particular by varying the magnitude of expression of sex-biased genes, throughout the lifetime of the same individual.

Keywords: alternative reproductive tactics; behavioural plasticity; neurogenomic state; RNA-seq; *Salaria pavo*; sex-biased gene expression

4.2. Introduction

Sexual dimorphisms in morphological and behavioural traits are ubiquitous in sexually reproducing organisms, and in most studied cases are paralleled by differential gene expression between the sexes for loci that are present in both males and females (Ellegren and Parsch, 2007; Mank, 2017, 2009). Thus, it has been hypothesized that sex biases in gene expression may play a central role in resolving sexual conflict and that sex-specific transcription regulatory mechanisms may be key in sexual selection (Pointer et al., 2013). This rationale can be extended to species with alternative reproductive tactics (ARTs) where multiple phenotypes may occur within one of the sexes, typically males, with varying degrees of sexual dimorphism. In ARTs at least two sexually mature (i.e. with functional testes) male morphs occur: (1) bourgeois males that fully express male secondary sex characters and compete for the monopolization of access to mates (e.g. by defending breeding territories), expressing male-typical behaviour (e.g. aggression, courtship); (2) parasitic males (aka sneakers) that do not express male secondary sex characters and behaviour, and exploit the investment of bourgeois males to get access to mates (e.g. sneaking fertilizations) (Taborsky et al., 2008). Hence, species with ARTs offer the possibility to test the hypothesis that differential expression of sex-biased genes is also a major genetic mechanism for the evolution of intra-sexual variation in reproductive tactics by providing inter-individual variation in the expression of behavioural and morphological sex-related traits. Moreover, in some species, female-mimicking sneaker males occur. These sneaker males are morphologically and behaviourally similar to females but have functional male gonads, and use their female resemblance to approach mating pairs in order to try to get parasitic fertilizations (Taborsky, 1994; Taborsky et al., 2008). Thus, these males have male and female traits expressed in the same phenotype. This raises the question of how regional differential regulation of gene expression may accommodate different sex phenotypes across different tissues and the role it may play in the evolution of ARTs. Interestingly in some cases, female-mimicking ARTs are expressed by the same individual at different life-history stages, which raises the additional hypothesis that temporal differential regulation

of gene expression may allow for the evolution of sequential ARTs within the same individual.

In recent years, the possibility of analysing gene expression profiles without a reference genome (e.g. RNA-seq; (Wang et al., 2009)) allowed us to test the abovementioned hypotheses in non-model organisms (Zuk and Balenger, 2014). For example, in two horned beetle species with ARTs (*Onthophagus taurus* and *O. nigriventris*) gene expression in different tissues was as divergent between alternative male morphs as it was between sexes (Snell-Rood et al., 2011). However, in the bulb mite, *Rhizoglyphus robini*, the number of sex-biased genes in whole-body expression analysis was much higher than the number of genes differentially expressed between male morphs (Stuglik et al., 2014). Within teleosts, the variability of reproductive systems is vast, and most studies have focused on the characterization of whole-brain gene expression patterns associated with discrete reproductive phenotypes. In all studies, gene expression differences between members of the same species that differ in plasticity of mating tactics reflect how they differ in behaviour and life-history (Aubin-Horth et al., 2005; Fraser et al., 2014), and these differences are more pronounced among male tactics than between sexes (Nugent et al., 2016; Partridge et al., 2016; Schunter et al., 2014; Stiver et al., 2015).

In the present study, we investigate the role of gene expression in the expression of female-mimicking male ARTs in the peacock blenny *Salaria pavo* (Blenniidae; Fig. 1) a littoral fish where female-mimicking males occur (Gonçalves et al., 1996). In this species, nest-holder (bourgeois) males are larger than females, have well-developed secondary sexual characters (SSCs; *viz.* a head crest and a sex-pheromone producing anal gland), which are used to attract females to their nests for spawning (Barata et al., 2008; Gonçalves et al., 2002), and provide sole parental care to eggs until hatching (Fishelson, 1963; Patzner et al., 1986). In a population where nest sites are scarce, and the operational sex ratio is female-biased, females become the courting sex (Almada et al., 1994; Saraiva et al., 2009). Female courtship consists of the expression of a transient nuptial colouration and in stereotypic movements directed towards nest-holder males (Almada et al., 1994). In this population, an alternative male mating tactic is also present, consisting of smaller and younger males behaving as female-mimics to get access to nests guarded by nest-holder males and sneak fertilizations (Gonçalves et al., 1996). Sneaker males switch to nest-holders at a later age (Fagundes et al., 2015), hence the same male expresses both male and female reproductive behaviour during his lifetime. Here, we specifically tested:

(1) if differences in sexual dimorphism in reproductive behaviour between sneaker and nest-holder males are paralleled by gene expression differences in the brain; and (2) if the sneakers' brain transcriptome is closer to that of nest-holder males, since both are sexually mature males and hence are expected to have masculinized brains, or to that of females, since both express female courtship and nuptial colouration. We have also studied males that are in transition from the sneaker to the nest-holder male phenotype. These transitional males are no longer reproductively active as sneakers (i.e. they no longer express female-mimicking behaviour typical of sneakers), but they also have not become nest-holders yet, since they lack the male secondary sex characters and do not defend a nest. Moreover, transitional males possess reduced testes which may result from their investment in somatic growth to become nest-holders, since nest-holders reproductive success depends on body size (Oliveira et al., 1999). Thus, transitional males offer the opportunity to have an extra sampling point in the intrasex dimorphism continuum. Since this is the first study reporting whole transcriptome sequencing for this species, we have also provided a brief characterization of the peacock blenny's transcriptome.

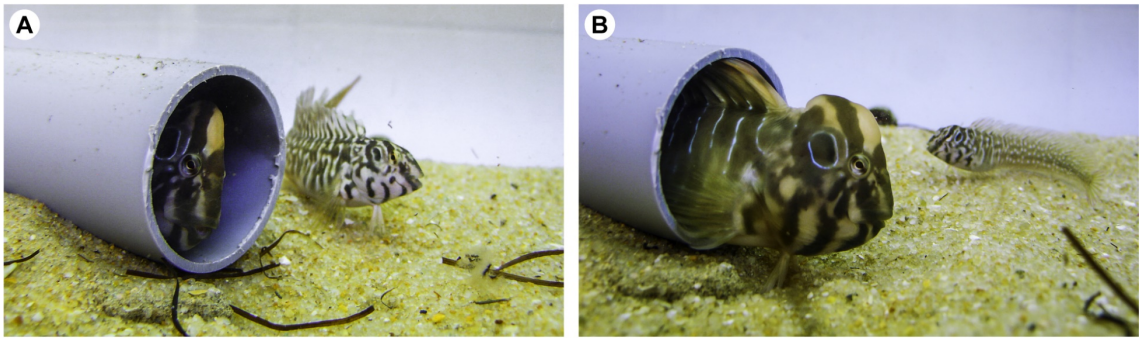


Fig. 1 – The peacock blenny (*Salaria pavo*) has two reproductive male morphs and sex-role reversal in courtship behaviour. Larger and older nest-holder males (inside the nest) with well-developed secondary sexual characters (e.g. head crest) are courted by (A) females and (B) smaller and younger sneaker males behaving as female-mimics. Both females and sneaker males while courting the nest-holder male display a conspicuous nuptial colouration that consists on a pattern of dark and light band across the anterior portion of the body.

4.3. Materials and methods

4.3.1. Ethics statement

All procedures were performed in accordance with accepted veterinary practice under a "Group-1" license issued by the Direcção-Geral de Veterinária, Ministério da

Agricultura, do Desenvolvimento Rural e das Pescas, Portugal (permit number 0421/000/000/2013).

4.3.2. Sample collection

Individuals were collected during the breeding season at Culatra Island (Ria Formosa Natural Park, 36°59′N, 7°51′W, Portugal; for a detailed description of the sampling area see (Almada et al., 1994)). During the breeding season, both females and sneaker males actively express courtship behaviour towards the nest-holder males, while nest-holder males defend nests and provide sole parental care to eggs (Almada et al., 1995; Gonçalves et al., 1996). The peacock blenny is an intertidal species, and hence, individuals were captured with hand nets during low tide; nest-holder males were collected from their nests, whereas females, sneaker males, and transitional males were collected from surrounding areas, where they seek shelter in small pools under tiles and other debris. As sneaker males are female mimics, we used sperm release upon gentle abdominal pressure as an identification criterion. Transitional males, which represent the shift from sneaker to nest-holder males, were identified by having poorly developed secondary sexual characters (SSCs) and a larger body size than sneakers, while not releasing sperm upon abdominal pressure. Fish were euthanized using a lethal dosage of the anaesthetic MS-222 (tricaine methanesulfonate; Sigma-Aldrich) followed by rapid severance of the spinal cord with a scalpel. Brain dissection occurred within 5 minutes after capture, and brains were immediately immersed in RNAlater (Sigma-Aldrich), kept at 4 ºC for 24 hr, and stored at –80 ºC until further processing. Fish were then dissected to assess gonadal sex by direct inspection of the gonads.

Thirty-seven individuals (nine females, nine nest-holder males, nine sneaker males and ten transitional males) were sampled for gene expression analysis, establishing two replicates of pools of 4 to 5 individuals for each phenotype. Among males, sneakers had the largest gonads relative to their body size (gonadosomatic index (GSI)), followed by nest-holders and then by transitional males (Welch's one-way ANOVA, $F_{2,11.0} = 47.708$, $p < 0.001$). This variation in GSI followed previously reported patterns (Gonçalves et al., 2008, 1996) confirming the reproductive state of examined male individuals (i.e. only nest-holder and sneaker males are sexually mature). All females used in this study were expected to be actively courting, as their GSI values indicate they were sexually mature (mean ± SEM = 10.65 ± 3.54; see (Patzner, 1983) for a description of the ovarian cycle

of this species), and in this species, all sexually mature females are known to exhibit courtship behaviour (Almada et al., 1995).

4.3.3. RNA extraction and sequencing

Total RNA was extracted from whole brain tissue for each individual of the four phenotypes. Samples were transferred to 1 ml of QIAzol lysis reagent (Qiagen) and the tissue was homogenized (on ice) with a sonicator. RNA was then extracted using the RNeasy Lipid Tissue Mini Kit (Qiagen) following the manufacturer's protocol. RNA quality and concentration for each sample were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific). Equal amounts of RNA from each individual were used to prepare the RNA pools corresponding to the replicated libraries for each phenotype (two libraries per phenotype).

Preparation and sequencing of the eight pooled RNA libraries were performed by The Genome Analysis Centre (TGAC, Norwich, United Kingdom). Upon arrival, samples' RNA integrity was verified on the Agilent 2100 Bioanalyzer. Libraries were prepared from polyA+ RNA and sequenced in two lanes on the Illumina HiSeq 2000 as paired-end 100 bp reads with an approximate insert size of 341 bp. Each library was sequenced at a high depth (Table S1, Appendix IV) to detect transcripts that can become underrepresented when analysing whole brain tissue expression. To improve the *de novo* assembly, four additional pooled libraries, one for each phenotype, were also prepared using RNA from the previous collection, and sequenced using Roche's 454 technology. Library preparation and single-end sequencing on a GS-FLX System were performed at Max Planck Institute (Berlin, Germany) using the manufacturer's protocol.

4.3.4. *De novo* assembly

Quality filtering and trimming of Illumina reads were carried using implemented plugins in Conveyor software (Linke et al., 2011), keeping only pairs of reads that had an overall Phred quality score superior to 20 and a minimum read length of 50 bp (Table S1, Appendix IV). In addition to the 454 data, a multi-tissue dataset previously sequenced and used for microsatellite mining (Cardoso et al., 2013) was also included in the analysis. All 454 reads were quality trimmed to have an overall Phred quality score of 15 and a minimum read length of 100 bp. Read quality was assessed using FastQC v0.7.2 (Andrews, 2010).

The *de novo* transcriptome assembly was carried out in two phases (Fig. S1, Appendix IV). First, all 454 reads, which represent the long reads in this work, were assembled with Newbler (454 Sequencing System Software v2.6; Roche) using the *-urt* option to recover more complete representations of transcripts as well as low coverage transcripts. The resultant 63,521 Newbler isotigs (N50 = 464 bp) were then given as input together with the Illumina reads to Trinity assembler v20110820 (Grabherr et al., 2011) with default parameters. After redundancy and exogenous contig removal (see Supporting information for more details), 577,532 putative transcripts with an N50 of 1,165 bp (Table S2, Appendix IV) were functionally annotated against UniProtKB (taxonomic division vertebrate databases, release April 2014) using BLASTX with an E-value cut-off of 10^{-5}, allowing a maximum of 10 hits per contig.

4.3.5. Differential expression analysis

For differential expression (DE) analysis of transcripts, pairs of reads from each of the eight sequenced libraries (2 pools from each male morph and female) were mapped separately against the filtered Trinity assembly using Bowtie v1.0.0 (Langmead et al., 2009) within RSEM v1.2.8 (Li and Dewey, 2011) with default settings. On average 77.8% of the pairs of reads from each library mapped back onto the reference assembly (Table S2, Appendix IV). Due to the high read depth of our dataset and the use of pooled individuals for sequencing, we expected that some of the 577,532 contigs present in the assembly would represent assembly mistakes and rare variants (e.g. genetic polymorphisms, introns, chimeras, sequencing errors). To remove this low-level expression noise from the analysis we employed a filter similar to the one used by Harrison and colleagues (2012), and discarded all contigs that did not have at least two fragments per million mappable reads (FPKM = 2, corresponding to 179,202 contigs retained; Fig. S2 and Table S2, see Appendix IV for more details).

Pairwise comparisons were made using the two libraries per phenotype as biological replicates in the R package DESeq v1.14.0 (Anders and Huber, 2010) by applying the default parameters. DESeq has a more data-driven parameter estimation, allowing a better performance in controlling the false-positive rate while also maintaining the power to detect differentially expressed transcripts when analysing experimental designs with a low number of replicates per condition (Soneson and Delorenzi, 2013; Zhang et al., 2014). Transcripts were considered to be differentially expressed with a fold change > 2 and a *P*

< 0.05 after Benjamini-Hochberg adjustment for multiple testing (Benjamini and Hochberg, 1995). For visualization of the global expression patterns of differentially expressed transcripts among phenotypes, variance stabilized transformed expression values were retrieved and Trinity scripts for heatmap generation using the hclust function in R adapted to produce a hierarchical clustering of Z-transformed expression values using Euclidean distance with complete linkage. The reliability of the inferred tree was assessed by 1,000 bootstrap resampling of the expression values using the R package Pvclust v1.2.2 (Suzuki and Shimodaira, 2006). A principal component analysis was also performed using the R package plot3D v1.1. Differences in transcript regulation bias towards up- or down-regulation among phenotypes were analysed using a two-way contingency table simulation statistics (ACTUS2, (Estabrook et al., 2002)), based on 1,000 simulations. Other statistical analyses and graphing presented in this work were obtained using R (R Core Team, 2014).

4.3.6. Transcriptional coexpression network analysis

To investigate the network of co-expressed genes (i.e. gene modules) across phenotypes a weighted coexpression network analysis (WGCNA) was conducted (R package WGCNA v1.51; (Langfelder and Horvath, 2008)). This analysis clusters together genes with highly correlated expression values across all samples into modules (Langfelder and Horvath, 2008). Variance stabilized transformed expression values from all transcripts that underwent differential expression analysis in DESeq were used. A block-wise weighted signed network approach was followed, using a maximum of 36,000 transcripts in each block, due to constraints in R memory allocation when working with large datasets. In brief, this method pre-clusters nodes into large blocks in an unsupervised way, where hierarchical clustering is applied independently defining the dendrogram of gene modules (Langfelder and Horvath, 2008). Afterwards, the module detection results are combined across blocks by merging modules whose eigengenes (i.e. module-weighted average expression profiles) are highly correlated (Langfelder and Horvath, 2008). The soft-thresholding power (β) was set to 17, corresponding to an R-squared of > 0.80 (Zhang and Horvath, 2005) (mean connectivity of 3,406.6; Fig. S3, Appendix IV), with a minimum module size of 30 transcripts. Modules were merged when eigengene dissimilarity between modules was < 0.20. Default settings were used for all other WGCNA parameters. Statistical significance of module correlation with each phenotype

was determined when $P < 0.05$ after a Benjamini-Hochberg adjustment for multiple testing. These modules were subsequently validated *in silico* by assessing the relationship between transcript significance for each phenotype and module membership (i.e. correlation of the module eigengene with the transcript expression profile).

4.3.7. Functional Annotation

Sequences and BLAST results for the selected level of differential expression analysis (FPKM = 2) were imported into Blast2GO v2.7 (Conesa et al., 2005), and mapping of Gene Ontology (GO) terms (Ashburner et al., 2000) proceeded against a local database (release April 2014). For the detected differential expressed contigs, samples' mapping results were manually checked in IGV v2.3.37 (Robinson et al., 2011) and annotations were further refined based on the characteristics of contig-hit alignments by cross-checking BLASTX results from UniProtKB and NCBI's nr database (E-value cut-off of 10^{-5}, release April 2014). After this step of manual curation, detected chimeras, contigs whose BLAST hits belonged to different genes that showed no significant overlap, were not considered for further analyses (N = 18). GO terms were annotated using an annotation cut-off of 65 and a GO weight of 10, which resulted in 111,325 GO terms annotated to 30,367 contigs and exported in tabular form. GO term enrichment for modules and contigs detected as differentially expressed were evaluated by GOstats v2.32 R package (Falcon and Gentleman, 2007) using a 'conditional' hypergeometric test with a $P < 0.01$. This method accounts for the hierarchical relationships of GO terms by removing genes from ancestor terms if they are significantly enriched in a child term. We did not apply any formal correction for multiple testing due to the implicit dependence between neighbouring GO terms, which do not comply with the independence of tests needed for correction of the *P*-values. Semantic similarity among phenotypes for enriched GO terms was computed using the Wang graph-based method available in the R package GOSemSim v1.30.3 (Yu et al., 2010) independently for each GO ontology, and the respective scores used to cluster samples according to their GO similarity using heatmap clustering. The relative contribution of GO enrichment data in terms of GO classes they represent was visualized using the GO-slim vocabulary and the web tool CateGOrizer (Hu et al., 2008).

4.4. Results

A total of 1.3 million single-end reads with an average size of 375 bp and 210 million paired-end reads with an average size of 93 bp remained after trimming and were used for *de novo* transcriptome assembly. After filtering lowly expressed transcripts, the reference assembly featured 179,202 transcripts with an N50 length of 1,646 bp (Table S2, Appendix IV). The annotation of the peacock blenny transcriptome assembly against the vertebrate division of UniProtKB revealed that only 45,994 contigs (*ca.* 25.7%) had a BLAST hit and that more than half of these contigs had its top hit from one of three species of fish, *Oreochromis niloticus* (N = 18,139), *Xiphophorus maculatus* (N = 5,190) and *Gasterosteus aculeatus* (N = 4,423).

4.4.1. Differential transcript expression across phenotypes

Six pairwise comparisons were performed using DESeq's negative binomial model with a false discovery cut-off of 5%, resulting in a total of 814 transcripts (corresponding to 0.45% of the transcriptome and 704 Trinity predicted genes) detected as being differentially expressed among the four phenotypes. Normalized expression profiles represented by hierarchical clustering were obtained considering all phenotypes (Fig. 2A). All biological replicates were clustered according to their respective phenotype. The profile of DE transcripts was more similar between nest-holder males and females than with transitional males or with sneaker males. Sneakers formed their own cluster apart from all other phenotypes. All clusters had high confidence, as illustrated by the bootstrap probability values (Fig. 2A). Principal components analysis of DE transcripts also showed a clear separation among male morphs, with the first three components accounting for 85.2% of the observed variance in transcript expression (Fig. 2B; Table S3, Appendix IV).

Regarding the pairwise comparisons (Table 1; Tables S4–S9, Appendix IV), the comparison between nest-holder males and females had the lowest number of differentially expressed transcripts (N = 155), whereas the comparison between sneakers and transitional males had the highest number of differentially expressed transcripts (N = 348). Overall, sneaker males differed from the other phenotypes in the expression of 644 transcripts (78.38% annotated), followed by transitional males (600 transcripts; 75.71% annotated), females (564 transcripts; 76.51% annotated) and finally nest-holder males (518 transcripts; 78.92% annotated).

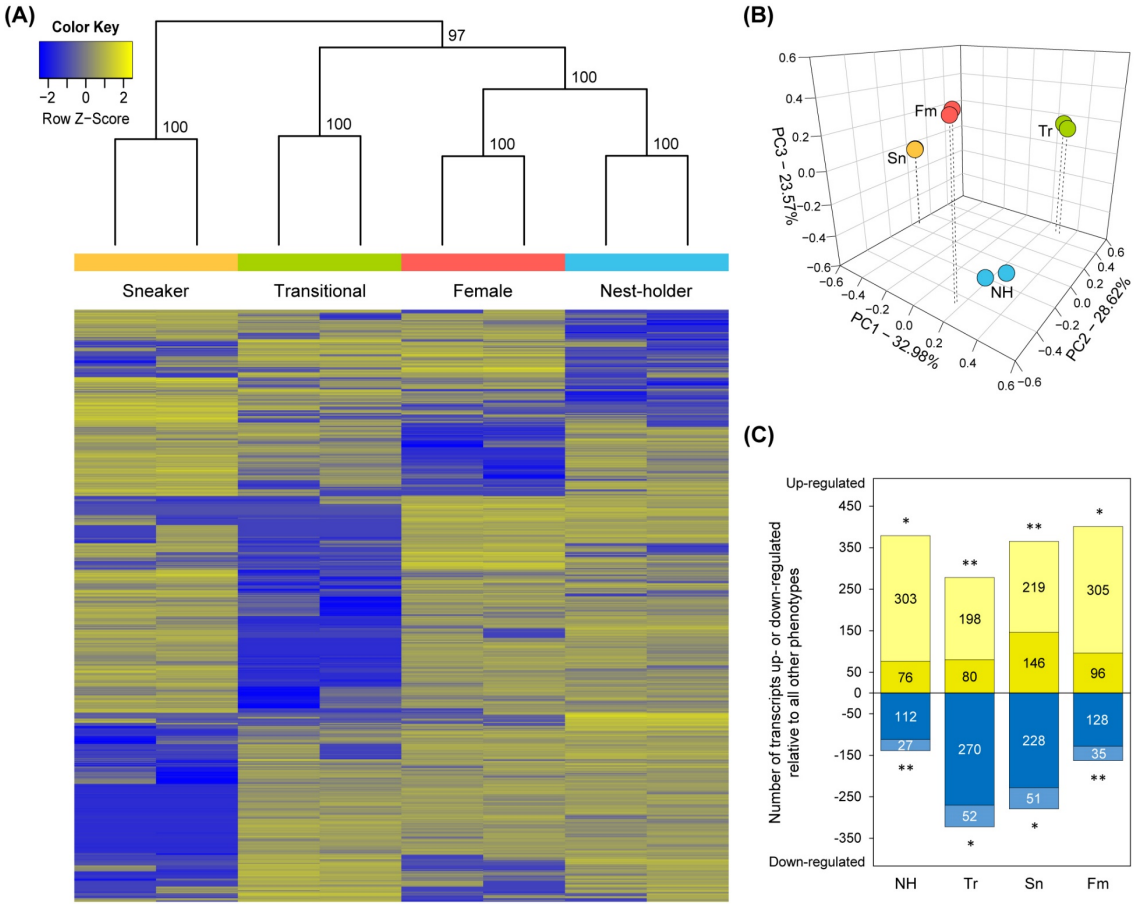


Fig. 2 – Differences in brain expression patterns across *Salaria pavo* phenotypes. (A) Heatmap of all transcripts differentially expressed ($N = 814$) between 2 or more phenotypes, either with or without functional annotation. Intensity of colour indicates relative expression levels of each transcript (rows) in each phenotype sample (columns), with blue representing down-regulated transcripts and yellow up-regulated transcripts. Similarity between phenotypes with hierarchical clustering can be seen above the heatmap with respective bootstrap values. (B) Principle Components Analysis (PCA) shows a clear separation between male morphs across the first three components, which account for 85.2% of the observed variance in transcript expression. (C) Total number of transcripts for which each phenotype showed significantly different expression from all other phenotypes. Differences in the degree of bias in expression for each phenotype toward up- or down-regulation were tested using a two-way contingency table simulation statistics (* cases detected as significantly big, ** cases detected as significantly small based on 1,000 simulated tables in ACTUS2, one tailed P-value < 0.05). Dark yellow and dark blue bars indicate the number of transcripts expressed exclusively by the phenotype. Light yellow and light blue bars indicate the number of transcripts expressed in two or more phenotypes regardless of the pairwise comparison. Each phenotype is colour coded in agreement with blue for nest-holder males (NH), green for transitional males (Tr), orange for sneaker males (Sn), and red for females (Fm).

Table 1 – Number of significantly expressed transcripts and percentage of annotated transcripts in each pairwise comparison between phenotypes of *Salaria pavo*.

	Total	Annotated
Nest-holder > Female	84	76.19%
Female > Nest-holder	71	74.65%
Nest-holder > Sneaker	155	76.77%
Sneaker > Nest-holder	92	84.78%
Nest-holder > Transitional	174	81.03%
Transitional > Nest-holder	59	79.66%
Sneaker > Female	106	75.47%
Female > Sneaker	173	76.88%
Transitional > Female	72	69.44%
Female > Transitional	195	78.97%
Sneaker > Transitional	187	85.56%
Transitional > Sneaker	161	64.60%

> indicates higher expression in the phenotype on the left; FDR adjusted significance value of 0.05.

A bias in the direction of expression of these transcripts was also detected for each phenotype. Looking into the total number of transcripts that were either up- or down-regulated in a phenotype, nest-holder males and females had more transcripts up-regulated and fewer transcripts down-regulated than expected by chance, whereas sneaker and transitional males had the opposite pattern with fewer transcripts up-regulated and more transcripts down-regulated than expected by chance (one-tailed P-value < 0.05; Fig. 2C). These results point to different mechanisms of gene regulation acting across phenotypes. However, when the number of transcripts that were only found to be either up- or down- regulated on a specific phenotype was examined, all phenotypes showed a bias towards down-regulation (Z-tests with $p < 0.05$: nest-holder, $z = -12.69$; transitional, $z = -13.64$; sneaker, $z = -10.63$; female, $z = -12.01$; Fig. 2C), indicating that negative regulation of gene expression was more specific within each phenotype than positive regulation.

4.4.2. Gene ontology (GO) enrichment analysis

GO analysis for up-regulated transcripts within each phenotype revealed distinct classes of enriched terms in male morphs (Fig. 3; Table S10, Appendix IV). Nest-holder males had a predominant enrichment for processes related to lipid metabolism and transmembrane transport of ions at the cell level. Females showed a predominant

enrichment for protein metabolism and, similarly to nest-holder males, transmembrane transport of ions. Sneaker males had a specific enrichment for processes related to morphogenesis, development, cell cycle and DNA metabolism. Finally, transitional males showed an enrichment primarily for regulation of cellular transport, cellular organization and biogenesis processes. Noteworthy, all phenotypes, with the exception of nest-holder males, had an enrichment for terms related to cell, cytoskeleton and organelle organization and biogenesis, whereas sneaker males were the only phenotype with enrichment for extracellular processes (Fig. 3). These enrichment results were further used to obtain measures of GO semantic similarity among phenotypes separately for each ontology (Fig. 3). For both 'Biological Process' and 'Molecular Function' nest-holder males and females had a GO enrichment more similar between themselves than with sneakers and transitional males. In contrast, for 'Cellular Component' enrichment transitional males were closer to nest-holders. Taken together, these results are in accordance with the gene expression level clustering.

Additionally, GO analysis for down-regulated transcripts within each phenotype also presented distinct classes of enriched terms in male morphs (Fig. 4; Table S11, Appendix IV). Transcripts down-regulated in nest-holder males had a predominant GO enrichment for processes related to cell and organelle organization and biogenesis, cell differentiation, and protein metabolism and modification at the cell and intracellular level. Down-regulated transcripts in females showed a predominant enrichment for processes related to transporter activity, metabolism and biosynthesis. Finally, both sneaker and transitional males had transcripts down-regulated with an enrichment for ion transport activity. Additionally, only transitional males presented an enrichment for processes related to lipid metabolism and development. GO semantic similarity analysis (Fig. 4), showed that overall sneaker and transitional males had a GO enrichment more similar between themselves than with females or nest-holder males, with the exception for 'Biological Process' where nest-holder males had the most dissimilar enrichment when compared with the remaining phenotypes.

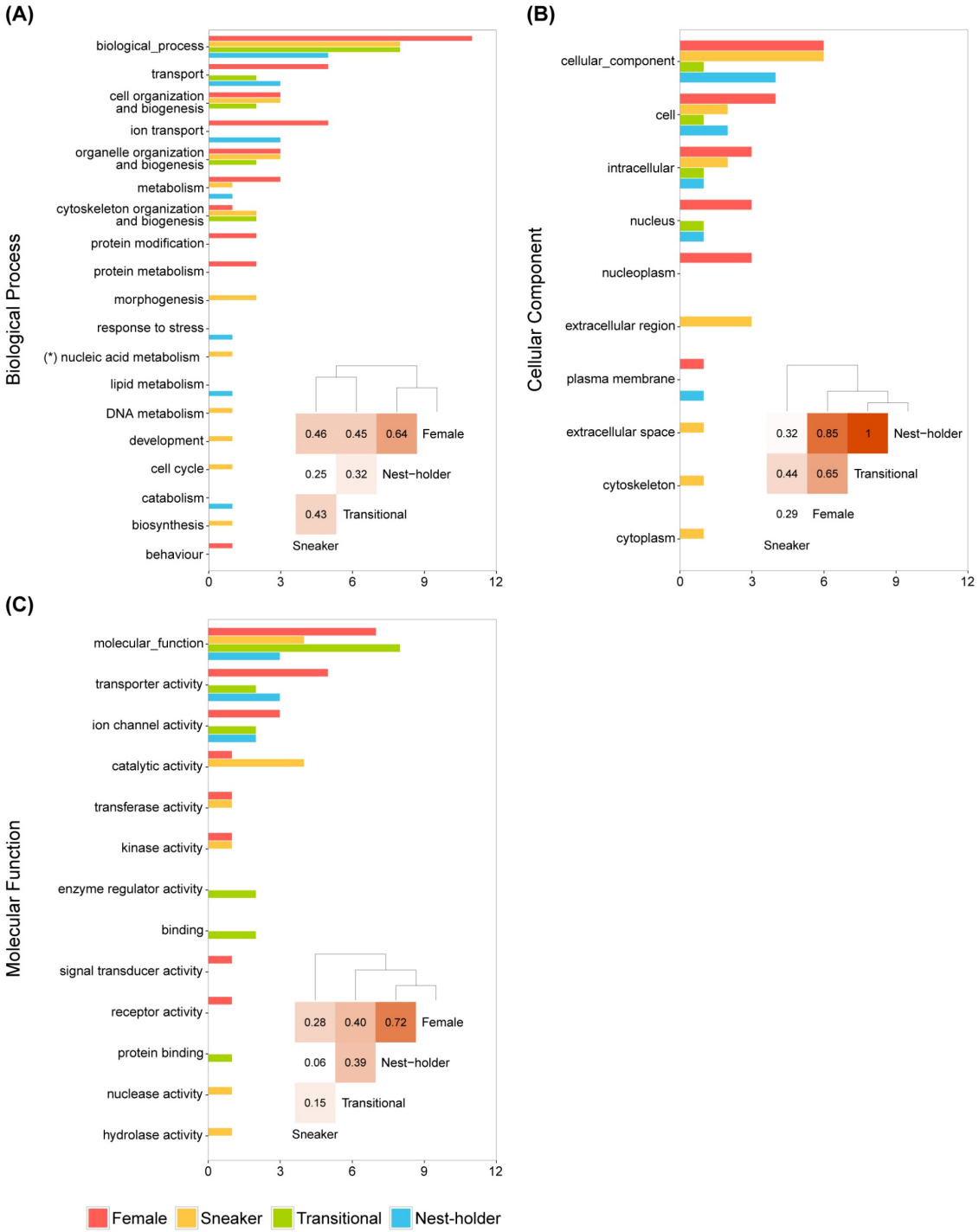


Fig. 3 – Representation of the relative contribution of enriched GO data in terms of GO classes for each ontology, A) Biological Process, B) Cellular Component, and C) Molecular Function. Enriched GO terms were obtained for up-regulated transcripts in each phenotype and mapped to a total of 127 GO slim ancestor terms with CateGOrizer. For each GO class, coloured bars represent the number of enriched terms for each phenotype, with blue for nest-holder males, green for transitional males, orange for sneaker males and red for females. Within each ontology, the heatmap clustering represents the semantic similarity scores for enriched GO terms among phenotypes computed using GOSemSim. Values range between 0 and 1, with higher values indicating greater similarity between groups of GO terms. (*) Complete term for GO class is 'nucleobase, nucleoside, nucleotide and nucleic acid metabolism'.

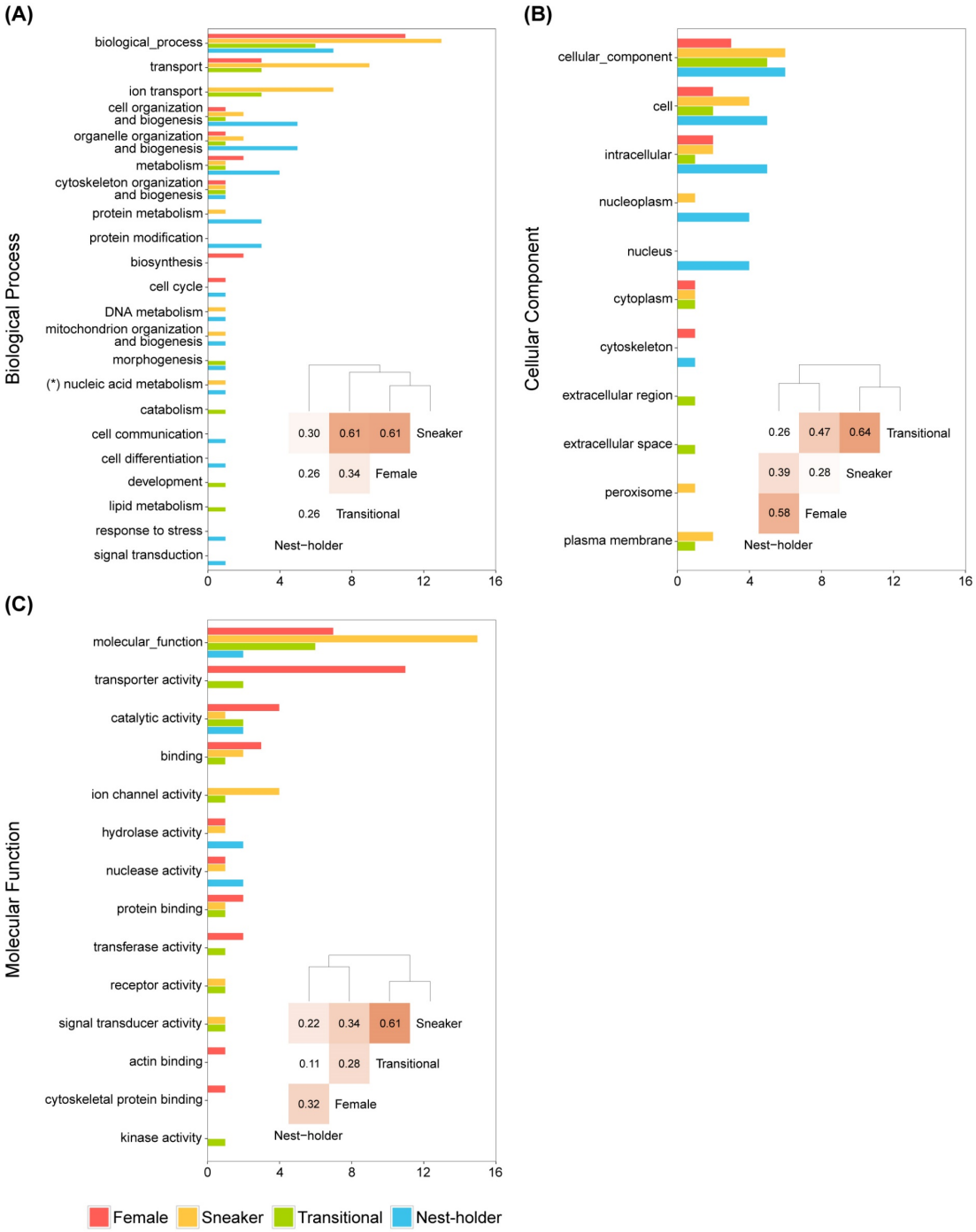


Fig. 4 – Representation of the relative contribution of enriched GO data in terms of GO classes for each ontology, A) Biological Process, B) Cellular Component, and C) Molecular Function. Enriched GO terms were obtained for down-regulated transcripts in each phenotype and mapped to a total of 127 GO slim ancestor terms with CateGOrizer. For each GO class, coloured bars represent the number of enriched terms for each phenotype, with blue for nest-holder males, green for transitional males, orange for sneaker males and red for females. Within each ontology, the heatmap clustering represents the semantic similarity scores for enriched GO terms among phenotypes computed using GOSemSim. Values range between 0 and 1, with higher values indicating greater similarity between groups of GO terms. (*) Complete term for GO class is 'nucleobase, nucleoside, nucleotide and nucleic acid metabolism'.

4.4.3. Demasculinization *versus* feminization of the brain transcriptome of sneakers and transitional males

To further test how the brain transcriptomes of sneakers and transitional males are related to the ones of either nest-holder males or females, we analysed sex-biased transcripts (i.e. transcripts that were up-regulated in either nest-holder males or females; Fig. 5). Both sneaker males and transitional males expressed nest-holder-biased transcripts (N = 84) at significantly lower levels than nest-holders (Fig. 5A), suggesting that the brains of these male morphs are transcriptionally demasculinized. Conversely, both sneaker and transitional males expressed female-biased transcripts (N = 71) at higher levels than nest-holder males (Fig. 5A), suggesting that their brains are transcriptionally feminized. Interestingly, there were no significant differences between sneaker and transitional males in the expression of either nest-holder-or female-biased transcripts (Fig. 5A).

Hierarchical clustering of expression levels showed that both sneaker and transitional males clustered together with high confidence with nest-holder males for nest-holder-biased transcripts, and with females for female-biased transcripts (Fig. 5B). This clustering pattern suggests that the brain transcriptome of these two male morphs can be interpreted as intersex rather than either typically male or female. Hierarchical clustering of expression levels also showed that among the three male morphs, transitional and nest-holder males clustered together for nest-holder-biased genes, suggesting a higher degree of demasculinization of the sneakers' brain transcriptome than of the transitional males. In contrast, transitional and sneaker males clustered together for female-biased genes, suggesting that their brain transcriptomes are equally feminized.

From the 71 female-biased transcripts, 37 were up-regulated in sneaker males of which 25 were also up-regulated in transitional males relative to nest-holder males (Tables S12-S13, Appendix IV). Thus, the 12 female-biased transcripts, which were exclusively up-regulated in females and sneaker males, are potential candidate genes to control the expression of female-like courtship behaviour. These candidate genes had an enrichment in GO terms for biological processes mainly involved in histone modifications and in the regulation of Ral protein signal transduction, which is involved in neural plasticity among other processes (Table S13c, Appendix IV).

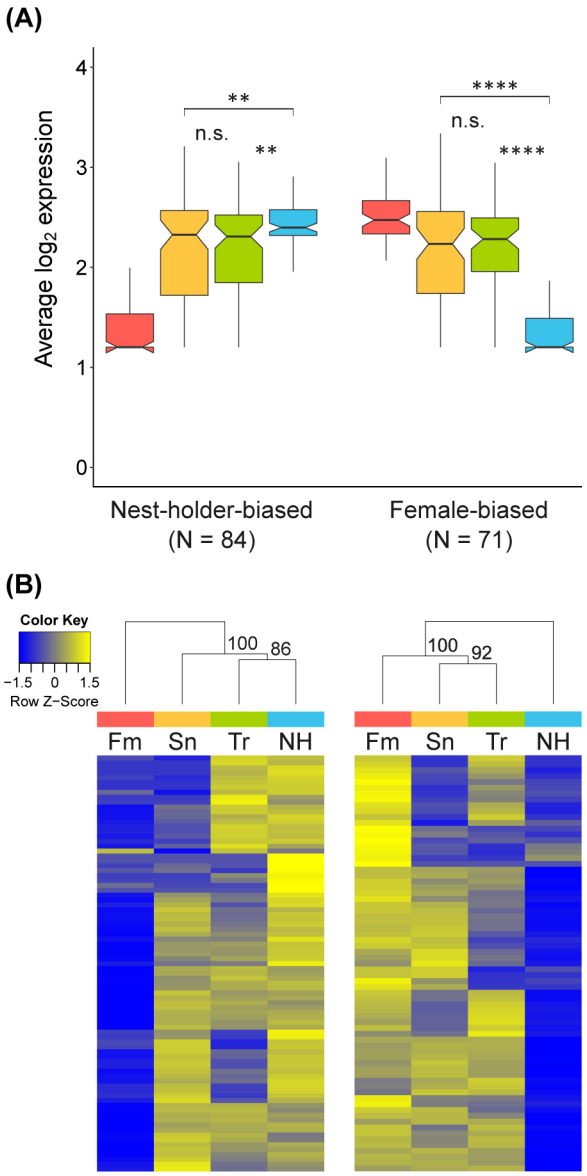


Fig. 5 – Expression patterns for sex-biased transcripts. Nest-holder-biased (N = 84) and female-biased (N = 71) transcripts in females (Fm, coloured in red), sneaker males (Sn, coloured in orange), transitional males (Tr, coloured in green) and nest-holder males (NH, coloured in blue) represented as average $\log_2$ variance-stabilized transformed expression values, A) notched boxplot, excluding outliers, with asterisks indicating significant P-values calculated by the Kruskal-Wallis test, followed by Nemenyi post-hoc test to perform multiple comparisons among male morphs (** $P < 0.01$, **** $P < 0.0001$), and B) hierarchical clustering of the four phenotypes with respective bootstrap values.

From the 84 nest-holder-biased transcripts, 48 were also up-regulated in sneakers of which 25 were also up-regulated in transitional males (Table S14 and S15, Appendix IV). Thus, the 23 transcripts that were exclusively up-regulated in nest-holder and sneaker males, represent candidate genes for the control of male reproduction since both nest-holders and sneakers, but not transitional males, are sexually mature. These transcripts had an enrichment in GO terms mainly related to cell cycle regulation (Table S15b, Appendix IV). From the 84 male nest-holder transcripts, 19 were exclusively up-regulated in nest-holder males (but not in either sneakers or transitional males) (Table S14, Appendix IV). Hence, these transcripts represent potential candidate genes for the control of male reproductive behaviour (i.e. nesting, breeding territoriality). From these

transcripts, only two had GO terms enriched to cell-cell contact and actin cytoskeleton (Table S15d, Appendix IV).

4.4.4. Patterns of gene co-expression modules among phenotypes

The WGCNA showed that the transcriptome could be clustered in 171 modules, with modules ranging in size from 30 to 41,342 transcripts (Fig. S4 and Table S16, Appendix IV). From these, only three modules could be significantly associated with one of the phenotypes after Benjamini-Hochberg correction: the plum3 module (Pearson's $r = 0.99$ with $P = 0.001$) with nest-holder males, and the sienna3 and salmon modules ($r_{sienna3} = 0.99$, $r_{salmon} = -0.96$, with $P_{sienna3} = 0.001$, $P_{salmon} = 0.04$) with sneaker males. All three modules showed a significant positive correlation between transcript significance between the phenotype and module membership, corroborating the importance of these modules for each of the male's phenotypes (Fig. S5, Appendix IV). Plum3 module included 296 transcripts, of which seven transcripts were also found up-regulated for nest-holder males. This module was mainly enriched for GO terms associated with energy production through the oxidative phosphorylation metabolic pathway (Table S17a, Appendix IV). Sienna3 module included 1,031 transcripts, of which 17 were also found as up-regulated in sneaker males. This module represented an enrichment for processes involved in the canonical *Wnt* signalling pathway, cytokinesis, glucose metabolism and carbohydrate transport (Table S17b, Appendix IV). The only module with a negative correlation was the salmon module containing 7,951 transcripts, of which 142 transcripts were also found down-regulated in sneaker males. This module was mainly enriched for processes involved in the regulation of G-protein coupled-receptor signalling pathway (Table S17c, Appendix IV).

4.5. Discussion

Here, we describe the development of a *de novo* transcriptome and brain gene expression profiles of wild-caught peacock blenny, *Salaria pavo*, an intertidal fish with sex-role reversal in courtship behaviour (i.e. females are the courting sex) and presence of alternative reproductive male morphs (Fagundes et al., 2015; Gonçalves et al., 1996).

4.5.1. Brain transcriptomic architecture of alternative reproductive tactics

We found that each peacock blenny phenotype has a distinct transcriptomic profile (Fig. 2A), indicating that distinct behavioural repertoires are associated with distinct neurogenomic states, which differentiate not only sex but also male morphs. These results are consistent with previous studies that have already described this parallel between specific behavioural states and neurogenomic states at different timescales (reviewed in (Cardoso et al., 2015; Harris and Hofmann, 2014; Zayed and Robinson, 2012)), and with previous reports of specific brain transcriptomic profiles among species with ARTs for alternative male morphs, particularly in teleost fishes (Aubin-Horth et al., 2005; Fraser et al., 2014; Nugent et al., 2016; Partridge et al., 2016; Schunter et al., 2014; Stiver et al., 2015). For example, in bluegill sunfish (*Lepomis macrochirus*), the three male ARTs, encompassing two distinct life histories, presented distinct neurogenomic profiles, with sneaker males having the most divergent expression profile, while parental males were further discriminated in the analysis relatively to their reproductive state (i.e. spawning vs. non-spawning; (Partridge et al., 2016)). In sailfin mollies (*Poecilia latipinna*), differentially expressed transcripts were also associated with each ART: in which small males and large males have fixed alternative tactics, always displaying sneaking and courtship behaviour, respectively, whereas intermediate males display courtship or sneaking behaviour depending on the male composition of their social environment (Fraser et al., 2014). Interestingly, in sailfin mollies, the differentially expressed transcripts in the brain associated with the plastic male tactic tended to exhibit larger and more robust genetically regulated changes than socially regulated changes (Fraser et al., 2014). Thus, in both fixed and plastic ARTs, the expression of alternative morphs with specific behavioural phenotypes within the same sex seems to be achieved through differential gene expression in the brain. However, when one compares the available brain RNA-seq transcriptome data for teleost species with ARTs, no clear pattern emerges. Whereas in some species sneaker males exhibit the most distinctive transcriptome (e.g. *L. macrochirus* (Partridge et al., 2016); *Tripterygion delaisi* (Schunter et al., 2014); present study), in other species nest-holder males are the most differentiated phenotype (e.g. *Symphodus ocellatus* (Nugent et al., 2016)). Similarly, the lists of differentially expressed genes for functionally equivalent phenotypes (e.g. sneakers) across species do

not share significant numbers of transcripts, suggesting that ARTS may have evolved in different species through species-specific genetic architectures.

In contrast to the morph-specific profiles of differentially expressed genes, the patterns of gene co-expression modules were only specific for nest-holder and sneaker males. Indeed, the nest-holder phenotype was associated with a gene co-expression module enriched for GO terms associated with energy production, which may reflect the metabolic demand of reproduction in nest-holder males, who have to defend breeding territories for extended periods of time, in which they do not exit the nest for feeding, hence presenting a sharp decline in their body condition over the breeding season (Gonçalves and Almada, 1997). On the other hand, the sneaker phenotype was associated with (i) the up-regulation of a gene co-expression module involved in the canonical *Wnt* signalling pathway and metabolisms, which may reflect their developmental stage; and (ii) with a down-regulation of a gene co-expression module enriched for processes involved in the regulation of G-protein-coupled receptor (GPCR) signalling pathway, which may reflect a mechanisms to make the brain of sneakers less responsive to male gonadal signals (e.g. sex steroids), given that sneakers have mature gonads. In fact, despite the fact that steroid hormones exert their actions by binding to intracellular nuclear receptors, which then bind to DNA to initiate gene transcription (i.e. steroid-inducible transcription factors), steroids can also activate GPCRs in the cell membrane, or even located intracellularly (e.g. G-protein-coupled oestrogen receptor), to initiate rapid non-genomic effects (Gaudet et al., 2015; Losel et al., 2003; Maggiolini and Picard, 2010). The lack of specific gene modules for transitional males and females can be due to the complex dataset that generated a large number of gene co-expression modules. Nonetheless, the analysis revealed the same pattern as the DE analysis among females and male morphs and, for the three significant modules discussed above, a partial overlap was present between transcripts belonging to the morph-specific gene modules and the direction of expression of differentially expressed genes in those morphs.

Overall, these results suggest that the brain transcriptome reflects better reproductive plasticity rather than sex-dimorphism. However, other factors present in the experimental design may have affected the observed patterns as well, namely: age differences between phenotypes; analysis of whole brain gene expression data; use of pooled samples; and the low number of biological replicates. For example, the use of whole brain samples may have masked regional differences in gene expression between phenotypes. Nevertheless, whole brain transcriptome analyses have been very informative as a first approach to

characterize behavioural phenotypes namely in the ART literature (Aubin-Horth et al., 2005; Fraser et al., 2014; Nugent et al., 2016; Partridge et al., 2016; Schunter et al., 2014; Stiver et al., 2015). Similarly, the use of sample pooled replicates offers the advantage of incorporating information from more individuals into DE analysis while still allowing for the estimation of biological variability (Todd et al., 2016), although greater power in the statistical analysis would have been gained by using more biological replicates (Robles et al., 2012). Finally, differences among male morphs may be influenced by differences in age, as sneaker males are younger than nest-holder males and transitional males lay between the two. Hence, in this study it is not possible to disentangle the effect of age from that of reproductive tactic. However, it should be noted that the expression of a specific tactic is one of the most prominent phenotypic characteristics of each age class, and that in other studies, where it was possible to assess the effect of age and age-dependent behavioural tactic to the transcriptome, the latter was the most relevant (e.g. differences in gene expression associated with the transition from nurse to forager in honey bees are largely independent of natural age-related changes (Whitfield et al., 2006)). In summary, despite the possibility of more detailed studies in the future, the present results already reveal ART-specific brain gene expression profiles in the peacock blenny.

4.5.2. Potential role of epigenetic mechanisms in alternative reproductive tactics

Although we do not aim to discuss in detail the list of genes differentially expressed between the different morphs of the peacock blenny, it is important to note the presence of enzymes responsible for post-translational modification of histones (i.e. chromatin rearrangements), and for repression of DNA transcription (i.e. DNA methylation), among the differentially expressed transcripts, hence providing some of the first evidence of the relevance of epigenetic marks in this species and more broadly in the context of ARTs. The role of epigenetic mechanisms in the expression of within-species differential behavioural phenotypes has already been highlighted in other contexts, namely in caste differentiation among social insects (e.g. (Yan et al., 2014)), and recently in relation to social status dependent behaviours of fish (Lenkov et al., 2015). In particular in our study, *dnmt3a*, encoding an enzyme responsible for *de novo* methylation of DNA (Okano et al., 1999), was upregulated in females, transitional and sneaker males, suggesting a common

role for this enzyme in the modulation of behaviour among these phenotypes (e.g. decreased aggression in all three and/or and display of female courtship in females and sneakers) when compared with nest-holders. This enzyme has been recently found to be required for brain feminization in female rat pups within the hormonally sensitive period of sexual differentiation (Nugent et al., 2015). During this period, aromatization of testosterone (T), synthesized by male testes, to oestradiol (E2) within the developing neurons, leads to a suppression of dnmt3a activity, which in turn enables male-specific neuronal organizational effects to take place and a consequent expression of male sexual behaviours in the adult (Nugent et al., 2015). A similar mechanism of brain feminization may be present in the peacock blenny. In this species, gonadal steroids regulate different aspects of reproductive behaviour in the different morphs (Gonçalves et al., 2014, 2007; Oliveira et al., 2001), and sneaker males have lower levels of brain aromatase mRNA (Gonçalves et al., 2010) and of brain aromatase activity (Gonçalves et al., 2008) than nest-holder males. Thus, neuronal intracellular levels of E2 are expected to be low in sneakers in comparison with nest-holder males, consistent with the observed up-regulation of *dnmt3a* in sneakers. Further experiments are necessary to test this hypothesis further.

4.5.3. Genetic architecture of female mimicry in sneaker males

Despite marked sex differences in sexual behaviour between nest-holders and females, in our study, relatively few genes were expressed differentially between the sexes, differences being more pronounced among morphs within the same sex (*i.e.* nest-holder vs. sneaker vs. transitional males). Thus, it is possible that the observed sex differences in reproductive behaviour are due to these few male- or female-biased genes. Therefore, in order to test the hypothesis that the lack of male reproductive behaviour (i.e. nesting and territorial defence) in sneakers was due to a demasculinization of their brain transcriptome, whereas their expression of female-like behaviour was linked to a feminization of their brain transcriptome, we looked into the expression of male- and female-biased genes across morphs. The fact that both sneaker and transitional males have a lower expression of male-biased brain transcripts than nest-holder males reflects a demasculinization of the brains of these morphs, which may reflect their lack of expression of male reproductive behaviour. On the other hand, the fact that both sneaker and transitional males have a higher expression of female-biased brain transcripts than

nest-holder males suggests a feminization of these male morphs. Thus, the brain transcriptome of both sneaker and transitional males is intersex, rather than typically male (or female). However, as from these two male morphs only sneaker males are sexually active and express female-like reproductive behaviour (i.e. courtship and nuptial colouration) the feminization of the sneaker's brain transcriptome (that also occurs in transitional males) cannot be associated with the expression of female-like behaviour by these males. Therefore, female-courtship expressed by females and by sneaker males seems to be associated with different brain transcriptomic architectures. This result is further supported by the lack of shared gene modules between sneakers and females in our WGCNA analysis. However, it has been recently shown that individual dimorphically expressed genes in specific regions of the brain may control one or a few components of a sex-typical behaviour in mice (Xu et al., 2012; Yang and Shah, 2014). Therefore, two alternative hypotheses may explain the genetic basis of female mimicry in the peacock blenny sneaker males: (i) despite their kinematic and structural similarities, the courtship behaviour expressed by females and sneaker males is controlled by different genetic programmes; or (ii) despite the lack of differences in the average expression of female-biased genes between sneakers and transitional males, there are specific female-biased genes, which are also biased in sneakers but not in transitional males, that regulate female courtship in both sneakers and females. In support of the latter hypothesis, we found 12 transcripts that were exclusively up-regulated in sneakers and females, which have a functional annotation that suggests an involvement on neural plasticity, in some cases with interactions with oestrogen-responsive elements. Interestingly, the mate search component of female courtship in this species has been shown to be oestrogen-dependent (Gonçalves et al., 2014). Thus, the hypothesis of a shared genetic factor underlying the expression of female courtship in both females and sneakers is the most parsimonious of the two presented above. Newly available techniques of genome editing (e.g. CRISPR/Cas9, (Heidenreich and Zhang, 2016; Lee et al., 2016)) open the door to testing the role of specific candidate genes in the control of female courtship behaviour in both females and sneakers.

4.6. Conclusions

In summary, our results are consistent with previous studies in species with ARTs and show that at the brain level, intrasexual behavioural plasticity is accompanied by

broader changes in gene expression than intersexual differences in sex-dimorphism, which is in clear contrast to what has been found in gonad (Dean et al., 2017; Harrison et al., 2015; Pointer et al., 2013) and whole body (Stuglik et al., 2014) transcriptional variation. However, the lack of reference genomes for most teleosts limits much of the work on ARTs possibly leaving other major regulatory and structural mechanisms unexplored (e.g. supergene in the ruff *Philomachus pugnax*; (Küpper et al., 2015; Lamichhaney et al., 2015)). Furthermore, our results for the peacock blenny also indicate that at the brain level, reproductive plasticity can be explained by varying the magnitude of sex-biased gene expression. A shift in expression of sex-biased genes has been found in response to sexual selection in mating behaviour (Hollis et al., 2014; Immonen et al., 2014) and has been considered one of the evolutionary processes for differential male and female adaptation. Our results suggest that a similar process may mediate the evolution of alternative reproductive morphs within one of the sexes, in our case males, that explore new areas of the sex-biased transcriptional landscape.

4.7. Acknowledgements

S.D.C. would like to thank Oliver Rupp from the Bioinformatics and Systems Biology group for the advices and assistance during the assembly of the dataset. This study was supported by the research grants PTDC/MAR/69749/2006 and EXCL/BIA-ANM/0549/2012 from the Portuguese Foundation for Science and Technology (FCT), and grant no. 012/2012/A1 from the Macao Science and Technology Development Fund (FDCT). During the writing of this manuscript S.D.C. was being supported by a Ph.D. fellowship (SFRH/BD/89072/2012) from FCT.

4.8. Data Accessibility

Both raw data and transcriptome assembly were deposited in BioProject portal at NCBI (PRJNA329073).

4.9. Author Contributions

SDC, DG, AVMC and RFO were responsible for the design and implementation of the study. Assembly of the sequence data was performed by SDC and AG, and subsequent

analyses by SDC. SDC and RFO wrote the manuscript with contributions from all authors.

4.10. References

Almada V.C., Gonçalves E.J., Oliveira R.F. & Santos A.J., 1995. Courting females: ecological constraints affect sex roles in a natural population of the blenniid fish Salaria pavo. *Animal Behaviour*, *49*, 1125–1127.

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Anders S. & Huber W., 2010. Differential expression analysis for sequence count data. *Genome Biology*, *11*, R106.

Andrews S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M. & Sherlock G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*, 25–29.

Aubin-Horth N., Landry C.R., Letcher B.H. & Hofmann H.A., 2005. Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 1655–1662.

Barata E.N., Serrano R.M., Miranda A., Nogueira R., Hubbard P.C. & Canário A.V.M., 2008. Putative pheromones from the anal glands of male blennies attract females and enhance male reproductive success. *Animal Behaviour*, *75*, 379–389.

Benjamini Y. & Hochberg Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.

Cardoso S.D., Gonçalves D., Robalo J.I., Almada V.C., Canário A.V.M. & Oliveira R.F., 2013. Efficient isolation of polymorphic microsatellites from high-throughput sequence data based on number of repeats. *Marine Genomics*, *11*, 11–16.

Cardoso S.D., Teles M.C. & Oliveira R.F., 2015. Neurogenomic mechanisms of social plasticity. *Journal of Experimental Biology*, *218*, 140–149.

Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M. & Robles M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676.

Dean R., Wright A.E., Marsh-Rollo S.E., Nugent B.M., Alonzo S.H. & Mank J.E., 2017. Sperm competition shapes gene expression and sequence evolution in the ocellated

wrasse. *Molecular Ecology*, *26*, 505–518.

Ellegren H. & Parsch J., 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*, 689–698.

Estabrook G., Almada V., Almada F. & Robalo J., 2002. Analysis of conditional contingency using ACTUS2 with examples from studies of animal behavior. *acta ethologica*, *4*, 73–80.

Fagundes T., Simões M.G., Saraiva J.L., Ros A.F.H., Gonçalves D. & Oliveira R.F., 2015. Birth date predicts alternative life history pathways in a fish with sequential reproductive tactics. *Functional Ecology*, *29*, 1533–1542.

Falcon S. & Gentleman R., 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*, 257–258.

Fishelson L., 1963. Observations on littoral fishes of Israel. I. Behaviour of Blennius pavo Risso (Teleostei: Blenniidae). *Israel Journal of Zoology*, *12*, 67–80.

Fraser B.A., Janowitz I., Thairu M., Travis J. & Hughes K.A., 2014. Phenotypic and genomic plasticity of alternative male reproductive tactics in sailfin mollies. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132310.

Gaudet H.M., Cheng S.B., Christensen E.M. & Filardo E.J., 2015. The G-protein coupled estrogen receptor, GPER: The inside and inside-out story. *Molecular and Cellular Endocrinology*, *418*, 207–219.

Gonçalves D., Alpedrinha J., Teles M. & Oliveira R.F., 2007. Endocrine control of sexual behavior in sneaker males of the peacock blenny Salaria pavo: effects of castration, aromatase inhibition, testosterone and estradiol. *Hormones and Behavior*, *51*, 534–541.

Gonçalves D., Costa S.S., Teles M.C., Silva H., Ingles M. & Oliveira R.F., 2014. Oestradiol and prostaglandin F2 regulate sexual displays in females of a sex-role reversed fish. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20133070.

Gonçalves D., Saraiva J., Teles M., Teodósio R., Canário A.V.M. & Oliveira R.F., 2010. Brain aromatase mRNA expression in two populations of the peacock blenny Salaria pavo with divergent mating systems. *Hormones and Behavior*, *57*, 155–161.

Gonçalves D., Teles M., Alpedrinha J. & Oliveira R.F., 2008. Brain and gonadal aromatase activity and steroid hormone levels in female and polymorphic males of the peacock blenny Salaria pavo. *Hormones and Behavior*, *54*, 717–725.

Gonçalves D.M., Barata E.N., Oliveira R.F. & Canário A.V.M., 2002. The role of male visual and chemical cues on the activation of female courtship behaviour in the sex-role reversed peacock blenny. *Journal of Fish Biology*, *61*, 96–105.

Gonçalves E.J. & Almada V.C., 1997. Sex differences in resource utilization by the peacock blenny. *Journal of Fish Biology*, *51*, 624–633.

Gonçalves E.J., Almada V.C., Oliveira R.F. & Santos A.J., 1996. Female mimicry as a mating tactic in males of the blenniid fish Salaria pavo. *Journal of the Marine*

Biological Association of the United Kingdom, *76*, 529–538.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652.

Harris R.M. & Hofmann H.A., 2014. Neurogenomics of behavioral plasticity., in: Landry, C.R., Aubin-Horth, N. (Eds.), Ecological Genomics. Advances in Experimental Medicine and Biology. Springer, Dordrecht, pp. 149–168.

Harrison P.W., Mank J.E. & Wedell N., 2012. Incomplete sex chromosome dosage compensation in the Indian meal moth, Plodia interpunctella, based on de novo transcriptome assembly. *Genome Biology and Evolution*, *4*, 1118–1126.

Harrison P.W., Wright A.E., Zimmer F., Dean R., Montgomery S.H., Pointer M.A. & Mank J.E., 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4393–4398.

Heidenreich M. & Zhang F., 2016. Applications of CRISPR-Cas systems in neuroscience. *Nature Reviews Neuroscience*, *17*, 36–44.

Hollis B., Houle D., Yan Z., Kawecki T.J. & Keller L., 2014. Evolution under monogamy feminizes gene expression in Drosophila melanogaster. *Nature Communications*, *5*, 689–698.

Hu Z.-L., Bao J. & Reecy J.M., 2008. CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. *Online Journal of Bioinformatics*, *9*, 108–112.

Immonen E., Snook R.R. & Ritchie M.G., 2014. Mating system variation drives rapid evolution of the female transcriptome in Drosophila pseudoobscura. *Ecology and Evolution*, *4*, 2186–2201.

Küpper C., Stocks M., Risse J.E., dos Remedios N., Farrell L.L., McRae S.B., Morgan T.C., Karlionova N., Pinchuk P., Verkuil Y.I., Kitaysky A.S., Wingfield J.C., Piersma T., Zeng K., Slate J., Blaxter M., Lank D.B. & Burke T., 2015. A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, *48*, 79–83.

Lamichhaney S., Fan G., Widemo F., Gunnarsson U., Thalmann D.S., Hoeppner M.P., Kerje S., Gustafson U., Shi C., Zhang H., Chen W., Liang X., Huang L., Wang J., Liang E., Wu Q., Lee S.M.-Y., Xu X., Höglund J., Liu X. & Andersson L., 2015. Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics*, *48*, 84–88.

Langfelder P. & Horvath S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

Langmead B., Trapnell C., Pop M. & Salzberg S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, R25.

Lee H.B., Sundberg B.N., Sigafoos A.N. & Clark K.J., 2016. Genome Engineering with TALE and CRISPR Systems in Neuroscience. *Frontiers in Genetics*, *7*, 47.

Lenkov K., Lee M.H., Lenkov O.D., Swafford A. & Fernald R.D., 2015. Epigenetic DNA Methylation Linked to Social Dominance. *PloS ONE*, *10*, e0144750.

Li B. & Dewey C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Linke B., Giegerich R. & Goesmann A., 2011. Conveyor: a workflow engine for bioinformatic analyses. *Bioinformatics*, *27*, 903–911.

Losel R.M., Falkenstein E., Feuring M., Schultz A., Tillmann H.-C., Rossol-Haseroth K. & Wehling M., 2003. Nongenomic Steroid Action: Controversies, Questions, and Answers. *Physiological Reviews*, *83*, 965–1016.

Maggiolini M. & Picard D., 2010. The unfolding stories of GPR30, a new membrane-bound estrogen receptor. *Journal of Endocrinology*, *204*, 105–114.

Mank J.E., 2017. The transcriptional architecture of phenotypic dimorphism. *Nature Ecology & Evolution*, *1*, 0006.

Mank J.E., 2009. Sex Chromosomes and the Evolution of Sexual Dimorphism: Lessons from the Genome. *The American Naturalist*, *173*, 141–150.

Nugent B.M., Stiver K.A., Alonzo S.H. & Hofmann H.A., 2016. Neuroendocrine profiles associated with discrete behavioural variation in Symphodus ocellatus, a species with male alternative reproductive tactics. *Molecular Ecology*, *25*, 5212–5227.

Nugent B.M., Wright C.L., Shetty A.C., Hodes G.E., Lenz K.M., Mahurkar A., Russo S.J., Devine S.E. & McCarthy M.M., 2015. Brain feminization requires active repression of masculinization via DNA methylation. *Nature Neuroscience*, *18*, 690–697.

Okano M., Bell D.W., Haber D.A. & Li E., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, *99*, 247–257.

Oliveira R.F., Almada V.C., Forsgren E. & Gonçalves E.J., 1999. Temporal variation in male traits, nesting aggregations and mating success in the peacock blenny. *Journal of Fish Biology*, *54*, 499–512.

Oliveira R.F., Carneiro L. a, Gonçalves D.M., Canario a V & Grober M.S., 2001. 11-Ketotestosterone inhibits the alternative mating tactic in sneaker males of the peacock blenny, Salaria pavo. *Brain, Behavior and Evolution*, *58*, 28–37.

Partridge C.G., MacManes M.D., Knapp R. & Neff B.D., 2016. Brain Transcriptional Profiles of Male Alternative Reproductive Tactics and Females in Bluegill Sunfish. *PLoS ONE*, *11*, e0167509.

Patzner R.A., 1983. The reproduction of Blennius pavo (Teleostei, Blenniidae). I. Ovarial cycle, environmental factors and feeding. *Helgoländer Meeresuntersuchungen*, *36*, 105.

Patzner R.A., Seiwald M., Adlgasser M. & Kaurin G., 1986. The reproduction of Blennius

pavo (Teleostei, Blenniidae) V. Reproductive behavior in natural environment. *Zoologischer Anzeiger*, *216*, 338–350.

Pointer M.A., Harrison P.W., Wright A.E. & Mank J.E., 2013. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genetics*, *9*, e1003697.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G. & Mesirov J.P., 2011. Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26.

Robles J.A., Qureshi S.E., Stephen S.J., Wilson S.R., Burden C.J. & Taylor J.M., 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*, *13*, 484.

Saraiva J.L., Barata E.N., Canário A.V.M. & Oliveira R.F., 2009. The effect of nest aggregation on the reproductive behaviour of the peacock blenny Salaria pavo. *Journal of Fish Biology*, *74*, 754–762.

Schunter C., Vollmer S. V, Macpherson E. & Pascual M., 2014. Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics*, *15*, 167.

Snell-Rood E.C., Cash A., Han M. V, Kijimoto T., Andrews J. & Moczek A.P., 2011. Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles. *Evolution*, *65*, 231–245.

Soneson C. & Delorenzi M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14*, 91.

Stiver K.A., Harris R.M., Townsend J.P., Hofmann H.A. & Alonzo S.H., 2015. Neural Gene Expression Profiles and Androgen Levels Underlie Alternative Reproductive Tactics in the Ocellated Wrasse, Symphodus ocellatus. *Ethology*, *121*, 152–167.

Stuglik M.T., Babik W., Prokop Z. & Radwan J., 2014. Alternative reproductive tactics and sex-biased gene expression: the study of the bulb mite transcriptome. *Ecology and Evolution*, *4*, 623–632.

Suzuki R. & Shimodaira H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, *22*, 1540–1542.

Taborsky M., 1994. Sneakers, Satellites, and Helpers: Parasitic and Cooperative Behavior in Fish Reproduction. *Advances in the Study of Behavior*, *23*, 1–100.

Taborsky M., Oliveira R.F. & Brockmann H.J., 2008. The evolution of alternative reproductive tactics: concepts and questions, in: Oliveira, R.F., Taborsky, M., Brockmann, H.J. (Eds.), Alternative Reproductive Tactics: An Integrative Approach. Cambridge University Press, Cambridge, pp. 1–21.

Todd E. V., Black M.A. & Gemmell N.J., 2016. The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology*, *25*, 1224–1241.

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

Whitfield C.W., Ben-Shahar Y., Brillet C., Leoncini I., Crauser D., Leconte Y., Rodriguez-Zas S. & Robinson G.E., 2006. Genomic dissection of behavioral maturation in the honey bee. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 16068–16075.

Xu X., Coats J.K., Yang C.F., Wang A., Ahmed O.M., Alvarado M., Izumi T. & Shah N.M., 2012. Modular genetic control of sexually dimorphic behaviors. *Cell*, *148*, 596–607.

Yan H., Simola D.F., Bonasio R., Liebig J., Berger S.L. & Reinberg D., 2014. Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews Genetics*, *15*, 677–688.

Yang C.F. & Shah N.M., 2014. Representing sex in the brain, one module at a time. *Neuron*, *82*, 261–278.

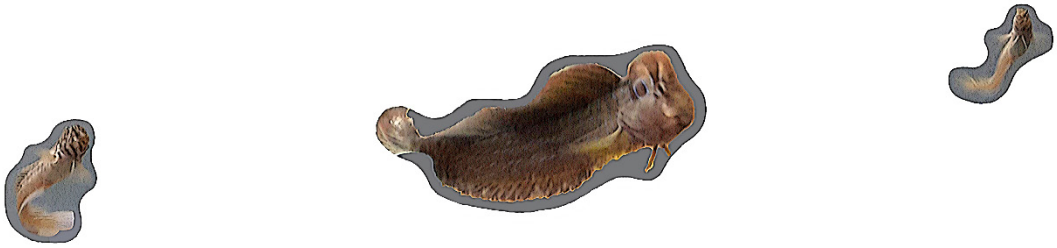
Yu G., Li F., Qin Y., Bo X., Wu Y. & Wang S., 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, *26*, 976–978.

Zayed A. & Robinson G.E., 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annual Review of Genetics*, *46*, 591–615.

Zhang B. & Horvath S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article17.

Zhang Z.H., Jhaveri D.J., Marshall V.M., Bauer D.C., Edson J., Narayanan R.K., Robinson G.J., Lundberg A.E., Bartlett P.F., Wray N.R. & Zhao Q.-Y., 2014. A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. *PLoS ONE*, *9*, e103207.

Zuk M. & Balenger S.L., 2014. Behavioral ecology and genomics: new directions, or just a more detailed map? *Behavioral Ecology*, *25*, 1277–1282.



CHAPTER 5 |

GENOMIC PATTERNS OF PLASTIC SEX ROLES

Cardoso S.D., Gonçalves D., Saraiva J. & Oliveira R.F., 2019. Transcriptomic architecture of courtship sex role reversal in the peacock blenny *Salaria pavo*. *Chapter not submitted for publication*

Transcriptomic architecture of courtship sex role reversal in the peacock blenny *Salaria pavo*

Sara D. Cardoso [1,2], David Gonçalves [3], João Saraiva [4], Rui F. Oliveira [1,2,5]

[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal; [2] ISPA – Instituto Universitário, Lisbon, Portugal; [3] Institute of Science and Environment, University of Saint Joseph, Macau, China; [4] CCMAR – Centro de Ciências do Mar, Universidade do Algarve, Faro, Portugal; [5] Champalimaud Research, Fundação Champalimaud, Lisbon, Portugal

5.1. Abstract

Phenotypic plasticity allows an organism to maximize its fitness in a variety of ecological settings, and differences in reproductive traits within and among populations can reveal which genetic and genomic mechanisms underpin this variation in response to changes in the environment. One such example is courtship behaviour, a central component of animal mating, whose expression may be facilitated or constrained in both females and males depending on variable ecological factors that may shift sexual selection pressures. In the peacock blenny *Salaria pavo*, the intensity of mating competition varies among populations and is largely explained by nest site availability during the breeding season. Using two different target tissues, forebrain and gonad, we focus on the genomic regulation of sex roles in courtship behaviour between females and males from two populations under different sexual selection regimes. The Italian population lives in rocky shores, where nesting sites are abundant and nest-holder males defend a courting territory around the entrance of the nest with females assuming a passive role in courtship (i.e. 'conventional' sex roles). In contrast, the Portuguese population lives in a coastal lagoon with scarce nesting sites, and as a consequence nest-holder males rarely leave their nets having an almost passive role in courtship while females display courtship behaviour and nuptial colouration (i.e. reversed sex roles). Here we show that at the forebrain level, variation in gene expression segregates individuals by population rather than by sex, indicating that plasticity in behaviour across populations is associated with variation in neurogenomic expression. On the other hand, at the gonad level, variation in gene expression segregates individuals by sex and then by population, indicating that sexual selection is also acting at the intrasexual level, particularly in nest-holders by paralleling differences in gonad investment. Additionally, the courting sexes in each population did not share a common gene network that

could underlie their propensity to court the opposite sex, hence pointing towards a complex interaction between brain and gonads.

Keywords: courtship behaviour; reversed sex roles; conventional sex roles; forebrain; gonad; transcriptomic states; *Salaria pavo*

5.2. Introduction

Phenotypic plasticity, the ability of a genotype to produce different phenotypes in response to varying environmental conditions, is widespread in nature and has been considered a potential adaptation for thriving in new environments by bringing populations into the premises of an adaptive peak (Price et al., 2003; Schlichting and Pigliucci, 1998; West-Eberhard, 2003). Individuals that show a plastic response, whether it is morphological, physiological or behavioural, have a higher fitness than those that not (West-Eberhard, 1989), and hence these changes may become 'states' that can be observed at the population-level (West-Eberhard, 2003). Recent studies have shown how breeding strategies are predominantly the outcome of the dynamics between both female and male strategies modulated accordingly to resource dispersion and abundance (Gowaty and Hubbell, 2005). These dynamic fluctuations in space and time between sexes promote biases in the relative abundance between males and females that are available to mate at any point in time (i.e. operational sex ratio (OSR; (Emlen and Oring, 1977; Kvarnemo and Ahnesjö, 1996)) generating plastic phenotypes. As such, the rationale behind the evolution and expression of polygamous mating systems, which according to classic evolutionary theory were determined based on the degree of monopolization of mates (Emlen and Oring, 1977), has been revisited to encompass plastic behavioural phenotypes in females and males in response to these ecological factors within and between populations of the same species. Such examples of mating plasticity have been widely demonstrated in fish, and include presence of alternative reproductive tactics (ARTs; e.g. *Salaria pavo* (Almada et al., 1994) and *Symphodus ocellatus* (Alonzo et al., 2000)), mate choice (e.g. *Symphodus ocellatus* (Alonzo, 2008) and *Xiphophorus nigrensis* (Lynch et al., 2012)), parental care (e.g. *Gasterosteus aculeatus* (Kent and Bell, 2018)), and sex role shift in courtship (e.g. *Salaria pavo* (Almada et al., 1995) and *Gobiusculus flavescens* (Amundsen, 2018; Wang et al., 2014)).

From a holistic genomic perspective, the development of new sequencing technologies (e.g. RNA-seq; (Wang et al., 2009)) have allowed to study non-model organisms and expand

hypothesis on how genetic based patterns of response can be modified by sexual selection in natural populations. Most importantly, we can now make inferences not only at species level but also within and between populations, where we observe the majority of plastic phenotypes, to underpin the genetic and genomic pathways being modulated. From the abovementioned examples, discrete reproductive phenotypes (i.e. ARTs) have been studied in recent years, with special emphasis on the neurogenomic mechanisms that showed more pronounced differences among male tactics than between sexes (e.g. (Cardoso et al., 2018; Nugent et al., 2016; Partridge et al., 2016; Todd et al., 2018). Furthermore, at the gonad level sperm competition has been shown to modulate gene expression and sequence evolution among male tactics in the ocellated wrasse (*Symphodus ocellatus*; (Dean et al., 2017)). In contrast, courtship behaviour, a central component of animal mating, has not been so extensively investigated at the genomic level in non-model organisms. Recently in the peacock blenny *Salaria pavo*, the genetic component of courtship behaviour was investigated and a small set of genes was found in common up-regulated in the courting phenotypes (i.e. females and sneaker males) when compared with the passive phenotype (i.e. nest-holder males), which could be mediating this behaviour (Cardoso et al., 2018). Furthermore, it has been shown a genomic conflict between courtship and aggression in threespined sticklebacks (*Gasterosteus aculeatus*; (Sanogo and Bell, 2016)), a finding in line with other studies in model organisms (reviewed in (Anderson, 2016)). Therefore, more studies elucidating the possible genomic mechanisms mediating roles in courtship behaviour are necessary.

In the present study, we investigate the role of gene expression in sex role shift in courtship behaviour between two populations of the peacock blenny *Salaria pavo* (Blenniidae; Fig. 1). This small intertidal fish is usually found in rocky shores of the Mediterranean and adjacent Atlantic region (Zander, 1986). In this species a strong sexual dimorphism is present with nest-holder males being larger than females and exhibiting conspicuous secondary sexual characters (SSCs; *viz.* a head crest and a sex-pheromone producing anal gland), which are used to attract females to their nests for spawning (Barata et al., 2008; Gonçalves et al., 2002). Moreover, in this species nest-holder males provide sole parental care to eggs until hatching (Fishelson, 1963; Patzner et al., 1986). In the peacock blenny *Salaria pavo*, the intensity of mating competition varies among populations and is mostly explained by nest site availability which has an impact on mate availability during the breeding season. A population at the Adriatic Sea, in the Gulf of Trieste, occurs in an area with a rocky substrate where nests are available in abundance.

Nest-holder males establish nests in rock crevices or holes and aggressively defend a territory around the nest. Furthermore, nest-holder males take the initiative in courtship and the frequency of male courtship displays is higher than the frequency of female courtship displays, with females usually assuming a passive role in courtship responding with changes in colouration and few displays before they enter the nest to spawn (i.e. 'conventional' sex roles (Saraiva et al., 2012)). In contrast, a population at Ria Formosa, Culatra island (southern Portugal) occurs in a mudflat area where the only substrates available for nesting are artificial reefs, such as bricks and tiles used by clam culturists to delimit fields. The scarcity of nest sites promotes strong male-male competition for nests, such that only the most competitive males are able to acquire a nest, which consequently leads to a female-biased operational sex ratio and sex role reversal in courtship behaviour in this population (Almada et al., 1994; Saraiva et al., 2012, 2009). After the breeding season starts, males' territory is restricted to the nest itself, with females competing for access to these nests by displaying elaborate courtship behaviours, that consists of the expression of a transient nuptial colouration and in stereotypic movements, more often than nest-holder males (Almada et al., 1995, 1994; Saraiva et al., 2009).

Interestingly, under similar laboratory conditions, nest-holder males from both populations did not differ in sexual or agonistic behaviours towards females, demonstrating that males from the sex reversed population are also able to express significant levels of courtship (Saraiva et al., 2011). However for females, courtship behaviour reflected the differences observed between populations, but in this case the possibility of not having been given enough information to dissociate their respective environmental from laboratory perception of egg presence in the nest was a confounding factor preventing conclusive results (Saraiva et al., 2011). Together these results indicate that the expression of courtship behaviour is plastic in this species, and hence being mediated by differential regulation of gene expression. Here, we specifically tested: (1) if differences in sexual dimorphism in courtship behaviour had a genetic signature and (2) whether these differences were greater at either the intra- or inter-sexual level. Since we have previously shown for the peacock blenny that at the whole brain level females and nest-holder males had a similar neurogenomic expression profile (Cardoso et al., 2018), in this study we focused on the forebrain regions and included the gonad, aiming for a snapshot of the 'brain-gonadal' axis.

Fig. 1 – The peacock blenny (*Salaria pavo*) presents plastic sex roles in courtship behaviour. (A) In coastal lagoon populations (e.g. Culatra island at Ria Formosa, Portugal) nest sites are scarce and highly aggregated, usually found in artificial reefs, and (B) individuals display reversed sex roles in courtship behaviour. Whereas, (C) in rocky shore populations (e.g. Gulf of Trieste, Italy) nest sites are abundant sparsely distributed, and (D) individuals display conventional sex roles in courtship behaviour.

5.3. Materials and methods

5.3.1. Ethics statement

All procedures were performed in accordance with accepted veterinary practice under a "Group-1" license issued by the Direcção-Geral de Veterinária, Ministério da Agricultura, do Desenvolvimento Rural e das Pescas, Portugal (permit number 0421/000/000/2013).

5.3.2. Sample collection and preparation

Fish were collected during the breeding season of 2014 at Culatra Island (Ria Formosa Natural Park, 36°59′N, 7°51′W, Portugal, and at Trieste seashore (Gulf of Trieste, 45°44'56.54"N, 13°39'25.91"E, Italy; for a detailed description of the sampling site locations please see (Almada et al., 1994) and (Saraiva et al., 2012) respectively). Sampling of individuals from both populations occurred within twelve days (i.e. 10th of July at Trieste; 21st and 22nd of

July at Culatra island) using snorkelling or SCUBA, contingent on tide conditions, during mid-morning when the tide was high. All individuals were caught in their spawning territories, after verifying through behavioural observations that both females and nest-holder males were actively displaying mating behaviours. Upon capture, each fish was brought to the surface, euthanized using an overdose of MS-222 (tricaine methanesulfonate anaesthetic; Sigma-Aldrich), and assigned to individual bags that were submersed on an ice bath until dissection in laboratory (Post-mortem interval (PMI) for Trieste = 4h; PMI for Culatra = 6h). Several studies have reported that RNA degradation is minimal for PMIs of several hours (e.g. (Cheviron et al., 2011; Seear and Sweeney, 2008)), which was also the case in this study (Table S1, Appendix V). Due to field conditions, we tried to minimize this effect as far as possible by keeping the target tissues in their natural conditions, enclosed within the body, at the possible lowest temperature of 0 to 4ºC. After dissection, brain macroareas and gonad samples were kept in RNAlater (Sigma-Aldrich) at 4 ºC up to one week, then stored at –80 ºC until RNA extraction. Twenty individuals (five females and five nest-holder males from each population) were selected for gene expression analysis of the forebrain and gonads, establishing five replicates per phenotype. The choice to use the forebrain (i.e. telencephalon and diencephalon) in this study aimed to focus only on the regions containing neural circuits involved in the regulation of social behaviour and decision-making that are most likely driving phenotypic behavioural differences (O'Connell and Hofmann, 2011). Among nest-holders, males from Trieste population had the largest gonads relative to their body size when compared with males from Culatra population (gonadosomatic index (GSI); Kruskal-Wallis test, *chi-squared* = 4.81, *df* = 1, *P-value* = 0.028), which is in accordance to previously reported patterns for the two populations (Saraiva et al., 2010). Whereas for females no difference in their GSI was detected between populations (Kruskal-Wallis test, *chi-squared* = 0.53, *df* = 1, *P-value* = 0.465).

RNA extraction was performed under standardized conditions using Qiagen's RNeasy Lipid Tissue Mini Kit for forebrain samples and RNeasy Lipid Tissue Midi Kit for gonad samples. Briefly, samples were transferred to QIAzol lysis reagent (volume used for forebrain samples = 0.5 ml; V_{testis} = 1 ml; V_{ovary} = 5 ml) and homogenized (on ice) with a pestle using cycles of 45 s until the tissue was no longer visible. RNA was then extracted following Qiagen's protocol with a DNase treatment on column (RNase-Free DNase Set, Qiagen). Samples' RNA quality and concentration were determined using NanoDrop 1000 spectrophotometer and Agilent 2100 Bioanalyzer (Table S1, Appendix V).

5.3.3. Library preparation and RNA sequencing

Preparation and sequencing of the 40 RNA libraries (forebrain and gonad for each of 5 biological replicates per phenotype) were conducted at Eurofins Genomics GmbH (Ebersberg, Germany). Strand-specific cDNA libraries were prepared from RNA selected using polyA+ purification, and each sample was individually barcoded and run across each of five lanes, in a pool of 64 total libraries (additional libraries from a separate project), on the Illumina HiSeq 2500 with chemistry v4 to eliminate technical variance. On average, 13.3 million 125 bp paired-end reads were recovered per sample before quality control (Table S1, Appendix V).

Read quality was assessed with FastQC v0.11.5 (Andrews, 2010), and reads quality trimmed with cutadapt v1.11 (Martin, 2011). Specifically, reads were trimmed when their ends had a Phred score <20, and remaining possible Illumina adapters removed. Reads were removed post-filtering if either read pair was < 50 bases in length. After trimming, there were on average 12.9 million paired-end reads per sample (Table S1, Appendix V).

5.3.4. *De novo* transcriptome assembly, mapping, annotation and normalization

A *de novo* transcriptome assembly was constructed using Trinity assembler v2.1.1 (Grabherr et al., 2011) with default parameters on the combined pool of reads from both tissues. Although a previous multi-tissue transcriptome existed for this species (Cardoso et al., 2018), in this study we chose to perform a new assembly contemplating the genetic information from both populations focusing on the two target tissues. After redundancy and exogenous contig removal (for more details see (Cardoso et al., 2018)), each individual sample was mapped to the Trinity assembly using Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) within RSEM v1.3.0 (Li and Dewey, 2011) with default settings, to obtain expression levels for the 796,652 contigs. On average 70.5% of the pairs of reads from each library mapped back onto the assembly (Table S1, Appendix V). De novo transcriptome assemblies generate numerous noncoding (i.e. erroneously inferred isoforms), chimeric, or rare variants with low expression-level support that should be removed before downstream analysis. Therefore, a series of filters based on the level of expression of each contig were applied to remove this low-level expression noise from the analysis, as performed before for this species (Cardoso et al., 2018) and in others (e.g. (Dean et al., 2017; Harrison et al., 2015, 2012)). First, the best isoform for each Trinity

contig cluster was selected based upon expression level and, in the case of ties, isoform length. RSEM was then used to remap reads to the set of best isoforms to obtain accurate expression levels. Second, noncoding RNA was filtered from the assembly using BLASTN with an E-value cut-off of 10^{-10} between the set of best isoform and *Oreochromis niloticus* (Nile Tilapia) noncoding RNA (Ensembl release 89) (Aken et al., 2017). Third, contigs with <2 FPKM in 3/5 of the samples per phenotype were removed separately in each tissue set of expressed contigs (Fig. S1 and Fig. S2, Appendix V). This allowed retaining phenotype-specific contigs with reasonable confidence (i.e. expressed in at least three of the five biological replicates). Lastly, only contigs containing potential protein coding sequences (CDS) were kept in the assembly, through using TransDecoder v5.3.0 (Haas et al., 2013) option of 'TransDecoder.LongOrfs' in strand-specific mode. This ensures that only contigs with an open reading frame will be considered for analysis, since it is expected that they are driving the behavioural plasticity under the different selective regimes. The transcriptome for downstream analysis contains 28,433 contigs with an average length of 1,716 bp (N50 = 2,375).

Normalization of contigs' expression levels was performed separately for forebrain and gonad using a generalized linear model (GLM) fit to estimate size factors and dispersion across all 10 samples in DESeq2 v1.16.1 (Love et al., 2014) using default parameters, and 'regularized log' transformation transformed (*rlog*) values were obtained. Hierarchical clustering and pairwise correlations (Spearman's ρ) were used to identify possible outliers. For forebrain, not one sample showed significant deviations from either female distribution (Wilcoxon rank-sum test *P*-value = 0.492, mean pairwise correlations among females ρ = 0.99) or from male distribution (Wilcoxon rank-sum test *P*-value = 0.432, mean pairwise correlations among males ρ = 0.988). As well for gonad, not one sample showed significant deviations from either female distribution (Wilcoxon rank-sum test *P*-value = 0.846, mean pairwise correlations among females ρ = 0.961) or from male distribution (Wilcoxon rank-sum test *P*-value = 0.846, mean pairwise correlations among males ρ = 0.938).

5.3.5. Differential expression analysis and functional annotation

Six pairwise comparisons were considered to detect genes differentially expressed: i) differences between the sexes within each population (i.e. nest-holder male Culatra vs. female Culatra, and nest-holder male Trieste vs. female Trieste); ii) differences within sex between populations (i.e. nest-holder male Culatra *vs* nest-holder male Trieste, and female Culatra vs.

female Trieste); iii) differences between sexes not considering population of origin (i.e. nest-holder males vs. females); and lastly, iv) differences between populations not considering sex (i.e. Culatra vs. Trieste). Differential expression was quantified separately for forebrain and gonad by performing the pairwise contrasts using the Wald test for significance of GLM coefficients in DESeq2 v1.16.1 (Love et al., 2014), considering both a fold-change threshold of 2 and *P*-value < 0.05 after Benjamini-Hochberg adjustment for multiple testing (Benjamini and Hochberg, 1995). Similarly, edgeR v3.18.1 (Robinson et al., 2010), was also used to calculate differential expression using the TMM normalization method and the GLM likelihood ratio (LR) test for significance. Of the genes identified using DESeq2 method, 90.9% of the genes for forebrain and 98.8% of the genes for gonad were also identified as differentially expressed using edgeR using the same cut-off thresholds. These genes detected by two different methods, were retained as the final set of genes differentially expressed for each of the pairwise comparisons of interest. For visualization of the global expression patterns of differentially expressed transcripts among phenotypes, 'regularized log' transformed (*rlog*) values were retrieved and Trinity scripts for heatmap generation using the hclust function in R adapted to produce a hierarchical clustering of Z-transformed expression values using Euclidean distance with complete linkage. The reliability of the inferred tree was assessed by 1,000 bootstrap resampling of the expression values using the R package Pvclust v2.0.0 (Suzuki and Shimodaira, 2006). Other statistical analyses and graphing presented in this work were obtained using R (R Core Team, 2018).

Functional annotation was obtained by doing BLASTx searches against the vertebrate division of UniProtKB (March 2018) and NCBI nr (March 2018) databases using an E-value threshold of 1e-5 and minimum ratio coverage of 33% for either the query or the best-hit. Subsequently, BLAST results were imported into Blast2GO (Götz et al., 2008) where Gene Ontology (GO) mapping and Interpro scan were performed and used to annotate the reference genes using default parameters. GO term enrichment for genes detected as differentially expressed were evaluated by GOstats v2.42.0 (Falcon and Gentleman, 2007), using a 'conditional' hypergeometric test with a *P*-value < 0.01. This method accounts for the hierarchical relationships of GO terms by removing genes from ancestor terms if they are significantly enriched in a child term. Hence, a formal correction for multiple testing cannot be applied due to the implicit dependence between neighbouring GO terms, which do not comply with the independence of tests needed for correction of the p-values. Enriched GO-terms were

then slimmed in REVIGO (Supek et al., 2011) and treemaps produced for genes up-regulated in each comparison under analysis. Treemaps' box size correlates to the –log10 p-value of the GO-term enrichment, and box colour to the same upper-hierarchy GO-term.

5.3.6. Transcriptional coexpression network analysis

To investigate the network of co-expressed genes (i.e. gene modules) across phenotypes, expression values were transformed using the 'regularized log' transformation (*rlog*) function in DESeq2 and used to build a weighted coexpression network using WGCNA v1.63 (Langfelder and Horvath, 2008). All forebrain samples were considered to build one signed network of co-expressed genes using a soft-thresholding power (β) set to 9, corresponding to an R-squared of > 0.80 (Zhang and Horvath, 2005) (mean connectivity of 298), with a minimum module size of 30 transcripts. Modules were merged when eigengene dissimilarity between modules was < 0.25. Whereas for gonad samples, due to its inherently high sexual dimorphism, a signed consensus network analysis was obtained for female and male samples. The fit of scale-free topology failed to reach values above 0.8 for reasonable powers in female samples, and hence the soft-thresholding power (β) set to 20 based on the number of samples (mean connectivity of 125.3 for the male network and 95.4 for the female network). Modules were merged when eigengene dissimilarity between modules was < 0.25 and considering a minimum module size of 30 transcripts. Default settings were used for all other WGCNA parameters in both networks. Statistical significance of module correlation with traits of interest was determined when *P*-value < 0.05 after a Benjamini-Hochberg adjustment for multiple testing. These modules were subsequently validated *in silico* by assessing the relationship between transcript significance for each phenotype and module membership (i.e. correlation of the module eigengene with the transcript expression profile).

5.4. Results

A total of 518 million paired-end reads across all samples were retained after quality control and used for a *de novo* transcriptome assembly. After filtering, the assembly contained 28,433 genes, of which 86% had a functional annotation against a known protein in either UniProtKB or NCBI nr databases. From this pool of transcripts, 22,760 genes were expressed in the forebrain and 21,035 in the gonad, having in common 54.03% of genes.

5.4.1. Forebrain differential gene expression across phenotypes

Six pairwise comparisons considering sex and population of origin were made to obtain genes differentially expressed in each tissue, using a false discovery cut-off of 5% and $\log_2$ of the fold-change of 1. From the forebrain sample comparisons, a total of 1,181 genes were detected as differentially expressed across phenotypes, corresponding to 5.2% of the forebrain expressed genes. Normalized expression profiles represented by hierarchical clustering were obtained considering all phenotypes. Interestingly, phenotypes clustered by population, with females and nest-holder males from each population forming their own clusters, with principal component 1 (PC1) accounting for 53% of the observed variance in gene expression (Fig. 2 and Fig. 3A). When considering gene average expression, phenotypes could be further separated by sex with high confidence.

In relation to the pairwise comparisons within populations, the contrast between females and nest-holder males had the lowest number of genes differentially expressed (Table 1; Tables S2 and S3, Appendix V). Nest-holder males in both comparisons up-regulated galanin (*gal*) and islet amyloid polypeptide (*iapp*) genes, having in common an enrichment in processes related to regulation of receptor activity and hormone activity (Fig. 4A; Fig. S3-S4, Appendix V). Whereas females in both comparisons up-regulated the vasoactive intestinal peptide (*vip*) gene, originating the enrichment for biological processes related to prolactin secretion due to their shared overexpression of the gene (Fig. 4; Fig. S3-S4, Appendix V). Additionally, nest-holders from Culatra had an enrichment for processes involved in cGMP biosynthesis and guanylate cyclase activity, whereas nest-holders from Trieste possessed an enrichment for mitochondrial electron transport and females for processes involved in regulation of cell-cell adhesion and mechanically-gated ion channel activity (Fig. S3-S4, Appendix V).

In relation to the pairwise comparisons within sexes, the contrast between nest-holder males from each population had a total of 789 genes differentially expressed (82% annotated), with nest-holders from Culatra overexpressing 515 of the detected genes (Table 1; Table S4, Appendix V). In this contrast, nest-holder males from Culatra had an enrichment for cell chemotaxis, positive regulation of immune effector process and positive regulation of low-density lipoprotein particle receptor binding (Fig. S5A, Appendix V), whereas nest-holders from Trieste had an enrichment in processes involved in neurological system process, autocrine

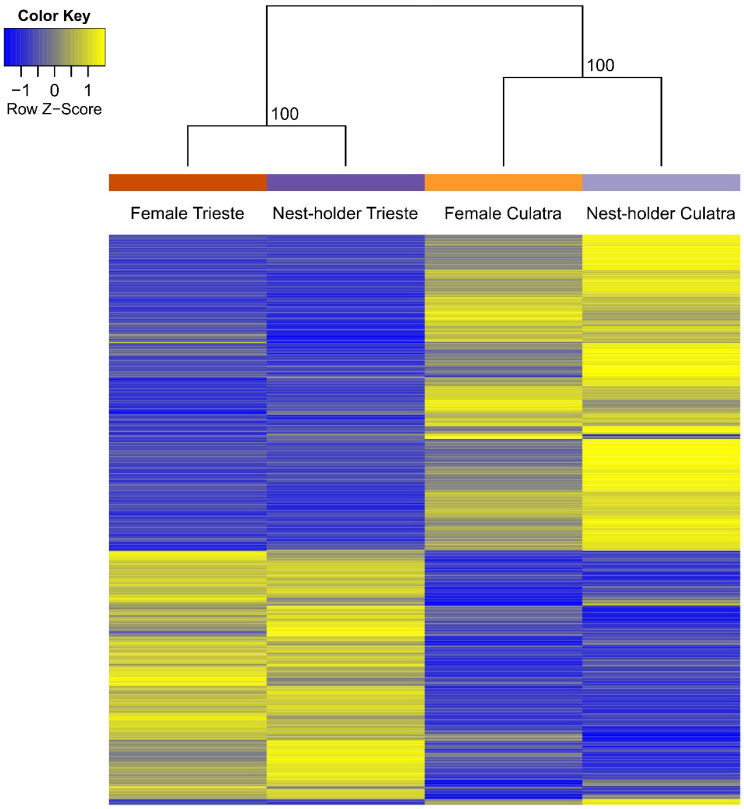


Fig. 2 – Differences in forebrain expression patterns across *Salaria pavo* phenotypes. Heatmap of average regularized expression values for each group, considering all genes differentially expressed (*N* = 1,181). Similarity between phenotypes with hierarchical clustering can be seen above the heatmap with respective bootstrap values. Each phenotype is colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

Table 1 –Number of significantly expressed transcripts for forebrain tissue and percentage of annotated transcripts in each pairwise comparison between phenotypes of *Salaria pavo*.

	Total	Annotated
Nest-holder Culatra > Female Culatra	20	85%
Female Culatra > Nest-holder Culatra	3	100%
Nest-holder Trieste > Female Trieste	3	100%
Female Trieste > Nest-holder Trieste	2	100%
Nest-holder Culatra > Nest-holder Trieste	515	85.24%
Nest-holder Trieste > Nest-holder Culatra	274	75.55%
Female Culatra > Female Trieste	162	66.05%
Female Trieste > Female Culatra	206	73.79%
Nest-holder > Female	1	100%
Female > Nest-holder	1	100%
Culatra > Trieste	446	77.13%
Trieste > Culatra	302	70.20%

> indicates higher expression in the phenotype on the left; FDR adjusted significance value of 0.05.

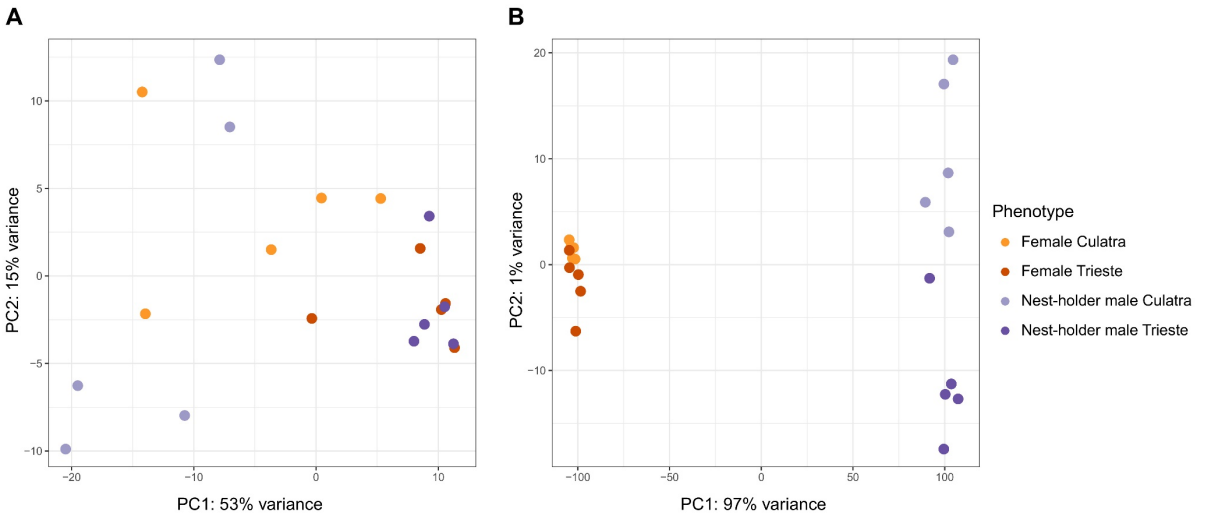


Fig. 3 – Clustering of samples visualised by principal component analysis (PCA) for the differential expressed genes obtained in A) Brain (N = 1,181) and B) Gonad (N = 14,785) pairwise comparisons. Expression values were transformed regularized log transformed transcriptome-wide gene expression data.

signalling and negative regulation of transcription by competitive promoter binding (Fig. S5B, Appendix V). On the other hand, the contrast between females from each population had a total of 368 genes differentially expressed (70% annotated), with females from Trieste overexpressing 206 of the detected genes (Table 1; Table S5, Appendix V). In this contrast, females from Culatra had an enrichment for negative regulation of glucocorticoid receptor signalling pathway, circadian regulation of gene expression, apoptotic DNA fragmentation, and monosaccharide biosynthesis (Fig. S6A, Appendix V), whereas females from Trieste had an enrichment in processes involved in cGMP biosynthesis, adrenergic receptor pathway and protein folding (Fig. S6B, Appendix V).

Looking into the global comparison between sexes (Table 1; Table S6, Appendix V), only two genes were found to be differentially expressed, galanin (*gal*) was up-regulated in nest-holder males and DnaJ heat shock protein family (Hsp40) member B5 (*dnajb5*) up-regulated in females. Lastly, the contrast between populations not considering sex produced the second largest number of differentially expressed genes (748 genes, 74% annotated; Table 1; Table S7, Appendix V). From this set of genes, 75% were shared either with the contrast within nest-holders or with the contrast within females.

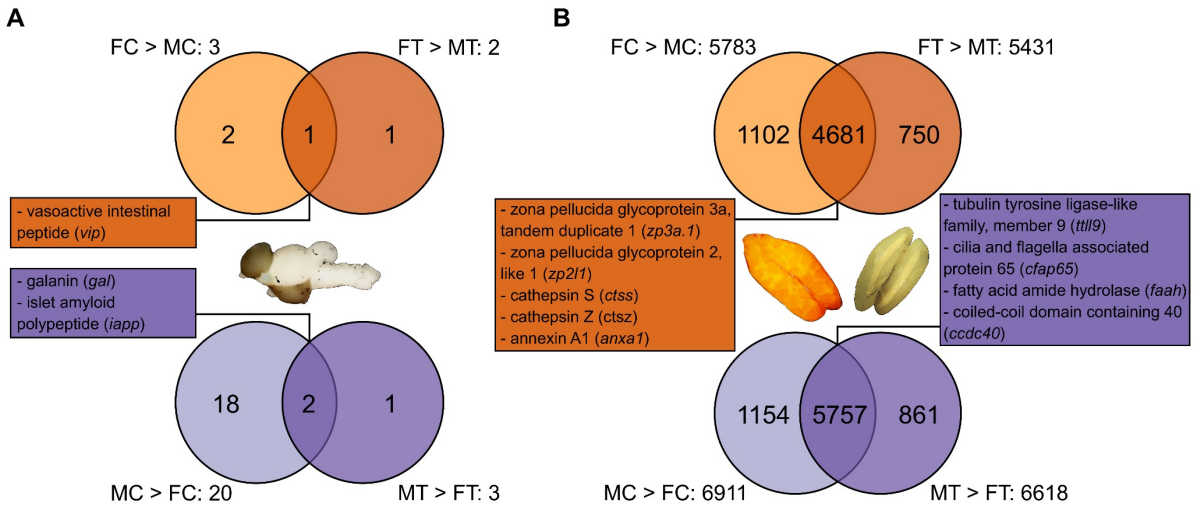


Fig. 4 – Number of genes differentially expressed in A) forebrain and B) gonad obtained in the pairwise comparisons performed within populations. For each comparison, ellipses indicate the proportion of genes up-regulated in common the other sex, with the indication of some genes with the highest fold-change in gonad contrasts. FC, female Culatra; FT, female Trieste; MC, nest-holder male Culatra; MT, nest-holder male Trieste.

5.4.2. Gonadal differential gene expression across phenotypes

Regarding the gonad sample comparisons, a total of 14,785 genes were detected as differentially expressed, corresponding to 70.3% of the gonadal expressed genes. Normalized expression profiles represented by hierarchical clustering, showed, as expected, that samples clustered first by sex, and afterwards by population with high confidence (Fig. 5). Principal component analysis comprising the genes detected as differentially expressed, clearly separated samples by sex, with PC1 accounting for 97% of the observed variance in gene expression, and also by population. From this analysis, we can also observe that male expression between populations is more divergent than female expression.

In relation to the pairwise comparisons within populations, the contrast between females and nest-holder males had the highest number of genes differentially expressed (Table 2; Tables S8 and S9, Appendix V). Genes with highest fold-change in both comparisons for nest-holders were involved in spermatogenesis, such as fatty acid amide hydrolase (*faah*) and tubulin tyrosine ligase-like family, member 9 (*ttll9*) genes, having in common an enrichment in processes related to microtubule cytoskeleton organization and regulation of alternative mRNA splicing, via spliceosome (Fig. 4B; Fig. S7-S8, Appendix V). Whereas females in both comparisons up-regulated genes involved in the degradation and remodelling of the

extracellular matrix, such as cathepsins, and having an enrichment for processes involved in the regulation of organ morphogenesis and cell-cell adhesion (Fig. 4B; Fig. S7-S8, Appendix V).

Concerning the pairwise comparisons within sexes, the contrast between nest-holder males from each population had a total of 2,194 genes differentially expressed (87% annotated), with nest-holders from Trieste overexpressing 1,526 of the detected genes (Table 2; Table S10, Appendix V). In this contrast, nest-holder males from Culatra had an enrichment for neuropeptide signalling pathway, positive regulation of voltage-gated sodium channel activity and resolution of meiotic recombinant intermediates (Fig. S9A, Appendix V), whereas nest-holders from Trieste had an enrichment in processes involved in organonitrogen compound catabolism, organic acid transport and cation homeostasis (Fig. S9B, Appendix V). On the other hand, the contrast between females from each population had a total of 862 genes differentially expressed (83% annotated), with females from Trieste overexpressing 539 of the detected genes (Table 1; Table S11, Appendix V). In this contrast, females from Culatra had an enrichment for regulation of receptor activity, endothelial cell-cell adhesion and galactosylceramide biosynthesis (Fig. S10A, Appendix V), whereas females from Trieste had an enrichment in processes involved in negative regulation of vascular permeability, adenylate cyclase-activating adrenergic receptor signalling pathway and negative regulation of cation channel activity (Fig. S10B, Appendix V).

Looking into the global comparison between sexes (Table 2; Table S12, Appendix V), corresponded to the second largest number of differentially expressed genes and had an overlap of 99% of the genes with the two previous pairwise comparisons performed within populations. Lastly, the contrast between populations not considering sex produced the smallest number of genes being differentially expressed at the gonad level (527 genes, 74% annotated; Table 2; Table S13, Appendix V).

5.4.3. Patterns of gene co-expression modules among phenotypes

A co-expression network for forebrain gene expression was obtained considering both sexes together, and as traits of interest ‘Sex’, ‘Population’ and ‘GSI’. The network comprised 33 modules, ranging in size between 36 (MEviolet) and 2,666 (MEturquoise) genes (Fig. S11, Appendix V). From the traits under analysis, only ‘Population’ had a significant correlation, after adjustment of multiple comparisons, with eight gene modules (Table S14A and Fig. S12

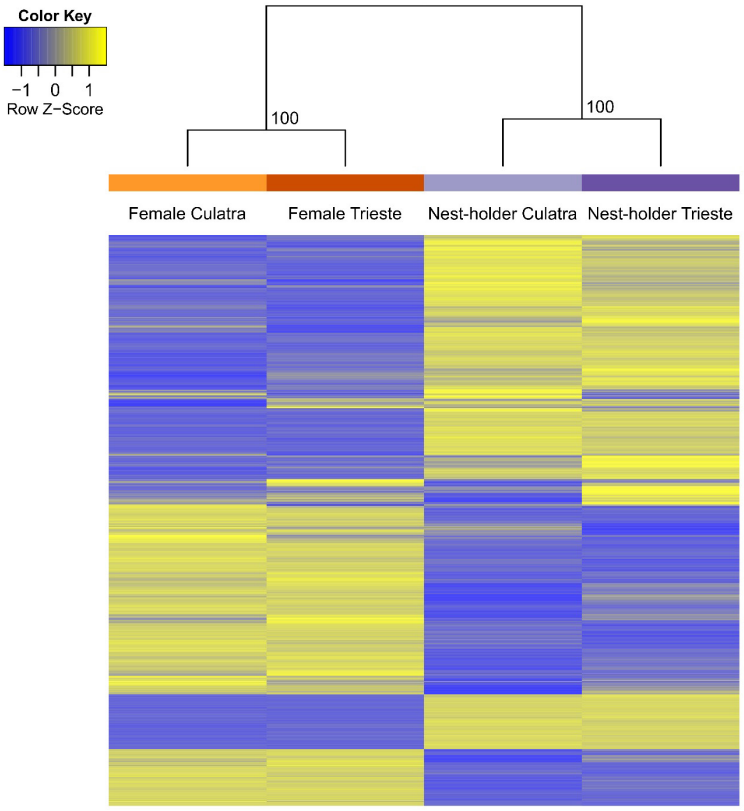


Fig. 5 – Differences in gonadal expression patterns across *Salaria pavo* phenotypes. Heatmap of average regularized expression values for each group, considering all genes differentially expressed (*N* = 14,785). Similarity between phenotypes with hierarchical clustering can be seen above the heatmap with respective bootstrap values. Each phenotype is colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

Table 2 – Number of significantly expressed transcripts for gonad tissue and percentage of annotated transcripts in each pairwise comparison between phenotypes of *Salaria pavo*.

	Total	Annotated
Nest-holder Culatra > Female Culatra	6,911	84.00%
Female Culatra > Nest-holder Culatra	5,783	91.08%
Nest-holder Trieste > Female Trieste	6,618	83.45%
Female Trieste > Nest-holder Trieste	5,431	91.16%
Nest-holder Culatra > Nest-holder Trieste	668	73.80%
Nest-holder Trieste > Nest-holder Culatra	1,526	92.86%
Female Culatra > Female Trieste	323	73.07%
Female Trieste > Female Culatra	539	88.68%
Nest-holder > Female	6,905	83.74%
Female > Nest-holder	5,620	90.91%
Culatra > Trieste	173	50.87%
Trieste > Culatra	354	85.88%

> indicates higher expression in the phenotype on the left; FDR adjusted significance value of 0.05.

and S13, Appendix V). Four modules had a positive correlation, indicating that these genes were up-regulated in both sexes in Culatra population, whereas four modules had a negative correlation representing genes up-regulated in both sexes in Trieste population (Table S14A, Fig. S12-S13, Appendix V). With the exception of module MEpaleturqoise, the remaining seven modules all contained genes differentially expressed in any of the comparisons for forebrain.

In line with the inherent sexual dimorphism for gonad tissue, a consensus WGCNA was performed for females and nest-holder males, considering 'Population' and 'GSI' as the traits of interest. The consensus network showed that the gonadal transcriptome could be clustered in 93 modules, with modules ranging in size from 49 (module MEmediumpurple4) to 3,547 genes (module MEgrey, corresponding to unassigned genes and hence not biologically relevant; Fig. S14, Appendix V). Once again, only 'Population' had a significant correlation with gene modules, after adjustment of multiple comparisons, in both sexes (Table S14B and S14C, Fig. S14-18, Appendix V). Looking into female correlated modules, seven modules had a positive correlation with 'Population', indicating that these genes were up-regulated in females from Culatra population, whereas ten modules had a negative correlation representing genes up-regulated in females from Trieste population (Table S14B and Fig. S15-S16, Appendix V). Regarding male correlated modules, 24 modules had a positive correlation with 'Population', indicating that these genes were up-regulated in males from Culatra population, whereas 18 modules had a negative correlation representing genes up-regulated in males from Trieste population (Table S14C, Fig. S17-S18, Appendix V). All of these modules contained genes differentially expressed in any of the comparisons for gonad.

Interestingly, only two modules were consensual for females and male analysis, MEturqoise having a positive correlation with 'Population' and 69.4% of its genes also found to be differentially expressed. This module had an enrichment for processes involved in ciliary basal body organization, Okazaki fragment processing involved in mitotic DNA replication and intermembrane lipid transfer activity. The second module, MEgreen had a negative correlation with 'Population' and 78.1% of its genes also found to be differentially expressed. This module had an enrichment for processes involved physiological muscle hypertrophy, response to testosterone and estradiol 17-beta-dehydrogenase activity.

5.5. Discussion

Here, we report patterns of regulatory expression in the forebrain and gonad transcriptomes of wild-caught peacock blenny, a species that presents variation in sex roles in courtship behaviour across populations (Almada et al., 1995; Saraiva et al., 2012). Previous work in this system has revealed low levels of whole-brain gene expression differences between females and nest-holder males within Culatra population (Cardoso et al., 2018), but mechanisms regulating courtship behaviour remained elusive.

The design of our study makes it possible to compare both inter- and intra-sexual transcriptional variation across two populations with different selective regimes, Culatra population where a female-biased operational sex ratio shifts the role in courtship behaviour to females, whereas in Trieste 'conventional' sex roles are observed. Previous work on whole-brain tissue showed greater intra-sexual than inter-sexual differences in transcription (Cardoso et al., 2018), which was also verified in this study. When nest-holder males were compared with females, the comparison for Culatra population showed a higher number of genes differentially expressed ($N = 23$) than the Trieste comparison ($N = 5$). In contrast, intra-sexual comparisons clearly showed an effect for 'Population' generating the greatest differences in gene expression, with nest-holder males from Culatra overexpressing 515 genes when compared with nest-holders from Trieste. These results once more demonstrate that at the brain level the pattern of gene expression reflects more the plasticity observed in sex roles, rather than sex differences (i.e. which population the individual belonged to rather than sex). On the other hand, variation in gene expression in the gonad showed a different clustering of the peacock blenny phenotypes. The greatest differences found were first by sex, and then by population (Fig. 3 and 5). This pattern, although partially expected, indicates that the gonad transcriptome is less plastic than the forebrain one. It also shows that nest-holders have the greatest divergence in gonadal gene expression across populations, indicating that males may be under sexual selection pressures in gamete production. A recent study has shown that male intrasexual competition leading to differential investment in absolute gonad mass is paralleled by gene expression and sequence evolution in the ocellated wrasse (Dean et al., 2017). Here a similar mechanism may be driving the observed divergence among nest-holders from Culatra and Trieste, since they invest differently in their gonads as shown by their GSIs.

No clear gene was found at the brain level that could be mediating courtship behaviour in the peacock blenny. Genes that were previously isolated as possible mediators of courtship behaviour in both females and sneaker males when compared to nest-holder males at Culatra population (Cardoso et al., 2018) were not detected in this study in the within-population comparison for Culatra. However, this assessment has to take into account the level of analysis of each study, whole-brain *versus* forebrain analyses. Interestingly, not one gene was found to be shared between females from Culatra and nest-holders from Trieste in their respective contrasts within population. Courtship behaviour in this species is structurally distinct in females and nest-holder males, whereas females in Culatra present a transient nuptial colouration and intense stereotypic movements displays towards nest-holder males (Almada et al., 1994), nest-holder males in Trieste display a conspicuous 'figure-8 swimming' courtship behaviour with colouration changes directed towards females (Patzner et al., 1986). Taken together, the results suggest that courtship behaviour in each sex is controlled by different genetic programmes. Of notice, females in Culatra up-regulated the neurexophilin 1 (*nxph1*) gene when compared to nest-holders. This gene is known to modulate synaptic function and strength through specific binding to the α-neurexin (Born et al., 2014). Nonetheless this gene was not differentially expressed intra-sexually in females, and thus points to other physiological mechanisms that may be mediating behavioural plasticity, such as hormones (Gonçalves et al., 2008).

Together these results reflect fundamental differences between somatic and gonad (*sensu* (Dean et al., 2017)) transcriptional variation. Although, somatic tissues, in this case forebrain, in general tend to show far less intersexual variation than the gonad (e.g. (Harrison et al., 2015; Todd et al., 2018)), in the latter case resulting from the profound functional and physiological differences and delivering male and female gametes, they present large differences in gene expression patterns regulating intrasexual variation. A recent study showed that under biased sex ratios, female traits evolved more than male traits (Fritzsche et al., 2016). These findings could have been translated in to our work under the form of differences in gene expression levels. However, we only detected an explicit down-regulation for females when compared against nest-holders.

5.6. Conclusions

In summary, our results show the importance of looking into females and males using different organ-level analyses to detect broad-scale expression differences that may indicate whether selective pressures may be acting differently across populations. Mechanisms underlying phenotypic plasticity between females and males have been under the spotlight in recent years in an attempt to explain the evolutionary processes for differential male and female adaptation (e.g. (Dean et al., 2017; Todd et al., 2018)). Although it has long been appreciated that natural populations possess genetic variation in the extent of plasticity, recent evidence suggests that epigenetic variation could also play an important role in mediating the emergence of such variation through the interactions between environment and genetic variation shaping phenotypic responses (Hu and Barrett, 2017; Ledón-Rettig, 2013). Even though, this species is wide-spread across the Mediterranean and adjacent Atlantic, studies have shown a very low divergence observed in *S. pavo*, although a genetic pattern between Atlanto-Mediterranean region has been detected indicating an asymmetric migration between populations (Almada et al., 2009; Castilho et al., 2017). Therefore, future work should investigate if such processes play a role in the peacock blenny *Salaria pavo*, specifically whether genetic divergence between populations is also a component in the observed differences in courtship behaviour in this species and which epigenetic mechanisms may be underlying the differences in gene expression. Moreover this finding should be complemented by looking into sex-biased genes, since it has been shown a shift in expression of these genes in response to sexual selection in mating behaviour (Hollis et al., 2014; Immonen et al., 2014).

5.7. Acknowledgements

S.D.C. would like to thank Magda Teles from the Integrative Behavioural Biology Group, for the assistance in verifying the dissections of the brain tissue. S.D.C and J.S. would like to thank Professor Serena Fonda-Umani from the University of Trieste for allowing the use of the wet laboratory at the University of Trieste, and Flávia Silva for her assistance in the sampling at Culatra. This study was supported by the research grants PTDC/MAR/69749/2006 and EXCL/BIA-ANM/0549/2012 from the Portuguese Foundation for Science and Technology (FCT), and grant no. 012/2012/A1 from the Macao Science and Technology Development Fund (FDCT). During the writing of this

manuscript S.D.C. was being supported by a Ph.D. fellowship (SFRH/BD/89072/2012) from FCT.

5.8. References

Aken B.L., Achuthan P., Akanni W., Amode M.R., Bernsdorff F., Bhai J., Billis K., Carvalho-Silva D., Cummins C., Clapham P., Gil L., Girón C.G., Gordon L., Hourlier T., Hunt S.E., Janacek S.H., Juettemann T., Keenan S., Laird M.R., Lavidas I., Maurel T., McLaren W., Moore B., Murphy D.N., Nag R., Newman V., Nuhn M., Ong C.K., Parker A., Patricio M., Riat H.S., Sheppard D., Sparrow H., Taylor K., Thormann A., Vullo A., Walts B., Wilder S.P., Zadissa A., Kostadima M., Martin F.J., Muffato M., Perry E., Ruffier M., Staines D.M., Trevanion S.J., Cunningham F., Yates A., Zerbino D.R. & Flicek P., 2017. Ensembl 2017. *Nucleic Acids Research*, *45*, D635–D642.

Almada V.C., Gonçalves E.J., Oliveira R.F. & Santos A.J., 1995. Courting females: ecological constraints affect sex roles in a natural population of the blenniid fish Salaria pavo. *Animal Behaviour*, *49*, 1125–1127.

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Almada V.C., Robalo J.I., Levy A., Freyhof J., Bernardi G. & Doadrio I., 2009. Phylogenetic analysis of peri-Mediterranean blennies of the genus Salaria: molecular insights on the colonization of freshwaters. *Molecular Phylogenetics and Evolution*, *52*, 424–431.

Alonzo S.H., 2008. Female mate choice copying affects sexual selection in wild populations of the ocellated wrasse. *Animal Behaviour*, *75*, 1715–1723.

Alonzo S.H., Taborsky M. & Wirtz P., 2000. Male alternative reproductive behaviours in a Mediterranean wrasse, Symphodus ocellatus: Evidence from otoliths for multiple life-history pathways. *Evolutionary Ecology Research*, *2*, 997–1007.

Amundsen T., 2018. Sex roles and sexual selection: lessons from a dynamic model system. *Current Zoology*, *64*, 363–392.

Anderson D.J., 2016. Circuit modules linking internal states and social behaviour in flies and mice. *Nature Reviews Neuroscience*, *17*, 692–704.

Andrews S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Barata E.N., Serrano R.M., Miranda A., Nogueira R., Hubbard P.C. & Canário A.V.M., 2008. Putative pheromones from the anal glands of male blennies attract females and enhance male reproductive success. *Animal Behaviour*, *75*, 379–389.

Benjamini Y. & Hochberg Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society,*

Series B, *57*, 289–300.

Born G., Breuer D., Wang S., Rohlmann A., Coulon P., Vakili P., Reissner C., Kiefer F., Heine M., Pape H.-C. & Missler M., 2014. Modulation of synaptic function through the α-neurexin-specific ligand neurexophilin-1. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E1274-1283.

Cardoso S.D., Gonçalves D., Goesmann A., Canário A.V.M. & Oliveira R.F., 2018. Temporal variation in brain transcriptome is associated with the expression of female mimicry as a sequential male alternative reproductive tactic in fish. *Molecular Ecology*, *27*, 789–803.

Castilho R., Cunha R.L., Faria C., Velasco E.M. & Robalo J.I., 2017. Asymmetrical dispersal and putative isolation-by-distance of an intertidal blenniid across the Atlantic-Mediterranean divide. *PeerJ*, *5*, e3195.

Cheviron Z.A., Carling M.D. & Brumfield R.T., 2011. Effects of Postmortem Interval and Preservation Method on RNA Isolated from Field-Preserved Avian Tissues. *The Condor*, *113*, 483–489.

Dean R., Wright A.E., Marsh-Rollo S.E., Nugent B.M., Alonzo S.H. & Mank J.E., 2017. Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. *Molecular Ecology*, *26*, 505–518.

Emlen S.T. & Oring L.W., 1977. Ecology, sexual selection, and the evolution of mating systems. *Science*, *197*, 215–223.

Falcon S. & Gentleman R., 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*, 257–258.

Fishelson L., 1963. Observations on littoral fishes of Israel. I. Behaviour of Blennius pavo Risso (Teleostei: Blenniidae). *Israel Journal of Zoology*, *12*, 67–80.

Fritzsche K., Booksmythe I. & Arnqvist G., 2016. Sex Ratio Bias Leads to the Evolution of Sex Role Reversal in Honey Locust Beetles. *Current Biology*, *26*, 2522–2526.

Gonçalves D., Teles M., Alpedrinha J. & Oliveira R.F., 2008. Brain and gonadal aromatase activity and steroid hormone levels in female and polymorphic males of the peacock blenny Salaria pavo. *Hormones and Behavior*, *54*, 717–725.

Gonçalves D.M., Barata E.N., Oliveira R.F. & Canário A.V.M., 2002. The role of male visual and chemical cues on the activation of female courtship behaviour in the sex-role reversed peacock blenny. *Journal of Fish Biology*, *61*, 96–105.

Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J. & Conesa A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*, 3420–3435.

Gowaty P.A. & Hubbell S.P., 2005. Chance, Time Allocation, and The Evolution of Adaptively Flexible Sex Role Behavior. *Integrative and Comparative Biology*, *45*, 931–944.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke

A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N. & Regev A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512.

Harrison P.W., Mank J.E. & Wedell N., 2012. Incomplete sex chromosome dosage compensation in the Indian meal moth, Plodia interpunctella, based on de novo transcriptome assembly. *Genome Biology and Evolution*, *4*, 1118–1126.

Harrison P.W., Wright A.E., Zimmer F., Dean R., Montgomery S.H., Pointer M.A. & Mank J.E., 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4393–4398.

Hollis B., Houle D., Yan Z., Kawecki T.J. & Keller L., 2014. Evolution under monogamy feminizes gene expression in Drosophila melanogaster. *Nature Communications*, *5*, 689–698.

Hu J. & Barrett R.D.H., 2017. Epigenetics in natural animal populations. *Journal of Evolutionary Biology*, *30*, 1612–1632.

Immonen E., Snook R.R. & Ritchie M.G., 2014. Mating system variation drives rapid evolution of the female transcriptome in Drosophila pseudoobscura. *Ecology and Evolution*, *4*, 2186–2201.

Kent M. & Bell A.M., 2018. Changes in behavior and brain immediate early gene expression in male threespined sticklebacks as they become fathers. *Hormones and Behavior*, *97*, 102–111.

Kvarnemo C. & Ahnesjö I., 1996. The dynamics of operational sex ratios and competition for mates. *Trends in Ecology & Evolution*, *11*, 404–408.

Langfelder P. & Horvath S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

Langmead B. & Salzberg S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Ledón-Rettig C.C., 2013. Ecological epigenetics: an introduction to the symposium. *Integrative and Comparative Biology*, *53*, 307–318.

Li B. & Dewey C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Love M.I., Huber W. & Anders S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550.

Lynch K.S., Ramsey M.E. & Cummings M.E., 2012. The mate choice brain: comparing

gene profiles between female choice and male coercive poeciliids. *Genes, Brain, and Behavior*, *11*, 222–229.

Martin M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10.

Nugent B.M., Stiver K.A., Alonzo S.H. & Hofmann H.A., 2016. Neuroendocrine profiles associated with discrete behavioural variation in Symphodus ocellatus, a species with male alternative reproductive tactics. *Molecular Ecology*, *25*, 5212–5227.

O'Connell L.A. & Hofmann H.A., 2011. Genes, hormones, and circuits: An integrative approach to study the evolution of social behavior. *Frontiers in Neuroendocrinology*, *32*, 320–335.

Partridge C.G., MacManes M.D., Knapp R. & Neff B.D., 2016. Brain Transcriptional Profiles of Male Alternative Reproductive Tactics and Females in Bluegill Sunfish. *PLoS ONE*, *11*, e0167509.

Patzner R.A., Seiwald M., Adlgasser M. & Kaurin G., 1986. The reproduction of Blennius pavo (Teleostei, Blenniidae) V. Reproductive behavior in natural environment. *Zoologischer Anzeiger*, *216*, 338–350.

Price T.D., Qvarnström A. & Irwin D.E., 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*, 1433–40.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Robinson M.D., McCarthy D.J. & Smyth G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.

Sanogo Y.O. & Bell A.M., 2016. Molecular mechanisms and the conflict between courtship and aggression in three-spined sticklebacks. *Molecular Ecology*, *25*, 4368–4376.

Saraiva J.L., Barata E.N., Canário A.V.M. & Oliveira R.F., 2009. The effect of nest aggregation on the reproductive behaviour of the peacock blenny Salaria pavo. *Journal of Fish Biology*, *74*, 754–762.

Saraiva J.L., Gonçalves D.M. & Oliveira R.F., 2010. Environmental modulation of androgen levels and secondary sex characters in two populations of the peacock blenny Salaria pavo. *Hormones and Behavior*, *57*, 192–197.

Saraiva J.L., Gonçalves D.M., Simões M.G. & Oliveira R.F., 2011. Plasticity in reproductive behaviour in two populations of the peacock blenny. *Behaviour*, *148*, 1457–1472.

Saraiva J.L., Pignolo G., Gonçalves D. & Oliveira R.F., 2012. Interpopulational variation of the mating system in the peacock blenny Salaria pavo. *Acta Ethologica*, *15*, 25–31.

Schlichting C.D. & Pigliucci M., 1998. Phenotypic Evolution: A Reaction Norm Perspective. Sinauer Associates Incorporated.

Seear P.J. & Sweeney G.E., 2008. Stability of RNA isolated from post-mortem tissues of Atlantic salmon (Salmo salar L.). *Fish Physiology and Biochemistry*, *34*, 19–24.

Supek F., Bošnjak M., Škunca N. & Šmuc T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS ONE*, *6*, e21800.

Suzuki R. & Shimodaira H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, *22*, 1540–1542.

Todd E. V, Liu H., Lamm M.S., Thomas J.T., Rutherford K., Thompson K.C., Godwin J.R. & Gemmell N.J., 2018. Female Mimicry by Sneaker Males Has a Transcriptomic Signature in Both the Brain and the Gonad in a Sex-Changing Fish. *Molecular Biology and Evolution*, *35*, 225–241.

Wang S., Ramsey M.E. & Cummings M.E., 2014. Plasticity of the mate choice mind: courtship evokes choice-like brain responses in females from a coercive mating system. *Genes, brain, and behavior*, *13*, 365–375.

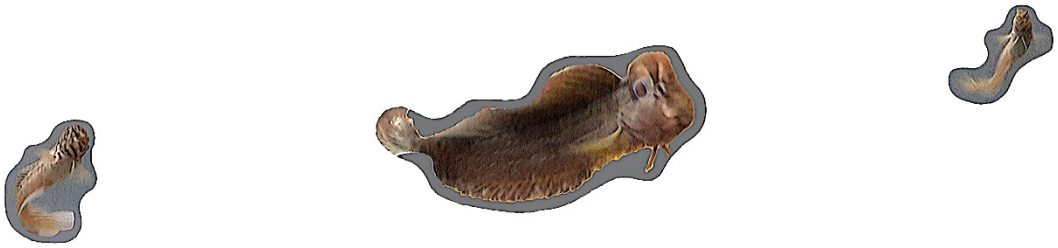
Wang Z., Gerstein M. & Snyder M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63.

West-Eberhard M.J., 2003. Developmental Plasticity and Evolution. Oxford University Press, New York.

West-Eberhard M.J., 1989. Phenotypic plasticity and the origins of diversity. *Annual Review of Ecology and Systematics*, *20*, 249–278.

Zander C.D., 1986. Blenniidae, in: Whitehead, P.J.P., Bauchot, M.L., Hureau, J.C., Nielsen, J., Tortonese, E. (Eds.), Fishes of the North-Eastern Atlantic and the Mediterranean. UNESCO, Paris, pp. 1096–1112.

Zhang B. & Horvath S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article17.



CHAPTER 6 |

GENERAL DISCUSSION & FUTURE DIRECTIONS

GENERAL DISCUSSION

The molecular basis of ecological variation is one of the fundamental research areas in the post-genomic era. By integrating a vast amount of information coming from RNA sequencing experiments, it yields the power to generate more comprehensive knowledge on the layers of complexity that operate and interact in the evolution and development of behavioural plasticity. The main aim of this thesis was to apply state-of-the-art RNA-Seq analyses in the peacock blenny *Salaria pavo* and examine which genetic and genomic mechanisms underpin the variation observed in this species in response to changes in the social and physical environment. In doing so, I pursued different methods of analysis and applied rigorous statistical methods to ensure high confidence in my findings. Taking advantage of the sequencing data, I also developed genetic markers that were used to complement the current understanding of ART dynamics in this species. This system, together with transcriptomic analyses, provides a unique opportunity to investigate the regulatory and genetic factors underlying phenotypic variation in the wild. The main findings are discussed in each chapter. Here, I discuss how these results fit together with previous work and provide some future perspectives.

6.1. Fertilization success of male ARTs

Microsatellite markers, due to their high polymorphism levels, are still widely used for estimating genetic diversity within populations and applied for parentage and kinship analysis (e.g. (Cogliati et al., 2014; Weber et al., 2018)). Recent studies comparing single-nucleotide polymorphisms (SNPs) and microsatellites point to the benefit of using randomly selected high-throughput sequencing-based SNP data to estimate genetic diversity within and differentiation among populations (Fischer et al., 2017). However, concordance can be found when comparing both SNP and microsatellite markers for assigning parentage, and therefore SNP data has been demonstrated to become only advantageous in studies that require high-throughput genotyping or that plan to address a range of ecological and evolutionary questions (Kaiser et al., 2017).

In this thesis, we opted for the development of a set of polymorphic microsatellite markers, to apply in the well-characterized population at Culatra Island and provide insight on how successful sneakers are in the fertilization of eggs. One of the strategies adopted in this study, had in its basis a *de novo* transcriptome assembly containing high

levels of genetic diversity since almost 27% of the evaluated 'complete' microsatellites were polymorphic *in silico*. Furthermore, this strategy of mining for variability *in silico* had been recently shown to expedite the screen for polymorphism when microsatellites were applied in the population (Hoffman and Nichols, 2011). However, in our study when selecting microsatellites based either on their polymorphism *in silico* or on their length (i.e. number of repeats) only the latter was found to be a predictor of polymorphic microsatellites, contrarily to what was found in the previous study by Hoffman and Nichols (2011). The results we obtained might be explained by three main factors: 1) microsatellites with longer repeat lengths are likely to be more variable due to higher mutation rates since replication slippage may increase in proportion to the number of repeats (Petit et al., 2005); 2) selected microsatellites with *in silico* polymorphism were located in coding regions, which could make these microsatellites functionally constrained (Li et al., 2004); and 3) differences in coverage across the transcriptome were present, especially for longer microsatellites that had only a mean coverage of 3.43 ± 2.75 reads. Nonetheless, the results show the advantage of combining the knowledge of number of repeats with other predictors of variability (i.e. *in silico* microsatellite polymorphism) in improving the rates of polymorphism when microsatellites are applied in natural populations.

The expression of alternative reproductive tactics (ARTs) in the peacock blenny is sequential and irreversible and dependent on body size at the time of the first breeding season (Fagundes et al., 2015). Hence, differences in fitness for male tactics are expected, and its quantification should help to understand better how these tactics evolve and are maintained in populations (Gross, 1996). In this system, reproductive success can be used as a proxy for fitness, and is composed of two components, mating success (i.e. how successful an individual is in obtaining mating events) and fertilization success (i.e. how successful an individual is in fertilizing eggs). Previous studies in teleost species with paternal care of eggs have shown great variability in the percentage of nests with parasitic fertilizations and as well in the success of each male ART (for a review see (Avise et al., 2002; Coleman and Jones, 2011)). Surprisingly, when taking into consideration the investment in female mimicry made by sneaker males in this species, our work reports that on average 95% of the eggs are fertilized by the nest-holder male, being one of highest rates of fertilization success reported for the male providing parental care. Inversely, in the plainfin midshipman fish (*Porichthys notatus*), a species where males can be guarders (aka type I males) that compete for nest sites and court females, or

sneakers (aka type II males) that attempt to steal fertilizations from the guarder males during spawning (Brantley and Bass, 1994), different patterns have been detected. Cogliati et al. (2013) reported for the plainfin midshipman fish guarder males a success of 63%, one of the lowest levels of fertilization success for the nesting tactic described so far in fish. However, fluctuations in time on the fertilization success rate of male ARTs in the midshipman fish were found (Cogliati et al., 2013). These results have been further supported with the comparison of two populations of the plainfin midshipman fish, where guarder males differed in their body size, that showed a remarkable conservation of mating patterns and consistent variations in fertilization success for the guarder male (52% and 58%) (Cogliati et al., 2014). In the peacock blenny, contrary to midshipman fish, the successes of the parasitic tactic is size dependent (Gonçalves et al., 2005; Gonçalves and Oliveira, 2003). Additionally, nest characteristics used by the nest-holder males as spawning sites (i.e. holes with only one partially obstructed entrance (Almada et al., 1994)) are likely an additional factor for the success of nesting tactic in the peacock blenny. Indeed, this factor has been already linked to male siring success in the cichlid *Lamprologus lemairii* (Ota et al., 2014). However, we could not test this factor in our current work. Here, we used nest-holder males' morphological traits and social network parameters as predictors of fertilization success, with only the number of sneakers present in the nest-holders' social networks was found to be a predictor of paternity loss, result that is in accordance with previous observations in this species (Gonçalves et al., 2003). In the work by Gonçalves and colleagues (2003), focal observations in the field showed that sneaker males associated with successful nests and not with ripe females. This results point to the stronger influence of the social environment than of morphological traits in the proportion of lost fertilizations by nest-holder males, however these results cannot be generalized to other species (see (Alonzo and Heckman, 2010)).

In comparison to the hundreds of species with documented alternative reproductive tactics, relatively few have been tested for the prediction of equal or unequal fitness in male ARTs. One explanation might be due to the challenges of measuring reproductive success for each male tactic, particularly when reproductive success fluctuates in both space and time (Cogliati et al., 2013; Neff and Clare, 2008). Although more studies are necessary for the peacock blenny, which should take into account multiple samplings throughout the breeding season and control for the total number of eggs and egg cohorts present in the nests, here we obtained the first measure of fertilization success for male ARTs in the peacock.

6.2. Genomic patterns associated with plastic reproductive phenotypes

In the last decade, transcriptomic (RNA-Seq) studies have shown that genomic expression varies greatly throughout an organism's body and through time. Understanding how, when and why this variability can be achieved is, therefore, essential to unravel key questions in evolution as well as in ecology (Alvarez et al., 2015). Most importantly, this plasticity in gene expression is paralleled by distinct behavioural phenotypes, so that distinct neurogenomic states correspond to different behavioural states (*sensu* (Zayed and Robinson, 2012)), and hence can be used as a first approach to characterize plastic phenotypes.

Working with next-generation sequencing technologies, in particular, RNA-Seq is still challenging when little genetic information is available for the organism under study, and a closely related species with genomic information is absent. In eukaryotes, the need to resolve genetic variability, gene structure and corresponding spliced isoforms from relatively short reads complicates *de novo* transcriptome assembly resulting in large assemblies that need to be methodically curated and reduced to only contain informative biological sequences, and hence facilitate alignment and downstream analysis steps. In order to accomplish this step, controlled methods to remove sequence noise from the assembly should be adopted, as such the FPKM method (for more details see Chapter 1). We have successfully applied this method in reducing the *de novo* assemblies obtained in this species without losing a significant percentage of aligned reads. In chapter 4, the initial assembly contained 577,532 contigs, an average read mappability across samples of 79.7% and only 17.6% of these contigs annotated, whereas the curated assembly comprised 179,202 contigs corresponding to a loss in average read mappability of only 1.9% of the reads and increment in annotation rate to 25.7%. Clearly, strategies equal or similar to this one are beneficial in reducing sequence expression noise and consequently improve gene expression results, and hence the same approach was also applied in the data obtained in chapter 5. A great variety of methods and bioinformatic tools have been developed to optimize the assemblies also in the absence of a reference genome, however, the ultimate goal of reconstructing complete gene transcripts has still not been wholly achieved. Even though some fragmentation present in these assemblies is present, which could to a bias in overexpression of larger transcripts, expression levels with RNA-Seq seem to be consistent and robust, as it has been shown for mammal species (Su et al., 2014). The discovery of new transcripts is one of the significant advantages of RNA-Seq,

however, if similar sequences are not present in the annotation databases no function will be associated with the gene. Especially for non-model species, this is reflected by the low percentage of transcripts with annotations from homologous genes, which still constitutes limitation of studying a non-model species at this moment, in the peacock blenny, we found only 25.7% of the brain transcripts annotated (chapter 4), whereas 42.4% of the transcripts constituting the shared transcriptome of brain and gonad transcripts were annotated, before filtering the assembly for sequences with coding regions (chapter 5). Despite the challenges and limitations, valuable evolutionary results can be drawn from the vast quantities of sequencing data produced by RNA-Seq, which allowed the analyses of genome-wide expression differences in the context of plasticity among reproductive phenotypes.

Peacock blenny is a classic system to investigate behavioural differences in male ARTs and sex role reversal in courtship behaviour. We started by first characterizing the overall expression differences among male ARTs (chapter 4). Our results showed that each peacock blenny phenotype has a distinct transcriptomic profile, which indicates that distinct behavioural repertoires are associated with distinct neurogenomic states, which differentiate not only sex but also male morphs. These results are consistent with previous reports of specific brain transcriptomic profiles among species with ARTs for alternative male morphs, particularly in teleost fishes ((Aubin-Horth et al., 2005; Fraser et al., 2014; Nugent et al., 2016; Partridge et al., 2016; Schunter et al., 2014; Stiver et al., 2015), Chapter 1). From these works, we can also generalize that expression of alternative morphs within the same sex in both fixed and plastic ARTs seems to be achieved through differential gene expression in the brain. Moreover, we also detected that behavioural plasticity (i.e. sneaker male) is accompanied by broader changes in brain gene expression, accounting for 79.1% of the genes detected as differentially expressed in any of the pairwise comparisons, when compared with the other phenotypes (73.7% for transitional males, 69.3% for females and 63.6% for nest-holders). Interestingly, nest-holders differentially up-regulated more genes in the pairwise comparisons against male morphs, which was further substantiated by the detected bias for a preferential up-regulation of gene expression than expected by chance, whereas sneaker and transitional males had a bias towards down-regulation in their global gene expression patterns. Functionally, these genes had an enrichment for metabolism and transmembrane transport of ions in nest-holders, and morphogenesis and development for sneaker males, representing their differences in developmental status. However, when one compares the available brain

RNA-Seq transcriptome data for teleost species with ARTs, no clear pattern emerges. Whereas in some species sneaker males exhibit the most distinctive transcriptome (e.g. *Lepomis macrochirus* (Partridge et al., 2016); *Tripterygion delaisi* (Schunter et al., 2014); present study), in other species nest-holder males are the most differentiated phenotype (e.g. *Symphodus ocellatus* (Nugent et al., 2016)*, Thalassoma bifasciatum* (Todd et al., 2018)). Interestingly, in the bluehead wrasses (*Thalassoma bifasciatum*) the authors show not only that the dominant terminal-phase males are the most differentiated phenotype at the forebrain level, but also a greater gonadal investment is made by sneaker males that entailed pervasive down-regulation of androgenesis genes (Todd et al., 2018), which is consistent with low androgen production in males lacking well-developed secondary sexual characters. Moreover, the lists of differentially expressed genes for functionally equivalent phenotypes (e.g. sneakers) across species do not share significant numbers of transcripts and genes, suggesting that ARTs may have evolved in different species through species-specific genetic architectures. This might be partially explained by the presence of distinct developmental windows and differences in shifts for behavioural traits among the studied species so far. For example, in *Tripterygion delaisi*, if the opportunity to acquire a nest arises during the breeding season, a sneaker male is capable of change into the territorial phenotype, which involves changes in body colouration and behavioural repertoire (De Jonge and Videler, 1989; Wirtz, 1978). Contrarily, in the peacock blenny, this shift in behaviour and morphology takes a longer period of time and usually occurs between breeding seasons, with the appearance of an intermediate developmental male morph (i.e. transitional males), where a trade-off between somatic growth and reproduction is present (Saraiva et al., 2010). Given the amount of RNA-Seq data available from these studies, one way forward would be to do a comparative analysis, controlling for developmental factors, so that the shared genetic modules underlying the different behavioural phenotypes could be determined and more precise inferences made. Along this line, Renn et al. (2017) obtained gene expression signatures for cichlids of the tribe Ectodini from Lake Tanganyika, that suggest the existence of deep molecular homologies underlying the convergent or parallel evolution of monogamy in different cichlid lineages from this tribe, which clearly demonstrates the usefulness of this approach to understand evolutionary processes for plastic traits.

Despite marked sex differences in sexual behaviour between nest-holders and females, relatively few genes were expressed differentially between the sexes at the brain level in both RNA-Seq studies, with the differences being more pronounced among

morphs within the same sex (i.e. nest-holder vs. sneaker vs. transitional males; Chapter 4), and intrasexual between populations (i.e. nest-holder Culatra vs. nest-holder Trieste and female Culatra vs. female Trieste; Chapter 5). Generally, these results are in accordance to previously discussed literature, which shows that the brain transcriptome reflects better reproductive plasticity rather than sex dimorphism. Additionally, it has been recently shown that individual dimorphically expressed genes in specific regions of the brain may control one or a few components of a sex-typical behaviour in mice (Xu et al., 2012; Yang and Shah, 2014). In both our works, we attempted to detect the possible genetic mechanisms underlying courtship behaviour, first in both females and sneakers from the sex role reserved population (Chapter 4), and second in females and nest-holders in both populations displaying different roles (Chapter 5). Overall, the genetic architecture of courtship behaviour in teleosts seems not to be such a 'simple' question when compared to studies in the fruit-fly (*Drosophila melanogaster*). Within invertebrates, *fruitless* gene (*fru*), which encodes a set of transcription factors, has been demonstrated through behavioural studies to be essential for courtship behaviour in the male fruit-fly and thought to be directing the developmental sex-specific neural circuitry that encodes this innate behavioural response (for a review see (Vernes, 2015; Yamamoto and Koganezawa, 2013)). In the first work (Chapter 4), we isolated a set of genes commonly up-regulated in females and sneaker males, but not in transitional males, when compared to nest-holder males, which had a functional enrichment for neural plasticity, in some cases with interactions with oestrogen-responsive elements. This result was in accordance with previous findings in this species, where the mate search component of female courtship has been shown to be oestrogen-dependent (Gonçalves et al., 2014), supporting the hypothesis of a shared genetic factor underlying the expression of female courtship in both females and sneakers. Conversely, in the second work (Chapter 5), at the forebrain level, we did not detect common genes being up-regulated by the courting phenotype (i.e. female Culatra and nest-holder Trieste) in relation to the passive phenotype (i.e. nest-holder Culatra and female Trieste). Recent studies in *Drosophila melanogaster* and mice have identified circuit nodes, constituted by relatively small groups of neurons, that have causal roles in the control of innate social behaviours such as aggression and mating, and also play a part in the encoding of internal states that promote these social behaviours (for a review see (Anderson, 2016)). Such states are probably encoded by the interaction between patterns of circuit activity and neuromodulation (e.g. hormones), which can change the intrinsic firing properties of

circuit neurons and alter effective synaptic strength, and consequently alter their output and give rise to behaviour (Marder, 2012). Along this line of thinking, Sanogo and Bell (2016) showed a genomic conflict between courtship and aggression in threespined sticklebacks. Remarkably, in our work, brain expression signatures clustered individuals without a clear segregation by sex, whereas gonadal expression signatures followed expected patterns, with clear cluster by sex and then by population. Moreover, under similar laboratory conditions, nest-holders from both populations displayed similar levels of courtship towards females, although the same was not established for females (Saraiva et al., 2011). These results taken together (i.e. behavioural and genomic expression) support the hypothesis of an interaction between circuit activity and neuromodulation. However, in our work, we cannot disregard the possibility of the intrasexual comparisons containing genes relevant in courtship behaviour, but further analyses are needed to disentangle the population effect in these results. Additionally, the level of analysis (whole brain (Chapter 4) and forebrain (Chapter 5)), may be masking regional differences (i.e. nodes of the neural circuit) in gene expression between courting and non-courting phenotypes.

6.3. Future directions

Gene expression studies are immensely useful tools for uncovering genes and functional pathways underlying complex traits such as social behaviour. Despite tremendous advances in recent years, significant challenges remain in this area. One issue, particularly within the brain, is the spatial restriction of gene expression: only a small population of cells may express the genes pivotal of influence. Globally, the findings presented in this thesis add to the increasing body of empirical evidence of mechanisms and gene expression patterns underlying plastic social phenotypes, specifically in the context of mating behaviours. Here, I will first make considerations on how to continue contributing to the mounting knowledge in the peacock blenny, and afterwards, general considerations regarding the field of ecological genomics.

The genetic basis of courtship behaviour in the peacock blenny *Salaria pavo*, and other teleost fishes remains elusive and more research work is necessary. The data gathered in Chapter 5, will be used to investigate the possibility of genetic differentiation being already present between Trieste and Culatra populations. To dissect this question, two different evolutionary approaches will be used, sequence divergence and

polymorphism analysis, as defined by Dean and colleagues (2017). Together with the differential expression results gathered in Chapter 5, will enable to consider various mechanisms that may be mediating sexual selection pressures in both nest-holder males and females in this species. Experimental approaches should also be considered to remove from the analysis confounding factors present when sampling individuals from their natural populations. However, working with this species in laboratory conditions is difficult due to its sensitivity to captivity conditions, and up until now, all the attempts to breed this species in captivity have been unsuccessful. Even so, controlled experiments involving female and sneaker males courtship behaviour is a straightforward approach that would bring more insight to female-mimicry in courtship behaviour. A second approach would be to test the effect of nest site availability in courtship behaviour using individuals from both populations (i.e. common garden experiment); although in the peacock blenny, it would have to be done with juveniles due to the limitations mentioned above. Additionally, different levels of analysis, brain macroareas vs. target brain nuclei, as well as different tissues, brain vs. gonads, should be considered for genomic analysis.

In general, the field of ecological genomics, and specifically for the work in the peacock blenny, would benefit with research advancing in three fronts: 1) comparative transcriptomics, which would reveal the conserved building blocks in the emergence of phenotypic traits; 2) integrative framework, incorporate research from multiple levels of analysis, such as the connectome, genome, epigenome, transcriptome, and proteome; 3) reverse genomics, by using genetic knock-down experiments such as interference RNA (RNAi) or the CRISPR-Cas system (for a review see (Boettcher and McManus, 2015)).

6.4. References

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Alonzo S.H. & Heckman K.L., 2010. The unexpected but understandable dynamics of mating, paternity and paternal care in the ocellated wrasse. *Proceedings of the Royal Society of London B: Biological Sciences*, *277*, 115–122.

Alvarez M., Schrey A.W. & Richards C.L., 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology*, *24*, 710–725.

Anderson D.J., 2016. Circuit modules linking internal states and social behaviour in flies and mice. *Nature Reviews Neuroscience*, *17*, 692–704.

Aubin-Horth N., Landry C.R., Letcher B.H. & Hofmann H.A., 2005. Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 1655–1662.

Avise J.C., Jones A.G., Walker D. & DeWoody J.A., 2002. Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. *Annual Review of Genetics*, *36*, 19–45.

Boettcher M. & McManus M.T., 2015. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell*, *58*, 575–585.

Brantley R.K. & Bass A.H., 1994. Alternative Male Spawning Tactics and Acoustic Signals in the Plainfin Midshipman Fish Porichthys notatus Girard (Teleostei, Batrachoididae). *Ethology*, *96*, 213–232.

Cogliati K.M., Mistakidis A.F., Marentette J.R., Lau A., Bolker B.M., Neff B.D. & Balshine S., 2014. Comparing population level sexual selection in a species with alternative reproductive tactics. *Behavioral Ecology*, *25*, 1524–1533.

Cogliati K.M., Neff B.D. & Balshine S., 2013. High degree of paternity loss in a species with alternative reproductive tactics. *Behavioral Ecology and Sociobiology*, *67*, 399–408.

Coleman S.W. & Jones A.G., 2011. Patterns of multiple paternity and maternity in fishes. *Biological Journal of the Linnean Society*, *103*, 735–760.

De Jonge J. & Videler J.J., 1989. Differences between the reproductive biologies of Tripterygion tripteronotus and T. delaisi (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour. *Marine Biology*, *100*, 431–437.

Dean R., Wright A.E., Marsh-Rollo S.E., Nugent B.M., Alonzo S.H. & Mank J.E., 2017. Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. *Molecular Ecology*, *26*, 505–518.

Fagundes T., Simões M.G., Saraiva J.L., Ros A.F.H., Gonçalves D. & Oliveira R.F., 2015. Birth date predicts alternative life history pathways in a fish with sequential reproductive tactics. *Functional Ecology*, *29*, 1533–1542.

Fischer M.C., Rellstab C., Leuzinger M., Roumet M., Gugerli F., Shimizu K.K., Holderegger R. & Widmer A., 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. *BMC Genomics*, *18*, 69.

Fraser B.A., Janowitz I., Thairu M., Travis J. & Hughes K.A., 2014. Phenotypic and genomic plasticity of alternative male reproductive tactics in sailfin mollies. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132310.

Gonçalves D., Costa S.S., Teles M.C., Silva H., Ingles M. & Oliveira R.F., 2014. Oestradiol and prostaglandin F2 regulate sexual displays in females of a sex-role reversed fish. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20133070.

Gonçalves D., Fagundes T. & Oliveira R., 2003. Reproductive behaviour of sneaker males of the peacock blenny. *Journal of Fish Biology*, *63*, 528–532.

Gonçalves D., Matos R., Fagundes T. & Oliveira R., 2005. Bourgeois males of the peacock blenny, Salaria pavo, discriminate female mimics from females? *Ethology*, *111*, 559–572.

Gonçalves D.M. & Oliveira R.F., 2003. Time spent close to a sexual partner as a measure of female mate preference in a sex-role-reversed population of the blenny Salaria pavo (Risso) (Pisces: Blenniidae). *acta ethologica*, *6*, 1–5.

Gross M.R., 1996. Alternative reproductive strategies and tactics diversity within sexes. *Trends in Ecology & Evolution*, *11*, 92–98.

Hoffman J.I. & Nichols H.J., 2011. A novel approach for mining polymorphic microsatellite markers in silico. *PloS ONE*, *6*, e23283.

Kaiser S.A., Taylor S.A., Chen N., Sillett T.S., Bondra E.R. & Webster M.S., 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources*, *17*, 183–193.

Li Y.-C., Korol A.B., Fahima T. & Nevo E., 2004. Microsatellites Within Genes: Structure, Function, and Evolution. *Molecular Biology and Evolution*, *21*, 991–1007.

Marder E., 2012. Neuromodulation of Neuronal Circuits: Back to the Future. *Neuron*, *76*, 1–11.

Neff B.D. & Clare E.L., 2008. Temporal variation in cuckoldry and paternity in two sunfish species (Lepomis spp.) with alternative reproductive tactics. *Canadian Journal of Zoology*, *86*, 92–98.

Nugent B.M., Stiver K.A., Alonzo S.H. & Hofmann H.A., 2016. Neuroendocrine profiles associated with discrete behavioural variation in Symphodus ocellatus, a species with male alternative reproductive tactics. *Molecular Ecology*, *25*, 5212–5227.

Ota K., Awata S., Morita M., Yokoyama R. & Kohda M., 2014. Territorial males can sire more offspring in nests with smaller doors in the cichlid Lamprologus lemairii. *Journal of Heredity*, *105*, 416–422.

Partridge C.G., MacManes M.D., Knapp R. & Neff B.D., 2016. Brain Transcriptional Profiles of Male Alternative Reproductive Tactics and Females in Bluegill Sunfish. *PLoS ONE*, *11*, e0167509.

Petit R.J., Deguilloux M.-F., Chat J., Grivet D., Garnier-Géré P. & Vendramin G.G., 2005. Standardizing for microsatellite length in comparisons of genetic diversity. *Molecular Ecology*, *14*, 885–890.

Renn S.C.P., Machado H.E., Unterwaditzer N., Sessa A.K., Harris R.M. & Hofmann H.A., 2017. Gene Expression Signatures of Mating System Evolution. *Genome*, *61*, 287–297.

Sanogo Y.O. & Bell A.M., 2016. Molecular mechanisms and the conflict between courtship and aggression in three-spined sticklebacks. *Molecular Ecology*, *25*,

4368–4376.

Saraiva J.L., Gonçalves D.M. & Oliveira R.F., 2010. Environmental modulation of androgen levels and secondary sex characters in two populations of the peacock blenny Salaria pavo. *Hormones and Behavior*, *57*, 192–197.

Saraiva J.L., Gonçalves D.M., Simões M.G. & Oliveira R.F., 2011. Plasticity in reproductive behaviour in two populations of the peacock blenny. *Behaviour*, *148*, 1457–1472.

Schunter C., Vollmer S. V, Macpherson E. & Pascual M., 2014. Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics*, *15*, 167.

Stiver K.A., Harris R.M., Townsend J.P., Hofmann H.A. & Alonzo S.H., 2015. Neural Gene Expression Profiles and Androgen Levels Underlie Alternative Reproductive Tactics in the Ocellated Wrasse, Symphodus ocellatus. *Ethology*, *121*, 152–167.

Su Z., Łabaj P.P., Li S., Thierry-Mieg J., Thierry-Mieg D., Shi W., Wang C., Schroth G.P., Setterquist R.A., Thompson J.F., Jones W.D., Xiao W., Xu W., Jensen R. V, Kelly R., Xu J., Shi L. et al., 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology*, *32*, 903–914.

Todd E. V, Liu H., Lamm M.S., Thomas J.T., Rutherford K., Thompson K.C., Godwin J.R. & Gemmell N.J., 2018. Female Mimicry by Sneaker Males Has a Transcriptomic Signature in Both the Brain and the Gonad in a Sex-Changing Fish. *Molecular Biology and Evolution*, *35*, 225–241.

Vernes S.C., 2015. Genome wide identification of Fruitless targets suggests a role in upregulating genes important for neural circuit formation. *Scientific Reports*, *4*, 4412.

Weber K., Hoelzl F., Cornils J.S., Smith S., Bieber C., Balint B. & Ruf T., 2018. Multiple paternity in a population of free-living edible dormice (Glis glis). *Mammalian Biology*, *93*, 45–50.

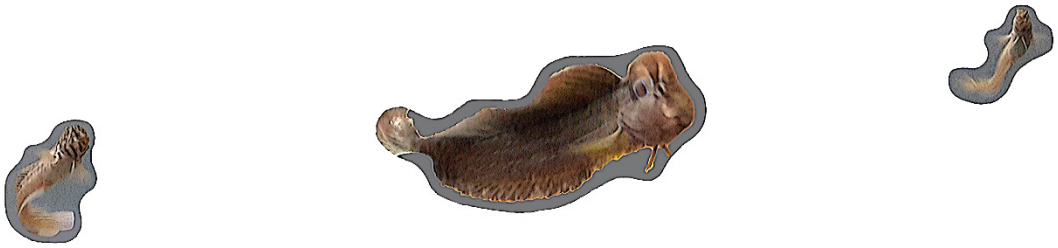
Wirtz P., 1978. The Behaviour of the Mediterranean Tripterygion Species (Pisces, Blennioidei). *Zeitschrift für Tierpsychologie*, *48*, 142–174.

Xu X., Coats J.K., Yang C.F., Wang A., Ahmed O.M., Alvarado M., Izumi T. & Shah N.M., 2012. Modular genetic control of sexually dimorphic behaviors. *Cell*, *148*, 596–607.

Yamamoto D. & Koganezawa M., 2013. Genes and circuits of courtship behaviour in Drosophila males. *Nature Reviews Neuroscience*, *14*, 681–692.

Yang C.F. & Shah N.M., 2014. Representing sex in the brain, one module at a time. *Neuron*, *82*, 261–278.

Zayed A. & Robinson G.E., 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annual Review of Genetics*, *46*, 591–615.



APPENDIX I |

NEUROGENOMIC MECHANISMS OF SOCIAL PLASTICITY

Cardoso S.D., Teles M.C. & Oliveira R.F., 2015. Neurogenomic mechanisms of social plasticity. *Journal of Experimental Biology, 218*, 140–149.
doi: 10.1242/jeb.106997

Neurogenomic mechanisms of social plasticity

Sara D. Cardoso[1,2,3], Magda C. Teles[1,2,3] and Rui F. Oliveira[1,2,3]

[1] ISPA – Instituto Universitário, Rua Jardim do Tabaco 34, 1149-041 Lisbon, Portugal; [2] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal; [3] Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Avenida Brasília, 1400-038 Lisbon, Portugal.

I.1. Abstract

Group-living animals must adjust the expression of their social behaviour to changes in their social environment and to transitions between life-history stages, and this social plasticity can be seen as an adaptive trait that can be under positive selection when changes in the environment outpace the rate of genetic evolutionary change. Here, we propose a conceptual framework for understanding the neuromolecular mechanisms of social plasticity. According to this framework, social plasticity is achieved by rewiring or by biochemically switching nodes of a neural network underlying social behaviour in response to perceived social information. Therefore, at the molecular level, it depends on the social regulation of gene expression, so that different genomic and epigenetic states of this brain network correspond to different behavioural states, and the switches between states are orchestrated by signalling pathways that interface the social environment and the genotype. Different types of social plasticity can be recognized based on the observed patterns of inter-versus intra-individual occurrence, time scale and reversibility. It is proposed that these different types of social plasticity rely on different proximate mechanisms at the physiological, neural and genomic level.

Keywords: Social behaviour, Behavioural flexibility, Neural plasticity, Behavioural states, Behavioural shifts, Epigenetics

I.2. Introduction

The ability to adapt to environmental changes is a ubiquitous characteristic of biological systems. According to classic evolutionary theory, adaptation is achieved by mechanisms that generate genetic diversity (e.g. mutation, recombination) upon which natural selection can act, such that only the organisms with higher fitness under the current environmental conditions will be likely to pass their genes on to the next

generation. Thus, adaptation by natural selection depends on heritable phenotypic variation produced by genetic variation. However, in situations where the rate of environmental change outpaces the rate of genetic evolutionary change, the need for adaptive change without genetic mutation emerges. In this scenario, the evolution of phenotypic plasticity is favoured, according to which environmental cues sensed by the organism lead the same genotype to produce different phenotypes depending on environmental conditions cues (Pigliucci, 2001; West-Eberhard, 2003). Thus, despite the fact that the contribution of non-heritable phenotypic variation to evolutionary change appears to be a paradox, the evolution of mechanisms that generate it must be a common evolutionary phenomenon (Frank, 2011). Different traits show different evolutionary changes in plasticity, in terms of the time lag to respond to the environmental cue and the magnitude of the response. Among animals, behavioural traits exhibit both more rapid and stronger plasticity than morphological traits, which makes behavioural plasticity a key adaptive response to changing environmental conditions (Pigliucci, 2001).

At the proximate level, behavioural plasticity depends on the development of a central nervous system that allows for rapid and integrated organismal responses in order to maintain homeostasis (or allostasis). Many of these responses are simple reflexes and fixed action patterns elicited by a stimulus in the environment, when it determinately predicts an appropriate response. However, when environmental complexity and ambiguity increase, the capacity to adaptively modify behaviour, as a function of experience (i.e. learning) and context, is needed. One of the most ambiguous components of the environment is the social domain, as it is made of other behavioural agents with an inherent level of unpredictability of their actions, with whom the individual needs to interact. Hence, the ability of animals to regulate the expression of social behaviour, so as to adapt their behavioural output to specific situations in a complex and variable social world, is expected to depend on the evolution of plastic responses. These allow the same genotype to produce different behavioural phenotypes (social plasticity), rather than to genetically determine rules controlling fixed responses. Thus, social plasticity should be viewed as a key ecological performance trait that impacts Darwinian fitness (Oliveira, 2009; Taborsky and Oliveira, 2012).

Here we propose a conceptual framework for understanding the genomic mechanisms of social plasticity that has the following premises.

(1) Observed intraspecific variation in behaviour can be characterized by specific behavioural states (*sensu* Zayed and Robinson, 2012), which are exhibited by different

behavioural phenotypes (ethotypes) characterized by the consistent expression of a behavioural profile (i.e. set of behaviours) for a given time period (see Fig. 1 for an example).

(2) These behavioural states are paralleled by specific neural states of a social decision-making network in the brain. This network is composed of two interconnected neural circuits, the social behaviour network (*sensu* Newman, 1999; see also Goodson, 2005) and the mesolimbic reward circuit, that together regulate the expression of social behaviour (O'Connell and Hofmann, 2012, 2011). Within this neural network, reciprocal connections are established between each pair of brain nuclei, such that information is encoded in a distributed and dynamic fashion. Therefore, each behavioural state is better reflected by the overall profile of activation across the network, rather than by the activity of a single node. Different combinations of activation across nodes, and variation in the strength of the connections among them, will generate an almost infinite variation in neural states that would produce equivalent behavioural states.

(3) Given that most nodes of the social decision-making network widely express receptors for neuromodulators (i.e. neuropeptides and amines) and hormones (i.e. sex steroids and glucocorticoids), the state of this network can be regulated by these molecules. This sensitivity of the social decision-making network to physiological modulators allows the integration of global organismic state into social decision making and hence the coordination between relevant brain and body states (Oliveira, 2009).

(4) At the molecular level, the neurobehavioural states mentioned above correspond to specific neurogenomic states (*sensu* Zayed and Robinson, 2012) characterized by distinct patterns of gene expression across the social decision-making network in the brain, such that individuals in different behavioural states exhibit different brain transcriptomes. Differential gene expression in the relevant neural network may change the weight of each node and/or the strength of the connectivity between them, therefore contributing to the generation of multiple network states.

According to this framework, social plasticity is achieved by rewiring or by biochemically switching nodes of the neural network underlying social behaviour in response to perceived social information. Therefore, at the molecular level, it depends on the social regulation of gene expression, so that different neurogenomic states correspond to different behavioural states and the switches between states are orchestrated by signalling pathways that interface the social environment and the genotype (Aubin-Horth and Renn, 2009; Taborsky and Oliveira, 2012). However, social plasticity may occur at

different temporal scales and vary in its degree of reversibility and within versus between individual occurrence, and this variation in the patterns of plasticity may require different proximate mechanisms.

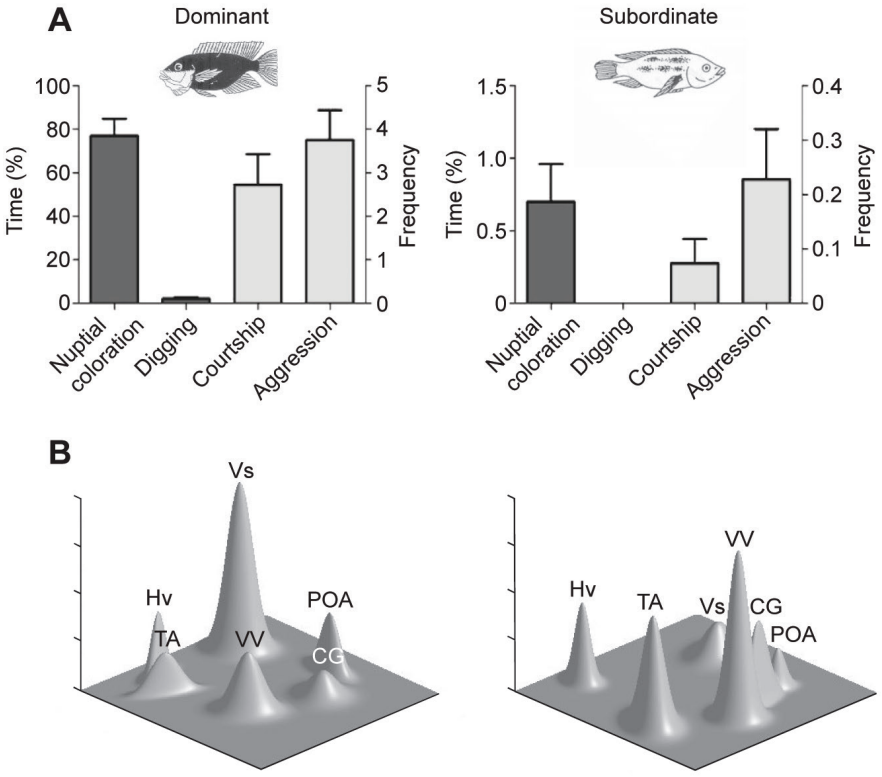


Fig. 1 – Example of paralleled behavioural and neurogenomic states in dominant and subordinate males of the African cichlid fish. (A) Dominant and subordinate males express distinct status-specific behavioural states, characterized by differences in nuptial coloration, digging a bower, courting females and aggressiveness [data from Almeida et al. (2014)]. (B) These divergent behavioural states are expected to be paralleled by different patterns of activation of the social behaviour network in the brain, achieved by differential gene expression. The social behaviour network is composed of six nodes in the forebrain and midbrain areas: supracommissural part of the ventral telencephalon (Vs, teleost homologue of the extended medial amygdala in tetrapods), preoptic area (POA), central grey (CG), ventral subdivision of the ventral telencephalon (VV, homologue of the lateral septum), nucleus anterior tuberis (TA, homologue of the ventromedial hypothalamus) and ventral zone of the periventricular hypothalamus (Hv, homologue of the anterior hypothalamus)

By considering two key parameters of individual variation in behaviour, reversibility and occurrence within versus between individuals, one might identify three types of social plasticity that are expected to rely on different proximate mechanisms and to be responding to different patterns of environmental predictability (Fig. 2). When intraspecific variation in social behaviour occurs only between individuals (i.e. fixed

alternative phenotypes, Fig. 2) it can be due either to a genetic polymorphism [e.g. genetically determined alternative reproductive tactics (Sinervo and Lively, 1996)], in which case it should not be viewed as a true case of social plasticity, or to developmental plasticity dependent on early social environment that is predictive of the adult environment and thus directs the individual towards alternative developmental pathways. When it occurs within the same individual, it can be irreversible with long-lasting behavioural states, such as changes in behaviour between different life-history stages (e.g. juvenile behavioural state versus adult behavioural state), or reversible usually with short-lived behavioural states within the same life-history stage (Kappeler and Kraus, 2010; Piersma and Drent, 2003; Schradin, 2013) (see Fig. 2). These three types of social plasticity potentially rely on different mechanisms at each of the levels discussed above (neural, physiological and genomic). Although classic examples of these different types of social plasticity come from different species, it should be stressed here that they are not mutually exclusive and thus may occur within the same species. For example, in an intertidal blenniid fish (peacock blenny, *Salaria pavo*) alternative tactics occur with smaller and younger males behaving as female mimics to sneak fertilizations and larger and older males being territorial and attracting females to spawn (Almada et al., 1994). When sneakers grow older they become territorial nest-holder males, which corresponds to a sequential plasticity pattern (Gonçalves et al., 1996). However, not all males go through the sneaker phase; depending on their birth date, males follow different developmental routes, with some of them growing directly into nest holders (Fagundes et al., 2015). Thus, the two alternative developmental pathways constitute an example of fixed alternative phenotypes that co-exist with the intra-individual developmental plasticity observed in sneakers.

At the neural level, social plasticity can be achieved by two major mechanisms of neural plasticity: structural reorganization or biochemical switching of relevant neural circuits (Zupanc and Lamprecht, 2000), depending on the time scale of the plasticity. Irreversible patterns of social plasticity (i.e. fixed alternative phenotypes or sequential plasticity), which involve long-lasting changes in behavioural states, can be achieved through different forms of structural modifications that might require addition or removal of cells in these circuits (i.e. neurogenesis or apoptosis), modification of the connectivity between different components of the network (i.e. changes in the dendrite and/or axon structures) or alteration of the responsiveness of the circuit (i.e. balance of neurotransmitter and/or neuromodulator receptors) (Oliveira, 2009; Zupanc and

Lamprecht, 2000). In contrast, reversible patterns of social plasticity (i.e. behavioural flexibility), which involve fast and transient changes between behavioural states, are explained more parsimoniously by biochemical switching of existing neural networks by different neuromodulatory molecules (i.e. neuropeptides, monoamines and hormones) that are released in a non-synaptic fashion and interact with receptors at multiple sites in the neural network (i.e. diffuse action), altering its functional properties by promoting either excitatory or inhibitory states (Oliveira, 2009; Zupanc and Lamprecht, 2000) (Table 1).

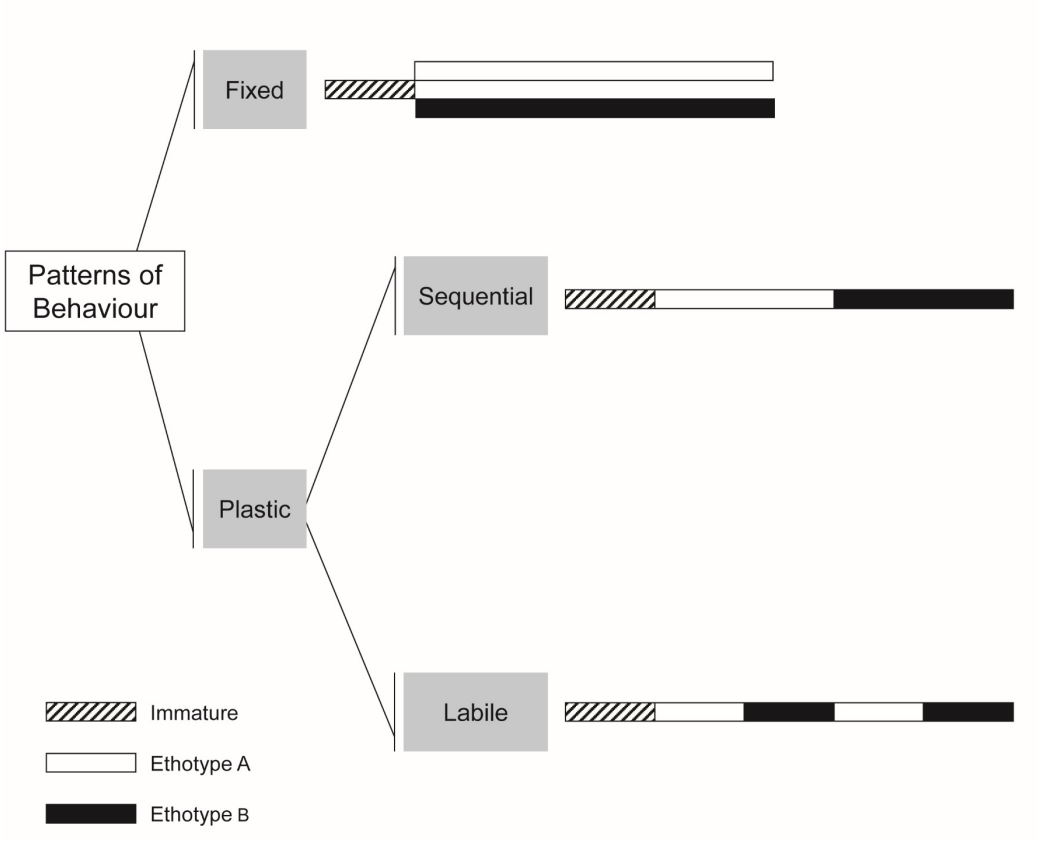


Fig. 2 – Schematic representation of different behavioural plasticity mechanisms. Based on two parameters of individual variation in behaviour, reversibility and occurrence within versus between individuals, three types of social plasticity emerge. When the behaviour is fixed and occurs between individuals, for example fixed alternative phenotypes or developmental plasticity, it depends on early social environment that directs the individual towards alternative and exclusive developmental pathways. When this variation occurs within the same individual, it can be sequential, with changes in behaviour between different life stages (e.g. juvenile versus adult behavioural states), or reversible, which encompass short-lived behavioural states within the same life-history stage, and depends on an interaction between the social environment and the internal state of the individual.

Similarly, at the physiological level, two types of effects of neuromodulators and hormones have been characterized and are expected to be involved in different patterns of plasticity depending on the time scale of their expression. Organizational effects occur early in development, typically within a critical or sensitive period during which the exposure to active molecules induces a long-lasting and irreversible differentiation of a behavioural state [e.g. masculinization of sexual behaviour by early exposure to androgens (Phoenix et al., 1959)]. In contrast, activational effects typically occur in fledged animals and are reversible and short-lived [e.g. triggering of sexual behaviour by administration of androgens to adult males (Moore et al., 1998; Phoenix et al., 1959)]. Thus, the integration of physiological parameters on social plasticity decisions is expected to be mediated by activational effects in the case of behavioural flexibility and by organizational or re-organizational effects in the case of fixed alternative phenotypes and sequential plasticity, respectively (Table 1).

Finally, at the genomic level, long-lasting and irreversible changes in behavioural states are expected to rely on epigenetic modifications (e.g. DNA methylation, histone modifications) of genes involved in social behaviour (e.g. oxytocin, vasopressin) or neural plasticity (e.g. the brain-derived neurotrophic factor gene *bdnf*) (Champagne and Curley, 2005; Curley et al., 2011; Szyf et al., 2008), whereas short-term and transient plasticity can be achieved by different neuronal activity-dependent mechanisms that change the neurogenomic state of the brain in response to perceived social stimuli (Aubin-Horth and Renn, 2009; Wolf and Linden, 2012) (Table 1). Therefore, irrespective of its temporal patterns, social plasticity relies on both temporal and spatial changes in gene regulation in the social decision-making network in the brain.

In this paper, we will review the available literature that supports this conceptual framework, with a particular focus on the neurogenomic mechanism of social plasticity. In particular, the following specific questions will be addressed. (1) What is the evidence that specific behavioural states are paralleled by specific neurogenomic states at different time scales? (2) Can the shifts between stable behavioural states also be characterized by specific but transient neurogenomic states? (3) What are the candidate neuromolecular mechanisms that mediate socially driven shifts between behavioural states? (4) Are epigenetic mechanisms only associated with irreversible plasticity patterns or do they also contribute to transient patterns of plasticity?

Table 1 – Physiological, neural and genomic mechanisms underlying different patterns of social plasticity.

	Fixed alternative phenotypes	Sequential (developmental) plasticity	Behavioural flexibility
Occurrence of alternative behavioural states within the same individual	No	Yes	Yes
Reversibility of alternative behavioural states	No	No	Yes
Time scale of alternative behavioural states	Long lasting	Long lasting	Short lived
Physiological mechanism	Organizational	(re)organizational	Activational
Neural mechanism	Structural plasticity	Structural plasticity	Biochemical switching
Genomic mechanism	Epigenetic	Epigenetic	Transient changes in gene expression

I.3. Genomic correlates of behavioural states

Several studies have demonstrated that different behavioural states are associated with different profiles of gene expression in the brain. In the genomic era, advances in technology have enabled us to identify gene modules [sets of co-regulated genes or proteins (Segal et al., 2004)] that reveal a unique gene expression pattern that reflects the biological phenotype of an individual. In this section, we present representative examples of associations between behavioural and neurogenomic states for the different patterns of social plasticity identified in the previous section.

I.3.1. Fixed alternative phenotypes

Atlantic salmon (*Salmo salar*) have a complex life cycle composed of an initial phase of birth and growth in freshwater, followed by migration to a seawater habitat where substantial body growth is achieved, before homing to the birthplace as a large fighting male to reproduce (Aubin-Horth et al., 2009; Fleming, 1998; Verspoor et al., 2007). These large males co-exist with other smaller males, known as mature male parr, that remain during their whole development in freshwater, where they mature and reproduce using an alternative mating tactic of ‘sneaking’ into the nest of migrating females. The development into one or other of these two male phenotypes is plastic and depends on size achieved and energy reserves accumulated during a critical period in the spring before autumn reproduction (Aubin-Horth and Dodson, 2004; Hutchings and Myers,

1994; Thorpe et al., 1998). Thus, depending on environment and internal conditions, any male can develop into one of these two irreversible phenotypes characterized by specific behavioural states: fighting male versus sneaker male. In order to study the molecular basis of this plastic trait, Aubin-Horth and colleagues (Aubin-Horth et al., 2005a) compared males of the same age (sneaker and immature males that will eventually become large fighting males) in a genome-wide approach. The microarray analysis revealed that 15% of the genes examined vary in expression between the two male types. Many of these differentially expressed genes are involved in processes such as growth, reproduction and neural plasticity. Genes related to cognition (learning and memory) and reproduction were upregulated in sneaker males, while genes related to cellular growth were upregulated in immature males (Aubin-Horth et al., 2005a, 2005b). Interestingly, even within a life history, for instance migrating males, differences were found between early and late migrants, indicating different genomic signatures at different life stages (Aubin-Horth and Renn, 2009).

I.3.2. Sequential (developmental) plasticity

A well-characterized example of developmental plasticity is provided by the distinct life stages and different behavioural tasks displayed by honey bees (*Apis mellifera*). During their development, bees assume different roles in their colony: (1) soon after eclosion, bees assume brood care functions (nursing); (2) after a week, they assume new roles, such as storing and processing food (e.g. turning nectar into honey); and (3) around 3 weeks of age, most bees begin foraging for pollen and nectar (Ben-Shahar, 2005; Robinson and Ben-Shahar, 2002; Whitfield et al., 2006, 2003). These different behavioural states are characterized by different profiles of gene expression in the bee brain. More than 85% of ~5500 analysed genes showed differences in expression associated with the transition from nurse to forager that are largely independent of natural age-related changes (Whitfield et al., 2006). Whitfield et al. (Whitfield et al., 2006, 2003) also showed that individual brain expression patterns are so dramatically different between life stages that they can be used to classify an individual honey bee as a nurse or as a forager with a very high accuracy rate. Like honey bees, fire ants live in colonies with thousands of workers but instead of having a single queen, fire ant colonies can have one or more. This tendency to have either one or more queens has a genetic basis and appears to be under the control of a single gene, general protein-9 (*Gp-9*). This genetic

factor determines whether workers tolerate a single fertile queen (monogyne social form, BB) or multiple queens (polygyne social form, Bb) in their colony (Wang et al., 2008). BB workers will only accept a single BB queen, and Bb workers will accept multiple Bb queens. BB workers become tolerant of multiple Bb queens when they are in colonies containing mostly Bb workers because they take on a Bb gene expression profile, showing that the BB genotype is more strongly affected by colony genotype (i.e. environment) than by their own genotype. In contrast, Bb workers do not change queen tolerance when they are in colonies containing mostly BB workers (Robinson et al., 2008). Another study on gene expression profiles between different castes of two fire ants species (*Solenopsis invicta* and *S. richteri*) revealed that genomic profiles are mostly influenced by developmental stage that exhibits a specific behavioural state than by caste membership, sex or species identity (Ometto et al., 2011). Between-species comparisons showed that workers have a considerable number of genes that are specifically upregulated or downregulated compared with males and queens. Moreover, workers also have more genes that are differentially expressed between species than do the other castes. Thus, much of the evolution of gene expression in ants may have occurred in the worker caste despite the fact that these individuals are largely or completely sterile. This can be explained by a combination of factors, including the fact that adult workers experience the most diverse environments and exhibit the broadest behavioural repertoires, and both queens and males have lost ancestral Hymenopteran feeding and self-maintenance (Ometto et al., 2011).

I.3.3. Behavioural flexibility

In the African cichlid fish *Astatotilapia burtoni*, males have evolved two distinct phenotypes: dominant males, which are brightly coloured, defend territories and actively court and spawn with females, and subordinate males, which have dull coloration similar to females, do not hold territories and are reproductively suppressed (Fernald and Hirata, 1977). These behavioural and phenotypic differences are reversible, and males change social status many times during their life depending on social context. Renn et al. (2008) examined whole-brain gene expression in dominant and subordinate males in stable hierarchies as well as in brooding females, and identified 171 genes that were differentially expressed between the two male types. Different expression profiles were also found between these male morphs in the sex steroid hormone receptors, where

dominant males had higher mRNA expression levels of androgen receptor alpha and beta (*ARα* and *ARβ*), and oestrogen receptor beta 1 and 2 (*ERβ1* and *ERβ2*), but not of oestrogen receptor alpha (*ERα*), compared with subordinate males (Burmeister et al., 2007).

I.4. Genomic correlates of behavioural transitions

Most examples of associations between neurogenomic states and social plasticity available in the literature have measured gene expression in stable behavioural states, rather than during phases of transition between behavioural states; hence, they do not reveal whether the observed differential gene expression is the mechanism behind the phenotypic change or whether it is only a consequence of the new phenotype (Aubin-Horth and Renn, 2009). However, a few recent studies on shifts between behavioural states suggest that major changes in gene expression are also involved in these transitions.

In the three-spined stickleback, *Gasterosteus aculeatus*, shifts between aggressive and non-aggressive states can be observed. During the breeding season, males defend nesting territories, and are especially aggressive towards other intruder males (Wootton, 1976). A resident–intruder paradigm revealed differential gene expression of hundreds of genes in the brain between males that were confronted by an intruder compared with those that were not (Sanogo et al., 2012). Four areas of the brain were analysed (telencephalon, diencephalon, cerebellum and brainstem) and the greatest number of differentially expressed genes was found in the diencephalon and cerebellum, and very few genes were found in the telencephalon, an unexpected result as most of the nuclei of the social decision-making network are located in the forebrain. These differences were region specific, and each brain region presented a distinct genomic response. For example, a set of genes that was upregulated in the diencephalon was downregulated in the cerebellum and in the brainstem. A *cis*-regulatory network analysis also identified transcription factors that consistently regulate genes in all brain regions and others that can upregulate or downregulate gene expression across brain regions (Sanogo et al., 2012).

In honey bees, environmental cues also trigger shifts between behavioural states. In the colony, the queen regulates many aspects of colony organization including the reproductive state of workers. This regulation is mainly done by a pheromone produced in the mandibular gland that inhibits ovary activation (Hoover et al., 2003; Le Conte and Hefetz, 2008; Slessor et al., 2005). However, this phenotype is reversible and dependent

on environmental cues. In queenless groups, worker bees develop ovaries and produce queen pheromone, but when introduced into a queenright group, there is a regression in ovary development and pheromone production (Malka et al., 2007). Genomic analysis of the mandibular gland in these two groups (queenright and queenless) identified 204 differentially expressed transcripts associated with protein catabolism and transport. These genes are likely candidates for regulating either social behaviour in queenright bees or reproductive competition behaviour in queenless workers (Malka et al., 2014).

Together, these studies illustrate that neurogenomic states are not only associated with behavioural states but also characteristic of phases of transition between states driven by social cues.

I.5. Shifting mechanisms

Following the conceptual framework presented above, transitions between behavioural states require shifts between their underlying neurogenomic states in response to relevant environmental cues perceived by the animal. At least three different neuronal activity-dependent molecular mechanisms can be proposed to translate the social information into a neurogenomic shift (Wolf and Linden, 2012) (Fig. 3).

A first mechanism consists of the activation (e.g. phosphorylation) of pre-existing proteins (e.g. phosphorylation of cAMP response element-binding protein, CREB to pCREB) that subsequently either act as transcription factors for immediate early genes (IEGs) or delayed response genes, or regulate intracellular signalling pathways [e.g. mitogen-activated protein kinase (MAPK) pathway]. An example of this activation comes from California mice (*Peromyscus californicus*). In this species, male aggressive behaviour increases in animals that are under short day photoperiods compared with animals in long day photoperiods (Trainor et al., 2008). A resident intruder test revealed that a phosphorylation mechanism of ERK1 and 2 (extracellular signal-regulated kinases 1 and 2) and CREB underlies these differences. A significant increase in phosphorylated ERK (pERK) expression in several brain regions known to regulate aggressive behaviour is induced by aggression tests under short days but not under long days. However, a very different pattern is observed with pCREB immunostaining, where aggressive behaviour decreased the number of pCREB-positive cells when mice were housed under long days but not short days. Together, this data set suggests that different phosphorylation

pathways are associated with the response to different environmental conditions (Trainor et al., 2010).

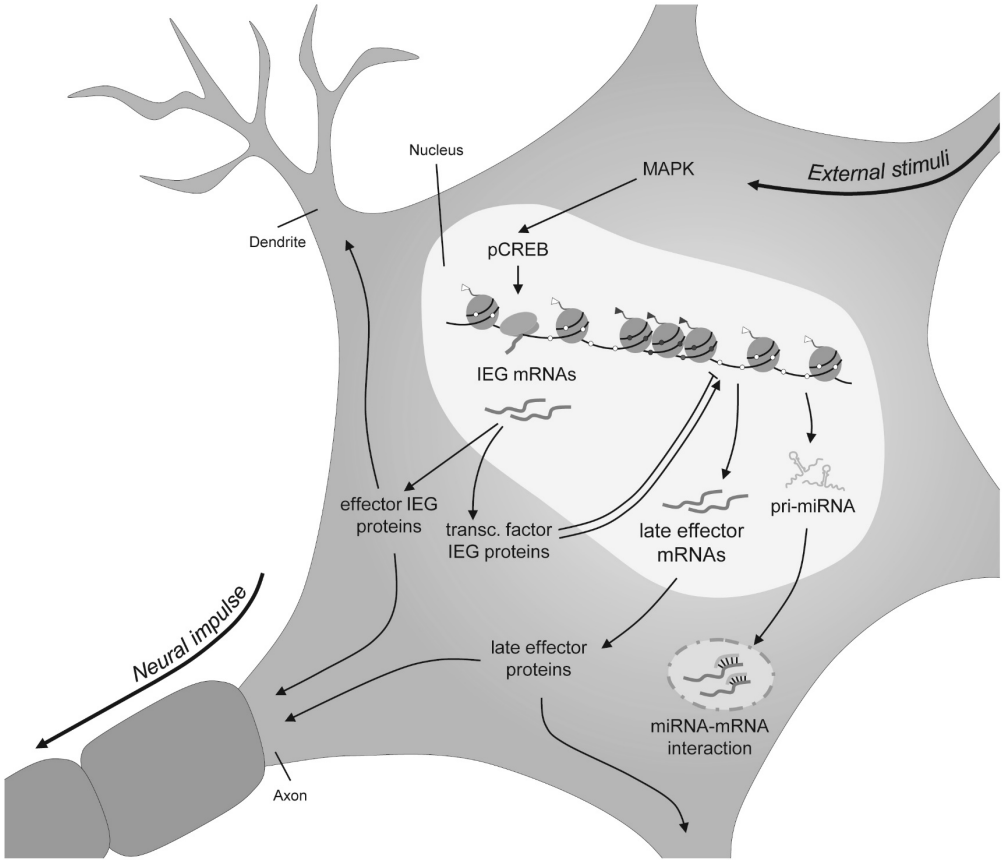


Fig. 3 – Schematic representation of neurogenomic shifting mechanisms. Social information (i.e. external stimuli) can trigger a neurogenomic shift through: (i) activation of pre-existing proteins (e.g. phosphorylation of cAMP response element-binding protein, pCREB) or intracellular signalling pathways (e.g. mitogen-activated protein kinases, MAPKs); (ii) immediate early gene expression (IEG) that will act as transcription factors for other genes (i.e. late effector genes) or as direct effector proteins; and/or (iii) transcription of microRNA genes (primary-microRNA, pri-miRNA) that will be processed into single-stranded mature microRNAs (miRNAs) and regulate mRNA (mRNA) transport, translation and degradation inside the cell. Modifications introduced into the nucleosome structures and the DNA strand (i.e. methylation) together provide different states of chromatin compaction, and therefore activate or inactivate gene expression.

A second mechanism is based on neuronal activity-dependent transcription factors that mediate IEG expression inside the nucleus. Activation of IEGs corresponds to the first genomic response given within minutes of a stimulus. Recently, Saha et al. (2011) were able to categorize IEGs into two groups depending on their expression time. For rapid IEGs, which are expressed within a few minutes, DNA polymerase II (Pol II) stalling was seen in the promoter region of these genes, whereas delayed IEGs, which are

expressed within an hour of the stimulus, largely lacked this poised Pol II (Saha et al., 2011; Saha and Dudek, 2013). This mechanism of stalling was shown to be pertinent in regulating the timing and dynamics of gene responses, as mRNA accumulation was not affected (Saha et al., 2011). Depending on their function, IEG proteins can act themselves as transcription factors (e.g. v-fos FBJ murine osteosarcoma viral oncogene homologue gene *c-fos* and early growth response protein 1 gene *egr-1*), returning to the nucleus to regulate the expression of specific genes, or as direct effector proteins (e.g. activity-regulated cytoskeleton-associated protein gene *arc* and homer homologue 1a gene *homer1a*), regulating synaptic function (Clayton, 2000). Two examples will be presented that illustrate the role of IEGs in the neurogenomic translation of social information. In *A. burtoni*, cues from the social context are crucial for males to switch between the subordinate and the dominant phenotype. Subordinate males can be induced to become dominants by removing all competing territorial males in their tank (social ascending paradigm). Conversely, dominant males can also be induced to become subordinate by exposing them to larger dominant males (social descending paradigm) (Francis et al., 1993; White et al., 2002). At the molecular level, these behavioural shifts are paralleled by changes in the expression of the IEGs. In socially ascending males, mRNA levels of the IEGs *c-fos* and *egr-1* increased in all nuclei of the social decision-making network within the first hour of becoming dominant. Importantly, this increased expression of IEGs was not found in either stable dominant or stable subordinate males (Burmeister et al., 2005; Maruska et al., 2013b). In contrast, in socially descending males, changes in IEG expression levels were nuclei specific both for *c-fos* and for *egr-1* but never simultaneously for both (Maruska et al., 2013a). Thus, depending on the direction of the behavioural transition, different IEG activation patterns can be observed in the social decision-making network, suggesting a complex regulatory system that translates the animal's perception of the social environment into transcriptional control of late response genes necessary for adaptive phenotypic changes. A second example of the role of IEGs in neurogenomic shifts is provided by studies in songbirds. In zebra finches (*Taeniopygia guttata*), hearing a conspecific song elicits the expression of the IEG *zenk* (aka *egr-1*) in the zebra finch auditory forebrain (Cheng and Clayton, 2004). This song-induced gene expression relies on the rapid phosphorylation of ERK, which varies depending on the novelty of the song: novel songs trigger a sharp increase in ERK phosphorylation, peaking within 1–2 min of song onset, required for *zenk* expression, whereas a familiar song leads to a persistent habituation for that specific song (Cheng and Clayton, 2004; Dong and

Clayton, 2009). Also dependent on the ERK/MAPK pathway is the expression of *arc*, an effector IEG that is co-localized with *zenk* and is thought to be necessary for dendritic spine outgrowth in postsynaptic signalling (Bramham et al., 2010; Velho et al., 2005). Thus, depending on the individuals' perception of the same song as either novel or familiar, IEGs can be activated with different fold changes mediating transitions between behavioural states in song learning (i.e. naive versus experienced). A third possible mechanism consists on the transcription of microRNAs (miRNAs) that regulate mRNA transport, translation and degradation for transcription factors or synaptic proteins. An example of this mechanism is the brain-expressed miR-133, recently found to play an important role in controlling behavioural aggregation in migratory locusts (*Locusta migratoria*) (Yang et al., 2014). In this species, considered a worldwide pest, it is possible to observe two main behavioural states, the solitary phase and the gregarious phase, which are density dependent and reversible (Kang et al., 2004). When present, miR-133 suppresses the expression of two genes involved in the dopamine pathway, *henna* and *pale*, thereby regulating the dopamine synthesis important for the phase transition of locusts in response to population density stress (Yang et al., 2014). It has been shown that decreasing dopamine production through miR-133 agomir resulted in the behavioural shift from the gregarious to the solitary phase, while miR-133 inhibition promoted gregarious-like behaviour of solitary locusts. Thus, miRNA plays an important role as an activational switch in this species acting a key mediator of a transition between behavioural states.

I.6. Epigenetics of social plasticity

Accumulating evidence has shown that epigenetic marks play a key role in maintaining behavioural states by keeping different neurogenomic states in the brain. Generally, epigenetic mechanisms operate either on the DNA sequence, via DNA methylation (Miranda and Jones, 2007) and/or binding of non-coding RNA (ncRNA) (Bernstein and Allis, 2005), or on chromatin, mainly via post-translational modification of histone proteins (Borrelli et al., 2008). Within cells, these mechanisms can interact with each other to inhibit or activate gene expression not just as an on and off switch but rather in a gradient fashion, determining which proteins are transcribed in which environmental contexts. These epigenetic changes have a significant role in controlling functional and structural molecular states, therefore enabling adaptive cellular expression

patterns during development and differentiation (i.e. stable modifications), or plastic changes in adult organisms (i.e. plastic modifications). In this section, we present examples where these epigenetic mechanisms were found to maintain behavioural states or to generate behavioural flexibility, not only during development but also in the adult stage.

I.6.1. DNA methylation

DNA methylation, which is normally associated with the inhibition of gene expression, occurs mainly at cytosine bases preceding guanines (CpG sites) that are converted to 5′-methylcystosine (5-mC) by DNA methyltransferase (DNMT) enzymes. These enzymes have two main enzymatic activities, maintenance methylation (DNMT1) necessary to preserve DNA methylation after every DNA replication cycle, and *de novo* methylation (DNMT3A/B), setting up DNA methylation patterns early in development (Bestor, 2000; Okano et al., 1999). In honey bees, a non-reversible segregation of behaviours between queens and workers can be observed during early life (see above) and is dependent on the royal jelly diet fed during larval development. In addition to the influence of nutrition on the epigenetic status of the queen's cells (for a review, see (Buttstedt et al., 2014)), Kucharski and colleagues (Kucharski et al., 2008) showed that silencing the expression of *Dnmt3* in newly hatched larvae induced a royal jelly-like effect on the larval development trajectory, suggesting a role for DNA methylation in storing epigenetic information in a context-dependent manner (i.e. dependent on nutrition).

Recent findings on the mechanism behind DNA demethylation have promoted a shift in our understanding of how changes in DNA methylation are coupled to cell differentiation, development and disease. Hydroxymethylation of 5-mC to 5-hmC (hydroxymethylcytosine) by TET (ten–eleven translocation) enzymes has been proposed as the mechanism behind methyl group removal from cytosine bases followed by DNA repair machinery (Pastor et al., 2013; Tahiliani et al., 2009). Along this line, Herb and colleagues (Herb et al., 2012) were able to show that a reversible behaviour in honey bees, switching back from foragers to nurses, is also accompanied by changes in the DNA methylation pattern. These authors found that otherwise genetically identical nurse and age-matched forager bees have different methylation patterns in 155 genes found in their brain cells. However, after inducing the role reversal, the methylation patterns in 57 genes of a total of 107 differentially methylated regions (DMRs) from the former foragers

changed to a nurse pattern (Herb et al., 2012). This finding suggests a subcaste-specific methylation signature that can assist in forming worker phenotypes. Also in adult male rat brains, Auger and colleagues (Auger et al., 2011) found that methylation patterns on some steroid-responsive genes were actively maintained by the presence of circulating steroid hormones (i.e. signal) that affect the expression of the socially relevant peptide vasopressin (AVP) and ERα within the bed nucleus of the stria terminalis (BNST). Castration (i.e. loss of signal) of these males dramatically reduced the expression of AVP while increasing ERα expression as a result of a shift in methylation state on the promoters of these genes, which could be prevented by testosterone replacement (i.e. signal restoration) (Auger et al., 2011). It would be interesting to see whether this mechanism is involved in seasonal reproduction in males or even in contexts of social dominance, where the social decision-making network is highly regulated by these hormones.

I.6.2. Histone modifications

Nucleosomes, around which DNA is wound, are structures composed of two copies of each of the core histones H2A, H2B, H3 and H4, and stabilized into high-order structures by the linker histone H1 forming the chromatin (Luger et al., 1997). The N-terminal tails of histones are exposed on the nucleosomal surface and are the target of numerous post-translational modifications (Borrelli et al., 2008; Keverne and Curley, 2008; Strahl and Allis, 2000). This provides a dynamic functional continuum between two states of chromatin compaction, which correlate primarily with active (euchromatin) and inactive (heterochromatin) states of gene expression (Berger, 2007).

Histone acetylation and phosphorylation are usually associated with active chromatin, because of the neutralization of the positive charge of the histone tail, playing an important role in integrating incoming signals (Barth and Imhof, 2010; Borrelli et al., 2008; Keverne and Curley, 2008). In prairie voles, *Microtus ochrogaster*, a socially monogamous species, pair-bond formation involves a selective affiliation of females to a partner (partner preference), which is induced by mating (Getz and Hofmann, 1986; Williams et al., 1992). Recently, the involvement of histone modifications in the regulation of pair bonding has been established (Wang et al., 2013). Female prairie voles treated with histone deacetylase inhibitors (HDACi) became bonded to their mates in the absence of a mating event. This was accompanied by a specific upregulation of oxytocin

receptor (*oxtr*) and vasopressin V1a receptor (*avpr1a*) in the nucleus accumbens (NAcc), which was associated with an increase in histone acetylation at their respective promoters, similar to that observed when untreated females were mated and formed a pair bond (Wang et al., 2013). However, HDACi treatment did not promote vasopressin and oxytocin receptor expression in the NAcc of female prairie voles that were not exposed to males, indicating that other factors related to social context are required to induce pair bonds. Also, in honey bees, further investigation of the royal jelly nutrition has shown that one of its major components, the fatty acid (*E*)-10-hydroxy-2-deconic acid (10-HDA), has HDACi activity and has the capacity to reactivate the expression of epigenetically silenced genes in mammalian cells (Spannhoff et al., 2011). The examples given here on the epigenetic effects of nutrition on honey bee larvae developmental dichotomy encompass some of the mechanisms responsible for the unique brain epigenomes detected by Lyko and colleagues (Lyko et al., 2010), where over 550 genes showed methylation differences between queen and worker behavioural phenotypes.

Histone methylation does not alter the positive charge of the amino acids but can be a marker for both active (e.g. trimethylation at lysine-4 residues of histone 3, H3K4me3) and inactive regions of chromatin (Keverne and Curley, 2008). Additionally, histone methylation has a lower turnover rate, which could facilitate the stabilization of gene expression in the absence of incoming signals (Barth and Imhof, 2010; Borrelli et al., 2008). We mentioned above the key role of poised Pol II in mediating rapid IEG expression. This mechanism of stalling is itself an example of an epigenetic mechanism, because it retains permissive epigenetic marks on promoters, making the chromatin accessible to polymerases and therefore allowing a robust transcription. In addition to Pol II having its largest subunit (RPB1) phosphorylated on serine-5 residues of the C-terminus, the promoter region is enriched with histones with high levels of H3K4me3 methylation (Saha et al., 2011; Telese et al., 2013).

I.6.3. ncRNA

ncRNA genes produce several classes of functional RNA molecules that are highly integrated into the molecular circuitry, making them important players in the regulation of gene expression through various mechanisms, such as RNA interference and gene silencing (Bernstein and Allis, 2005; Qureshi and Mehler, 2012). One such example of ncRNAs is miRNA, small single stranded molecules of ~22 nucleotides that modify gene

activity through post-transcriptional regulation of their mRNA targets. We mentioned this class of ncRNA above as a shifting mechanism in organisms, giving the example of the migratory locusts where miR-133 mediated the transition between the gregarious and solitary phases (Yang et al., 2014). They have also been recognized as having a role in the division of labour among honey bees. Greenberg and colleagues (Greenberg et al., 2012) compared the brain miRNA transcriptome of adult workers and found that miRNA expression was dependent upon social context, as several miRNAs were downregulated in honey bees that were nurses relative to foragers only when they were in colonies that contained foragers.

From integrative genomic studies elucidating the complex wiring of miRNA regulatory relationships, some preliminary conclusions can be drawn: miRNAs can act cooperatively or redundantly (Su et al., 2011), and miRNAs from one cluster tend to be involved in the same network module with a direct or indirect regulatory coordination seen in a dependent manner of miRNA cluster composition (i.e. same versus various families, respectively), suggesting distinct roles in biological processes (Wang et al., 2011). As for other transcribed RNAs, miRNA expression is itself regulated by epigenetic factors such as DNA methylation and chromatin structure (Iorio et al., 2010). An example of this interplay can be seen in the regulation of neural stem cells (NSCs) in adult neurogenesis (Shi et al., 2010; Szulwach et al., 2010). DNA methylation can also promote the biding of methyl-CpG-binding domain proteins (MBDs, e.g. MeCP2) (Hendrich and Bird, 1998) that can in turn recruit additional proteins to the locus, such as post-translational modification of histones and as well as the recruitment of Dnmt1. In rats, MeCP2 can epigenetically regulate specific miRNAs in adult NSCs such as miR-137, so that: (1) in the absence of MeCP2, miR-137 is overexpressed and promotes the proliferation of adult NSCs; and (2) in the presence of MeCP2 together with Sox2 (SRY-box containing gene 2 protein), a core transcription factor regulating stem cell self-renewal (Zappone et al., 2000), miR-137 expression is regulated, enhancing instead adult NSC differentiation (Szulwach et al., 2010). This structural reorganization (i.e. neurogenesis) is another path through which social plasticity in the neural decision-making network can be achieved.

Finally, it should be said that these three epigenetic mechanisms are expected to work together in the regulation of social plasticity, and therefore their coexistence should be observed. This can be illustrated by the transgenerational effects of maternal nurturing in rodents, which have long been the classic example of behavioural epigenetics (for review,

see (Curley et al., 2011)). In this behavioural paradigm, female offspring of mothers that showed high levels of licking/grooming will also display high levels of licking/grooming when they become mothers later in life. High licking/grooming offspring were found to have fewer methyl groups attached to the glucocorticoid receptor promoter of exon 17 and histone acetylation at the lysine-9 (K9) residue of H3 (H3K9) when compared with offspring of low licking/grooming mothers, showing therefore an increased hippocampal glucocorticoid receptor expression (Weaver et al., 2004). Thus, multiple epigenetic mechanisms act together to allow the long-term maintenance of gene expression differences (e.g. epigenetic marks) depending on early social environment.

I.7. Prospects

In this review we have shown how highly responsive the brain neurogenomic states are to the environment and at storing social information, channelling in this way transient as well as long-lasting adaptations. However, to date most studies on the relationship between neurogenomic and behavioural states have either used whole-brain gene expression data or restricted their analyses to a few areas of interest in the brain. Given the distributed nature of the putative social decision-making network in the brain, a future challenge will be to characterize the neurogenomic states across this network by studying simultaneously the patterns of gene expression of the multiple network nodes. This would extend the characterization of neurogenomics to a spatial level that might be quite relevant to the understanding of the diversity of potential behavioural states produced by the relevant brain networks.

In future research at this level of genome-behaviour analysis, it will also be necessary to distinguish between neurogenomic mechanisms that lead to a specific behavioural state and those that are maintaining that same state. As discussed above, different patterns of gene expression might be associated with stable behavioural states and others with behavioural shifts, and the neuromolecular mechanisms expected to underlie states and shifts are not the same. With the development of new sequencing technologies that are becoming available to behavioural ecologists, such as RNA-Seq for transcriptome analysis, Chip-Seq for specific activity-regulated genes and Bisulfite-Seq for methylome analysis, these hypotheses are becoming testable in the short-term.

I.8. Acknowledgements

R.F.O. would like to thank Hans Hoppeler and Catharine Rankin for the invitation to the JEB symposium on 'Epigenetics and Comparative Physiology' and Michaela Handel for all her assistance during the workshop.

I.9. Competing interests

The authors declare no competing financial interests.

I.10. Author contributions

All authors planned, wrote and revised the article.

I.11. Funding

S.D.C. (SFRH/BD/89072/2012) and M.C.T. (SFRH/BD/44848/2008) were supported by PhD fellowships from Fundação para a Ciência e a Tecnologia (FCT), Portugal. The writing of this review was supported by an FCT grant of excellence to R.F.O. (EXCL/BIA-ANM/0549/2012).

I.12. References

Almada V.C., Gonçalves E.J., Santos A.J. & Baptista C., 1994. Breeding ecology and nest aggregations in a population of Salaria pavo (Pisces: Blenniidae) in an area where nest sites are very scarce. *Journal of Fish Biology*, *45*, 819–830.

Almeida O., Gonçalves-de-Freitas E., Lopes J.S. & Oliveira R.F., 2014. Social instability promotes hormone-behavior associated patterns in a cichlid fish. *Hormones and Behavior*, *66*, 369–382.

Aubin-Horth N. & Dodson J.J., 2004. Influence of individual body size and variable thresholds on the incidence of a sneaker male reproductive tactic in Atlantic salmon. *Evolution*, *58*, 136–144.

Aubin-Horth N., Landry C.R., Letcher B.H. & Hofmann H.A., 2005a. Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 1655–1662.

Aubin-Horth N., Letcher B.H. & Hofmann H.A., 2009. Gene-expression signatures of Atlantic salmon's plastic life cycle. *General and Comparative Endocrinology*, *163*, 278–284.

Aubin-Horth N., Letcher B.H. & Hofmann H.A., 2005b. Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon.

Journal of Heredity, *96*, 261–278.

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

Auger C.J., Coss D., Auger A.P. & Forbes-Lorman R.M., 2011. Epigenetic control of vasopressin expression is maintained by steroid hormones in the adult male rat brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 4242–4247.

Barth T.K. & Imhof A., 2010. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences*, *35*, 618–626.

Ben-Shahar Y., 2005. The foraging gene, behavioral plasticity, and honeybee division of labor. *Journal of Comparative Physiology A*, *191*, 987–994.

Berger S.L., 2007. The complex language of chromatin regulation during transcription. *Nature*, *447*, 407–412.

Bernstein E. & Allis C.D., 2005. RNA meets chromatin. *Genes & Development*, *19*, 1635–1655.

Bestor T.H., 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics*, *9*, 2395–2402.

Borrelli E., Nestler E.J., Allis C.D. & Sassone-Corsi P., 2008. Decoding the epigenetic language of neuronal plasticity. *Neuron*, *60*, 961–974.

Bramham C.R., Alme M.N., Bittins M., Kuipers S.D., Nair R.R., Pai B., Panja D., Schubert M., Soule J., Tiron A. & Wibrand K., 2010. The Arc of synaptic memory. *Experimental Brain Research*, *200*, 125–140.

Burmeister S.S., Jarvis E.D. & Fernald R.D., 2005. Rapid behavioral and genomic responses to social opportunity. *PLoS Biology*, *3*, e363.

Burmeister S.S., Kailasanath V. & Fernald R.D., 2007. Social dominance regulates androgen and estrogen receptor gene expression. *Hormones and Behavior*, *51*, 164–170.

Buttstedt A., Moritz R.F.A. & Erler S., 2014. Origin and function of the major royal jelly proteins of the honeybee (Apis mellifera) as members of the yellow gene family. *Biological reviews of the Cambridge Philosophical Society*, *89*, 255–269.

Champagne F.A. & Curley J.P., 2005. How social experiences influence the brain. *Current Opinion in Neurobiology*, *15*, 704–709.

Cheng H.-Y. & Clayton D.F., 2004. Activation and habituation of extracellular signal-regulated kinase phosphorylation in zebra finch auditory forebrain during song presentation. *Journal of Neuroscience*, *24*, 7503–7513.

Clayton D.F., 2000. The genomic action potential. *Neurobiology of Learning and Memory*, *74*, 185–216.

Curley J.P., Jensen C.L., Mashoodh R. & Champagne F.A., 2011. Social influences on neurobiology and behavior: epigenetic effects during development.

Psychoneuroendocrinology, *36*, 352–371.

Dong S. & Clayton D.F., 2009. Habituation in songbirds. *Neurobiology of Learning and Memory*, *92*, 183–188.

Fagundes T., Simões M.G., Saraiva J.L., Ros A.F.H., Gonçalves D. & Oliveira R.F., 2015. Birth date predicts alternative life history pathways in a fish with sequential reproductive tactics. *Functional Ecology*, *29*, 1533–1542.

Fernald R.D. & Hirata N.R., 1977. Field study of Haplochromis burtoni: habitats and co-habitant. *Environmental Biology of Fishes*, *2*, 299–308.

Fleming I.A., 1998. Pattern and variability in the breeding system of Atlantic salmon (Salmo salar), with comparisons to other salmonids. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*, 59–76.

Francis R.C., Soma K. & Fernald R.D., 1993. Social regulation of the brain-pituitary-gonadal axis. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 7794–7798.

Frank S.A., 2011. Natural selection. II. Developmental variability and evolutionary rate. *Journal of Evolutionary Biology*, *24*, 2310–2320.

Getz L.L. & Hofmann J.E., 1986. Social organization in free-living prairie voles, Microtus ochrogaster. *Behavioral Ecology and Sociobiology*, *18*, 275–282.

Gonçalves E.J., Almada V.C., Oliveira R.F. & Santos A.J., 1996. Female mimicry as a mating tactic in males of the blenniid fish Salaria pavo. *Journal of the Marine Biological Association of the United Kingdom*, *76*, 529–538.

Goodson J.L., 2005. The vertebrate social behavior network: evolutionary themes and variations. *Hormones and Behavior*, *48*, 11–22.

Greenberg J.K., Xia J., Zhou X., Thatcher S.R., Gu X., Ament S.A., Newman T.C., Green P.J., Zhang W., Robinson G.E. & Ben-Shahar Y., 2012. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes, Brain and Behavior*, *11*, 660–670.

Hendrich B. & Bird A., 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology*, *18*, 6538–6547.

Herb B.R., Wolschin F., Hansen K.D., Aryee M.J., Langmead B., Irizarry R., Amdam G. V & Feinberg A.P., 2012. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, *15*, 1371–1373.

Hoover S.E.R., Keeling C.I., Winston M.L. & Slessor K.N., 2003. The effect of queen pheromones on worker honey bee ovary development. *Die Naturwissenschaften*, *90*, 477–480.

Hutchings J.A. & Myers R.A., 1994. The evolution of alternative mating strategies in variable environments. *Evolutionary Ecology*, *8*, 256–268.

Iorio M. V, Piovan C. & Croce C.M., 2010. Interplay between microRNAs and the epigenetic machinery: an intricate network. *Biochimica et Biophysica Acta*, *1799*,

694–701.

Kang L., Chen X., Zhou Y., Liu B., Zheng W., Li R., Wang J. & Yu J., 2004. The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 17611–17615.

Kappeler P.M. & Kraus C., 2010. Levels and mechanisms of behavioural variability, in: Kappeler, P. (Ed.), Animal Behaviour: Evolution and Mechanisms. Springer-Verlag, Heidelberg, pp. 655–684.

Keverne E.B. & Curley J.P., 2008. Epigenetics, brain evolution and behaviour. *Frontiers in Neuroendocrinology*, *29*, 398–412.

Kucharski R., Maleszka J., Foret S. & Maleszka R., 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, *319*, 1827–1830.

Le Conte Y. & Hefetz A., 2008. Primer pheromones in social hymenoptera. *Annual Review of Entomology*, *53*, 523–542.

Luger K., Mäder A.W., Richmond R.K., Sargent D.F. & Richmond T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*, 251–260.

Lyko F., Foret S., Kucharski R., Wolf S., Falckenhayn C. & Maleszka R., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology*, *8*, e1000506.

Malka O., Niño E.L., Grozinger C.M. & Hefetz A., 2014. Genomic analysis of the interactions between social environment and social communication systems in honey bees (Apis mellifera). *Insect Biochemistry and Molecular Biology*, *47*, 36–45.

Malka O., Shnieor S., Hefetz A. & Katzav-Gozansky T., 2007. Reversible royalty in worker honeybees (Apis mellifera) under the queen influence. *Behavioral Ecology and Sociobiology*, *61*, 465–473.

Maruska K.P., Becker L., Neboori A. & Fernald R.D., 2013a. Social descent with territory loss causes rapid behavioral, endocrine and transcriptional changes in the brain. *The Journal of Experimental Biology*, *216*, 3656–3666.

Maruska K.P., Zhang A., Neboori A. & Fernald R.D., 2013b. Social opportunity causes rapid transcriptional changes in the social behaviour network of the brain in an African cichlid fish. *Journal of Neuroendocrinology*, *25*, 145–157.

Miranda T.B. & Jones P.A., 2007. DNA methylation: the nuts and bolts of repression. *Journal of Cellular Physiology*, *213*, 384–390.

Moore M.C., Hews D.K. & Knapp R., 1998. Hormonal control and evolution of alternative male phenotypes: Generalizations of models for sexual differentiation. *American Zoologist*, *38*, 133–151.

Newman S.W., 1999. The medial extended amygdala in male reproductive behavior. A node in the mammalian social behavior network. *Annals of the New York Academy of Sciences*, *877*, 242–257.

O'Connell L.A. & Hofmann H.A., 2012. Evolution of a vertebrate social decision-making network. *Science*, *336*, 1154–1157.

O'Connell L.A. & Hofmann H.A., 2011. The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. *The Journal of Comparative Neurology*, *519*, 3599–3639.

Okano M., Bell D.W., Haber D.A. & Li E., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, *99*, 247–257.

Oliveira R.F., 2009. Social behavior in context: Hormonal modulation of behavioral plasticity and social competence. *Integrative and Comparative Biology*, *49*, 423–440.

Ometto L., Shoemaker D., Ross K.G. & Keller L., 2011. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Molecular Biology and Evolution*, *28*, 1381–1392.

Pastor W.A., Aravind L. & Rao A., 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Reviews in Molecular Cell Biology*, *14*, 341–356.

Phoenix C.H., Goy R.W., Gerall A.A. & Young W.C., 1959. Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology*, *65*, 369–382.

Piersma T. & Drent J., 2003. Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology & Evolution*, *18*, 228–233.

Pigliucci M., 2001. Phenotypic Plasticity: Beyond Nature and Nurture. John Hopkins University Press, Baltimore, MD.

Qureshi I.A. & Mehler M.F., 2012. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nature Reviews in Neuroscience*, *13*, 528–541.

Renn S.C.P., Aubin-Horth N. & Hofmann H.A., 2008. Fish and chips: functional genomics of social plasticity in an African cichlid fish. *The Journal of Experimental Biology*, *211*, 3041–3056.

Robinson G.E. & Ben-Shahar Y., 2002. Social behavior and comparative genomics: new genes or new gene regulation? *Genes, Brain and Behavior*, *1*, 197–203.

Robinson G.E., Fernald R.D. & Clayton D.F., 2008. Genes and social behavior. *Science*, *322*, 896–900.

Saha R.N. & Dudek S.M., 2013. Splitting Hares and Tortoises: A classification of neuronal immediate early gene transcription based on poised RNA polymerase II. *Neuroscience*, *247*, 175–181.

Saha R.N., Wissink E.M., Bailey E.R., Zhao M., Fargo D.C., Hwang J.-Y., Daigle K.R., Fenn J.. D., Adelman K. & Dudek S.M., 2011. Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature Neuroscience*,

14, 848–856.

Sanogo Y.O., Band M., Blatti C., Sinha S. & Bell A.M., 2012. Transcriptional regulation of brain gene expression in response to a territorial intrusion. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 4929–4938.

Schradin C., 2013. Intraspecific variation in social organization by genetic variation, developmental plasticity, social flexibility or entirely extrinsic factors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*, 20120346.

Segal E., Friedman N., Koller D. & Regev A., 2004. A module map showing conditional activity of expression modules in cancer. *Nature Genetics*, *36*, 1090–1098.

Shi Y., Zhao X., Hsieh J., Wichterle H., Impey S., Banerjee S., Neveu P. & Kosik K.S., 2010. MicroRNA regulation of neural stem cells and neurogenesis. *The Journal of Neuroscience*, *30*, 14931–14936.

Sinervo B. & Lively C.M., 1996. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, *380*, 240–243.

Slessor K.N., Winston M.L. & Le Conte Y., 2005. Pheromone communication in the honeybee (Apis mellifera L.). *Journal of Chemical Ecology*, *31*, 2731–2745.

Spannhoff A., Kim Y.K., Raynal N.J.-M., Gharibyan V., Su M.-B., Zhou Y.-Y., Li J., Castellano S., Sbardella G., Issa J.-P.J. & Bedford M.T., 2011. Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO reports*, *12*, 238–243.

Strahl B.D. & Allis C.D., 2000. The language of covalent histone modifications. *Nature*, *403*, 41–45.

Su W.-L., Kleinhanz R.R. & Schadt E.E., 2011. Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. *Molecular Systems Biology*, *7*, 490.

Szulwach K.E., Li X., Smrt R.D., Li Y., Luo Y., Lin L., Santistevan N.J., Li W., Zhao X. & Jin P., 2010. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *Journal of Cell Biology*, *189*, 127–141.

Szyf M., McGowan P. & Meaney M.J., 2008. The social environment and the epigenome. *Environmental and Molecular Mutagenesis*, *49*, 46–60.

Taborsky B. & Oliveira R.F., 2012. Social competence: an evolutionary approach. *Trends in Ecology & Evolution*, *27*, 679–688.

Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L. & Rao A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*, 930–935.

Telese F., Gamliel A., Skowronska-Krawczyk D., Garcia-Bassets I. & Rosenfeld M.G., 2013. “Seq-ing” insights into the epigenetics of neuronal gene regulation. *Neuron*, *77*, 606–623.

Thorpe J.E., Mangel M., Metcalfe N.B. & Huntingford F.A., 1998. Modelling the proximate basis of salmonid life-history variation, with application to Atlantic salmon, Salmo salar L. *Evolutionary Ecology*, *12*, 581–599.

Trainor B.C., Crean K.K., Fry W.H.D. & Sweeney C., 2010. Activation of extracellular signal-regulated kinases in social behavior circuits during resident-intruder aggression tests. *Neuroscience*, *165*, 325–336.

Trainor B.C., Finy M.S. & Nelson R.J., 2008. Rapid effects of estradiol on male aggression depend on photoperiod in reproductively non-responsive mice. *Hormones and Behavior*, *53*, 192–199.

Velho T.A.F., Pinaud R., Rodrigues P. V & Mello C. V, 2005. Co-induction of activity-dependent genes in songbirds. *European Journal of Neuroscience*, *22*, 1667–1678.

Verspoor E., Stradmeyer L. & Nielsen J.L., 2007. The Atlantic Salmon: Genetics, Conservation and Management. Blackwell Publishing Ltd, Oxford.

Wang H., Duclot F., Liu Y., Wang Z. & Kabbaj M., 2013. Histone deacetylase inhibitors facilitate partner preference formation in female prairie voles. *Nature Neuroscience*, *16*, 919–924.

Wang J., Haubrock M., Cao K.-M., Hua X., Zhang C.-Y., Wingender E. & Li J., 2011. Regulatory coordination of clustered microRNAs based on microRNA-transcription factor regulatory network. *BMC Systems Biology*, *5*, 199.

Wang J., Ross K.G. & Keller L., 2008. Genome-wide expression patterns and the genetic architecture of a fundamental social trait. *PLoS Genetics*, *4*, e1000127.

Weaver I.C.G., Cervoni N., Champagne F.A., D'Alessio A.C., Sharma S., Seckl J.R., Dymov S., Szyf M. & Meaney M.J., 2004. Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*, 847–854.

West-Eberhard M.J., 2003. Developmental Plasticity and Evolution. Oxford University Press, New York.

White S.A., Nguyen T. & Fernald R.D., 2002. Social regulation of gonadotropin-releasing hormone. *Journal of Experimental Biology*, *205*, 2567–2581.

Whitfield C.W., Ben-Shahar Y., Brillet C., Leoncini I., Crauser D., Leconte Y., Rodriguez-Zas S. & Robinson G.E., 2006. Genomic dissection of behavioral maturation in the honey bee. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 16068–16075.

Whitfield C.W., Cziko A.-M. & Robinson G.E., 2003. Gene expression profiles in the brain predict behavior in individual honey bees. *Science*, *302*, 296–299.

Williams J.R., Catania K.C. & Carter C.S., 1992. Development of partner preferences in female prairie voles (Microtus ochrogaster): the role of social and sexual experience. *Hormones and Behavior*, *26*, 339–349.

Wolf C. & Linden D.E.J., 2012. Biological pathways to adaptability – interactions between genome, epigenome, nervous system and environment for adaptive behavior. *Genes, Brain and Behavior*, *11*, 3–28.

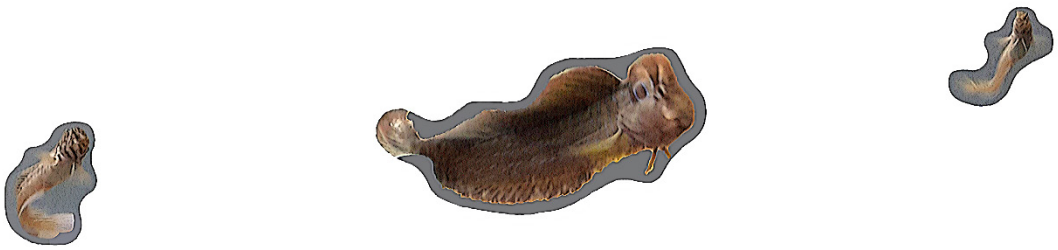
Wootton R.J., 1976. The Biology of Sticklebacks. Academic Press, London.

Yang M., Wei Y., Jiang F., Wang Y., Guo X., He J. & Kang L., 2014. MicroRNA-133 inhibits behavioral aggregation by controlling dopamine synthesis in locusts. *PLoS Genetics*, *10*, e1004206.

Zappone M. V, Galli R., Catena R., Meani N., De Biasi S., Mattei E., Tiveron C., Vescovi A.L., Lovell-Badge R., Ottolenghi S. & Nicolis S.K., 2000. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development*, *127*, 2367–2382.

Zayed A. & Robinson G.E., 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annual Review of Genetics*, *46*, 591–615.

Zupanc G.K.H. & Lamprecht J., 2000. Towards a cellular understanding of motivation: structural reorganization and biochemical switching as key mechanisms of behavioral plasticity. *Ethology*, *106*, 467–477.



APPENDIX II |

SUPPLEMENTARY INFORMATION CHAPTER 2

Supplementary information

Cardoso S.D., Faustino A.I., Costa S.S., Valério F., Gonçalves D. & Oliveira R.F., 2017. Social network predicts loss of fertilizations in nesting males of a fish with alternative reproductive tactics. *acta ethologica*, *20*, 59–68.
doi: 10.1007/s10211-016-0249-9

Table S1 – Microsatellite *in silico* characteristics and BLAST results.

This table contains the detailed BLASTX results, GO terms, enzyme commission number (EC number), microsatellite position in the gene (when possible), and the *in silico* characteristics observed for each one of the 28 microsatellites. The number of repeats reported in this table are the ones found by *in silico* mining, and not the ones submitted to GenBank. BLASTX results were obtained using an e-value of 1e-5 with a minimal *HSP (*high-scoring segment pair*) length cut-off* of 33 amino acids, against the NCBI non-redundant (nr) protein sequence database (release of May 2010).

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo01§	$(GT)_5$	2	2	TAP-binding protein	1,7e-14	**P:** antigen processing and presentation of endogenous peptide antigen via MHC class I **C:** membrane	3'UTR
Spavo02§	$(GA)_7C(GA)_4$	4	2	Ras association domain-containing protein 1	1,5e-21	**P:** intracellular signalling pathway; signal transduction **F:** metal ion binding; receptor activity; protein binding	3'UTR
Spavo03§	$(AC)_5$-$(GT)_5$	98	2	Hook homolog 1	4,8e-41	**P:** endosome to lysosome transport; spermatid development; early endosome to late endosome transport; endosome organization; protein transport; lysosome organization **F:** microtubule binding; identical protein binding; actin binding **C:** FHF complex; HOPS complex; microtubule	3'UTR
Spavo04§	$(AC)_5$	7	2	HAUS augmin-like complex subunit 5	8,3e-11	**P:** cell cycle; mitosis; spindle assembly; centrosome organization; cell division **C:** cytoplasm; microtubule organizing center; spindle; cytoskeleton; microtubule; HAUS complex	3'UTR

(Continued)

Table S1 (*Continuation*)

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo05§	$(AC)_7$	11	3	Transketolase-like protein 2 EC:2.2.1.1	1.3e-23	**P:** metabolic process **F:** transketolase activity	3'UTR
Spavo06§	$(TG)_8$	6	2	Mitogen-activated protein kinase 8 interacting protein 3-like	1,6e-19	**F:** kinase activity	3'UTR
Spavo07§	$(AC)_4G(AC)_{10}$	6	2	Type alpha 1 chain	1,1e-23	**P:** cell adhesion **F:** platelet-derived growth factor binding; extracellular matrix structural constituent **C:** collagen type VI; sarcolemma	3'UTR
Spavo08§	$(CA)_8$	5	3	Ubiquitin C-terminal hydrolase L1 EC:3.1.2.15, EC:3.4.19.0, EC:3.4.22.0	3,9e-44	**P:**ubiquitin-dependent protein catabolic process; multicellular organismal process; protein deubiquitination; behavior **F:** ubiquitin thiolesterase activity; omega peptidase activity; ubiquitin binding; cysteine-type endopeptidase activity **C:**cytoplasm	3'UTR
Spavo09§	$(AC)_7$	2	2	Tyrosyl-tRNA synthetase, cytoplasmic EC:6.1.1.1	2,3e-31	**P:** tyrosyl-tRNA aminoacylation **F:** tyrosine-tRNA ligase activity; protein binding; ATP binding; tRNA binding **C:** cytoplasm; nucleus	3'UTR

(Continued)

Table S1 (*Continuation*)

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo10§	$(AC)_8$	2	2	Sulphite oxidase EC:1.8.3.1	1,1e-13	**P:** response to nutrient; oxidation reduction **F:** electron carrier activity; sulfite oxidase activity; molybdenum ion binding; heme binding; molybdopterin cofactor binding **C:** mitochondrial intermembrane space; cytosol	3'UTR
Spavo11§	$(CT)_7$	14	2	UPF0538 protein C2orf76 homolog isoform 2	2,4e-20	**P:** biologiacal_process **F:** molecular_function	5'UTR
Spavo12§	$(AC)_7G(AC)_{11}$	6	2	ADP-ribosylation factor	3,2e-29	**P:** vesicle-mediated transport; positive regulation of growth rate; embryonic development ending in birth or egg hatching; small GTPase mediated signal transduction; hermaphrodite genitalia development; molting cycle, collagen and cuticulin-based cuticle; body morphogenesis; nematode larval development; inductive cell migration; intracellular protein transport **F:** GTP binding **C:** Golgi apparatus	5'UTR

(Continued)

Table S1 (*Continuation*)

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo13§	$(AC)_8$	3	3	Proteasome activator complex subunit 3 EC:1.4.3.4, EC:1.4.3.6	4,3e-123	**P:** negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle; regulation of apoptosis; oxidation reduction; anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process; cell adhesion; positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle; amine metabolic process **F:** peptidase activity; quinine binding; copper ion binding; p53 binding; MDM2 binding; identical protein binding; amine oxidase activity; proteasome activator activity **C:** integral to membrane; proteasome activator complex; cell surface; cytoplasm; nucleus; plasma membrane	3'UTR
Spavo14	$(AC)_{16}$	2	2	-	-	-	-
Spavo15§	$(AC)_6T(AC)_4$	6	2	Phosphoribosyl pyrophosphate synthetase-associated protein 1 EC:2.7.6.1	9,7e-43	**P:** nucleotide biosynthetic process **F:** ribose phosphate diphosphokinase activity; magnesium ion binding	3'UTR

(Continued)

Table S1 (*Continuation*)

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo16§	$(AC)_9$	10	2	Coagulin factor II (Prothrombin) EC:3.4.21.0	1,2e-158	**P:** cell surface receptor linked signaling pathway; platelet activation; positive regulation of protein amino acid phosphorylation; positive regulation of blood coagulation; proteolysis; positive regulation of release of sequestered calcium ion into cytosol; fibrinolysis; positive regulation of collagen biosynthetic process **F:** thrombospondin receptor activity; receptor binding; serine-type endopeptidase activity; calcium ion binding **C:** extracellular space	3'UTR
Spavo17§	$(TC)_6$	23	2	Adrenodoxin-like mitochondrial precursor	9,1e-36	**P:** transport; electron transport chain **F:** 2 iron, 2 sulfur cluster binding; metal ion binding; electron carrier activity **C:** mitochondrial matrix	5'UTR
Spavo18§	$(GA)_6$	3	2	ELAV-like protein 3 (Hu-antigen C)	1,2e-153	**P:** cell differentiation; nervous system development **F:** RNA binding; nucleotide binding **C:** ribonucleoprotein complex	3'UTR
Spavo19§	$(CA)_3T(CA)_6$	3	2	E3 ubiquitin-protein ligase HUWE1 EC:6.3.2.0	2,7e-113	**F:** acid-amino acid ligase activity **P:** protein modification process **C:** intracellular	3' UTR

(Continued)

Table S1 (*Continuation*)

Locus	Shortest allele *in silico*	Depth of coverage	*In silico* number of alleles	BLASTX / EC number	e-Value	GO terms	Repeat location
Spavo20 [c]	$(AGC)_{10}$	15	1	-	-	-	-
Spavo21	$(AATG)_{15}$	2	1	-	-	-	-
Spavo22	$(ATCC)_{14}$	5	1	-	-	-	-
Spavo23	$(CATT)_{8}$	3	1	-	-	-	-
Spavo24	$(CTGT)_{9}$	2	1	-	-	-	-
Spavo25	$(CTGT)_{9}$	4	1	-	-	-	-
Spavo26	$(GTTT)_{8}$	2	1	-	-	-	-
Spavo27	$(AAAC)_{20}$	3	1	-	-	-	-
Spavo28	$(CTATT)_{9}$	2	1	-	-	-	-

§ – Strategy 1

Table S2 – Microsatellite loci polymorphism characteristics for Formentera and Borovac populations.
This file contains the polymorphic characteristics of the 28 microsatellites developed from *Salaria pavo* unigenes in six individuals tested from Formentera and Borovac populations.

Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	Ta (ºC)	FORMENTERA				BOROVAC			
				Size (bp)	*k*	Ho	He	Size (bp)	*k*	Ho	He
Spavo01§ JQ619676	$(GT)_6$	F-CACCTCGAACAGTTGGCTTC GCTGCATTAGCCCAGATCC	58	387	1	-	-	385–387	2	0.33	0.33
Spavo02§ JQ619677	$(GA)_8C(GA)_4$	F-CCCTGGCTGATGTGACTCC ACTCTCCAGGTGTAAGGCAC	61	254–256	2	0.00	0.53	262–268	2	0.67	0.53
Spavo03§ JQ619678	$(AC)_6$-$(GT)_6$	F-GCACAAGTCGGCACTCAAG GCCAAGCCGAGTATGAAGC	60a	235	1	-	-	235	1	-	-
Spavo04§ JQ619679	$(AC)_6$	F-CCCACGTCTGTTCAGTTGAC GGAGTTGGCACATTCCGTG	58	264	1	-	-	258	1	-	-
Spavo05§ JQ619680	$(AC)_9$	F-ATCAGCGCGAAACACATCG ACTGCACTCAAGTCAAAGCC	56	183–189	3	1.00	0.73	195–199	2	0.67	0.53
Spavo06§ JQ619681	$(TG)_8$	F-GCTGGTCGATGGCAGAATG GCGTCGGAAATACCGTTCC	58	295–297	2	0.68	0.53	297	1	-	-
Spavo07§ JQ619682	$(AC)_4G(AC)_{10}$	F-CACGACAGCTGGTCTCAAC GGGCTCACCAGTCCCATTC	58	331–333	2	0.33	0.33	331–337	3	0.33	0.73
Spavo08§ JQ619683	$(CA)_9$	F-CGTGACTTCATGGCAAGGG TGTGTGGAAACGATATGTGC	58	225–231	4	0.67	0.80	231–237	2	0.33	0.33

(Continued)

Table S2 (*Continuation*)

Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	Ta (ºC)	FORMENTERA				BOROVAC			
				Size (bp)	*k*	Ho	He	Size (bp)	*k*	Ho	He
Spavo09§ JQ619684	$(AC)_8$	F-CGCTAAAAGGAGGCAACATC ACAGCGACGAGCTTCATCTT	61	198	1	-	-	198–200	2	1.00	0.60
Spavo10§ JQ619685	$(AC)_9$	F-AGAGTAGGGGTCCGTCGATT TGGCAGTGAGAAAGTGCAAG	61	153–161	5	1.00	0.93	173	1	-	-
Spavo11§ JQ619686	$(CT)_9$	F-GGTAGCGAGAGACGCAGAAG GGTAGACCAGCGGTCTGAAG	62	230–232	2	0.33	0.33	232	1	-	-
Spavo12§ JQ619687	$(AC)_7G(AC)_{12}$	F-GCTGTAAAACTGCGTGGACA GGACGTGAACCTGGAGAAGA	61	196–202	2	0.33	0.33	204–214	4	1.00	0.80
Spavo13§ JQ619688	$(AC)_{10}$	F-CCTCGCAGCAGTAACTCAGA TCCGTCTATGGAGGCTAACG	61b	136–144	2	0.33	0.33	138	1	-	-
Spavo14 JQ619689	$(AC)_{17}$	F-GGGGATCGAAATGTTTCACA CCACATGGAACCAACTTCCT	59	248–250	2	0.00	0.53	256–260	2	0.67	0.53
Spavo15§ JQ619690	$(AC)_6T(AC)_4$	F-CATGGCCTATCTGTTCCGC AGACCAACATCCCAGTCGC	58	240	1	–	–	240	1	-	-
Spavo17§ JQ619692	$(TC)_7$	F-TGTCAAGCTCACAGCGAC ATGGCACCCATGCTTCAGG	56a	216	1	–	–	216–218	2	0.33	0.33
Spavo16§ JQ619691	$(AC)_5T(AC)_5$	F-GTTCAGGATGACCCGGTGG TGTGTATGAGTTCCTGCCC	56	164–168	2	0.33	0.60	164	1	-	-

(Continued)

Table S2 (*Continuation*)

Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	Ta (ºC)	FORMENTERA				BOROVAC			
				Size (bp)	*k*	Ho	He	Size (bp)	*k*	Ho	He
Spavo18§ JQ619693	$(GA)_7$	F-CCATGACCAACTACGACGAG GGAGCTTAGGTCGCTCACC	62	175	1	–	–	175	1	-	-
Spavo19§ JQ619694	$(CA)_3T(CA)_7$	F-ACCTTCCAGCCTACGAGAGC TGTGTCAGGAGTAGGCAGACC	62	170–172	2	0.67	0.53	164	1	-	-
Spavo20 JQ619695	$(AGC)_{10}$	F-TGCTCGGCTCTACGGTTC CCCTCACAGAGTTCACGGG	60	227–233	3	0.33	0.73	221–224	2	0.33	0.33
Spavo21 JQ619696	$(AATG)_{15}$	F-TGTGTTGGTTTGAGACGGC CCTCAAAGACATTGGATGCG	60	298–314	3	1.00	0.73	302–338	4	0.67	0.80
Spavo22 JQ619697	$(ATCC)_{14}$	H-GGCAGAAGGAAACCTGGAC GGCCCTTGAAACTCCACTCT	61	143–189	3	0.33	0.73	195–211	4	1.00	0.87
Spavo23 JQ619698	$(CATT)_8$	H-CGACCCATTTCGGTTACAAG GAACGAGTAACGTGATGCTGA	61	249–257	3	0.67	0.73	245	1	-	-
Spavo24 JQ619699	$(CTGT)_9$	F-GCTCCAACAGAGATAAAACGCTCT TCACTGTAGGAACACGGGAAT	62	178–182	2	0.33	0.60	174–178	2	1.00	0.60
Spavo25 JQ619700	$(CTGT)_{10}$	H-GAGTGAGCCGGAGTGTTCTG GGCTAAACTGTGGCTGCCTA	62	232–236	2	0.33	0.33	228–232	2	0.67	0.53
Spavo26 JQ619701	$(GTTT)_9$	H-CACGTTGCCAATTCCAGTAG GAAGACGACAACCACTCTCAG	59	212	1	-	-	204	1	-	-

(Continued)

Table S2 (*Continuation*)

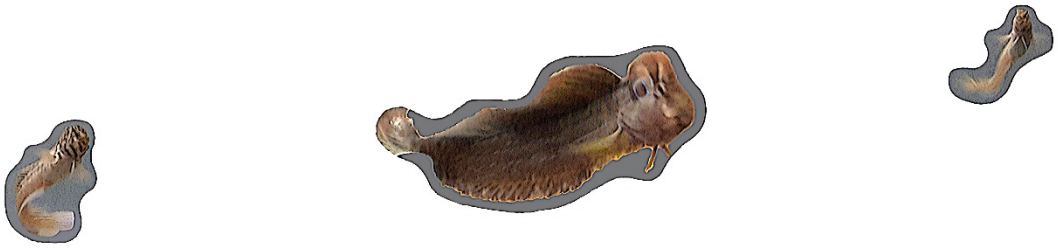
Locus GenBank No	Repeat motif	Primer Sequence (5'-3')	Ta (ºC)	FORMENTERA				BOROVAC			
				Size (bp)	*k*	Ho	He	Size (bp)	*k*	Ho	He
Spavo27	$(AAAC)_{13}$	F-GAGCTGGCGTTTCCCAAATA	59	185–189	2	0.33	0.33	177–220	5	1.00	0.93
JQ619702		ACGGCGTAGTGAGCATGTTG									
Spavo28	$(CTATT)_{10}$	H-GCAGAGTGACAATAAAGGACGA	59a	302–307	2	1.00	0.60	307–333	3	0.33	0.73
JQ619703		CCACAAGGCTCAGTTTGACA									

Ta (ºC) – annealing temperature; Ho – observed heterozygosity; He – expected heterozygosity; k – number of alleles; "F-" or "H-" at the 5′ end of the primer indicates FAM- or HEX-labelled primer; Hardy-Weinberg expectation deviations, *P<0.05, **P<0.001.

a – Mg = 1.0 mM.

b – Mg = 1.75 mM.

§ – Strategy 1



APPENDIX III |

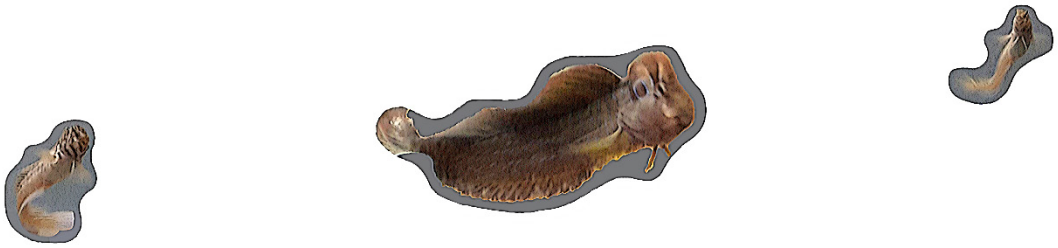
SUPPLEMENTARY INFORMATION CHAPTER 3

Supplementary information

Cardoso S.D., Faustino A.I., Costa S.S., Valério F., Gonçalves D. & Oliveira R.F., 2017. Social network predicts loss of fertilizations in nesting males of a fish with alternative reproductive tactics. *acta ethologica*, *20*, 59–68.
doi: 10.1007/s10211-016-0249-9

Table S1 – Summary of all Pearson correlations among morphometric and social variables (above diagonal) used as predictors of fertilization success in the multiple regression model, with respective unadjusted *P*-values (below diagonal).

	Standard length	Crest size	Relative anal gland area	EgoNet sneaker	EgoNet heterogeneity	Centrality
Standard length		**−0.68**	0.42	0.19	−0.42	−0.58
Crest size	**0.021**		−0.29	−0.60	0.16	0.40
Relative anal gland area	0.199	0.383		0.46	0.19	0.24
EgoNet sneaker	0.582	0.051	0.139		0.38	0.26
EgoNet heterogeneity	0.195	0.642	0.579	0.253		**0.78**
Centrality	0.062	0.227	0.487	0.445	**0.005**	



APPENDIX IV |

SUPPLEMENTARY INFORMATION CHAPTER 4

Supplementary information

Cardoso S.D., Gonçalves D., Goesmann A., Canário A.V.M. & Oliveira R.F., 2018. Temporal variation in brain transcriptome is associated with the expression of female mimicry as a sequential male alternative reproductive tactic in fish.
Molecular Ecology, *27*, 789–803.
doi: 10.1111/mec.14408

IV.1. Supplementary Methods

Assembly "clean-up"

To reduce sequence redundancy in the Trinity assembly, CD-HIT-EST v4.5.7 (Fu et al., 2012; Li and Godzik, 2006) was used with an alignment coverage for the shorter sequence of 1 and sequence identity of 0.99, resulting in an assembly reduction of 3.12%. Detection of exogenous contigs to the peacock blenny transcriptome that could represent bacterial, viral or fungal contamination and possible vectors that still persisted in the assembly was achieved by doing BLASTN searches against locally installed databases (NCBI UniVec and NCBI fungal, bacterial and viral genomes; January 2014). Contigs with successful hits (E-value cut-off of 10−5 and query coverage >=45%) were subsequently blasted against a local installed partial database of the UniProtKB (taxonomic division vertebrate databases, release April 2014) using BLASTX with an E-value cut-off of 10−5. By searching for sequence similarities between the detected exogenous contigs and vertebrate sequences we intended to avoid removing 'true' contigs from the peacock blenny transcriptome. The remaining 2,246 contigs without BLAST results were considered potential 'exogenous' sequences and therefore removed from the assembly before further analyses.

Transcriptome reduction

To remove from the assembly contigs with low read support, we applied different filter thresholds to the overall expression of each contig in the dataset and observed the behaviour of the normalization step with DESeq's rescaling factors. We computed the number of mappable Fragments Per Kilobase per Million fragments (FPKM; (Trapnell et al., 2010)) and filtered contigs using fragment coverage thresholds between 0, all contigs are kept, and 5, where contigs that failed to have at least a total representation of 5 mapped FPKMs across libraries were removed from the analysis. For each threshold, the raw count data matrices were extracted, and we assessed how the normalization of counts across the eight libraries varied using the implement methods in DESeq. Since most genes are expected not to be differentially expressed between phenotypes, especially between biological replicates, the differential expression analysis proceeded using the FPKM cut-off threshold for which we saw no improvement in the median of scaled counts between libraries when compared with the following threshold, cut-off threshold for transcript coverage of at least two fragments per million mappable reads (FPKM = 2; 179,202

contigs; Fig. S2). We proceeded with the analyses using this threshold, instead of using a higher cut-off value, since we obtained a significant reduction in the number of transcripts present in the assembly (*ca.* 69%) without either a substantial reduction in the number of mapped fragments (*ca.* 1.9%) or an increment in inter-library variability in fragment mappability (Table S2).

Supplementary References

Fu L., Niu B., Zhu Z., Wu S. & Li W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*, 3150–3152.

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

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J. & Pachter L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*, 511–515.

IV.2. Supplementary Figures

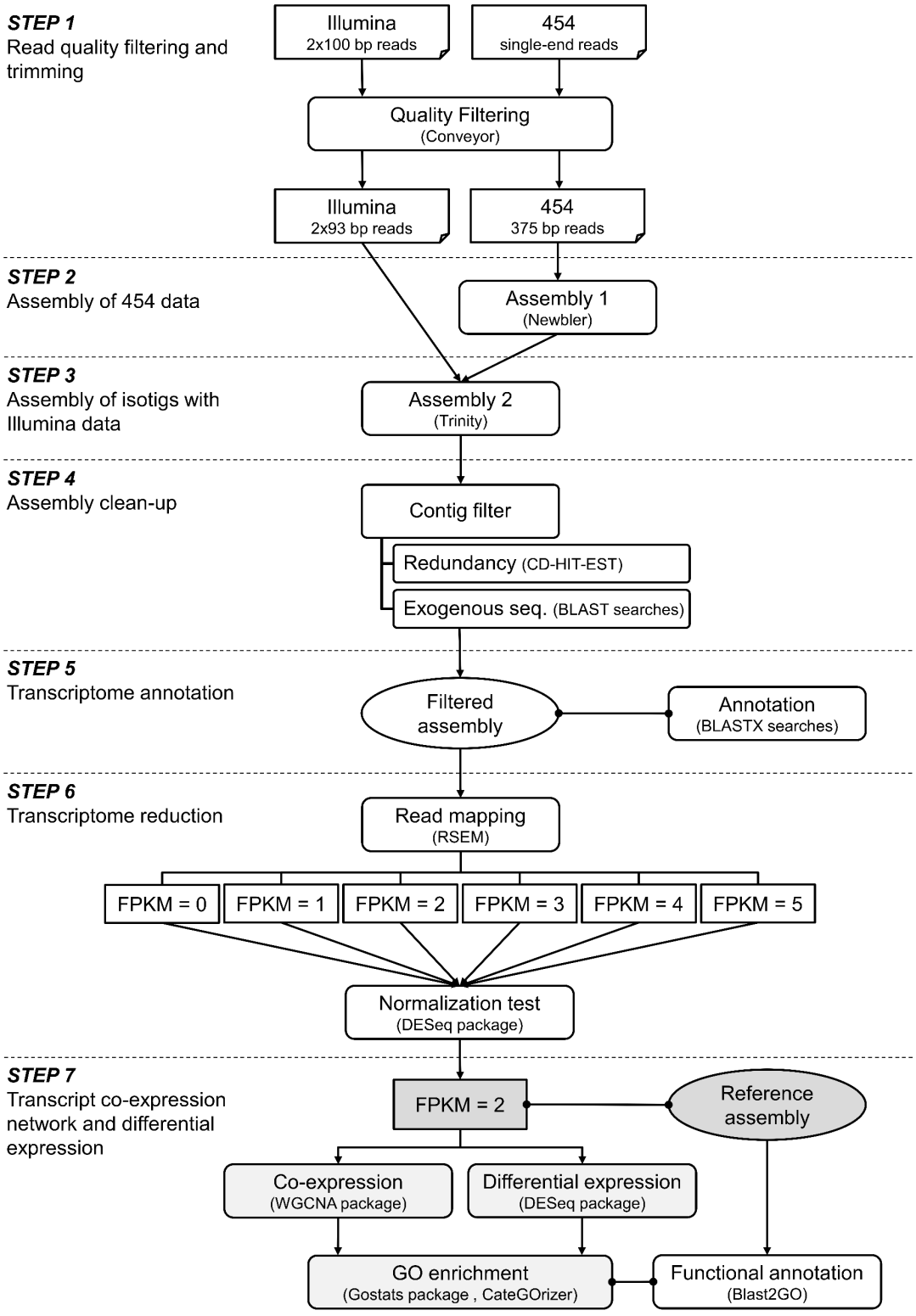


Fig. S1 – Data analysis workflow outline.

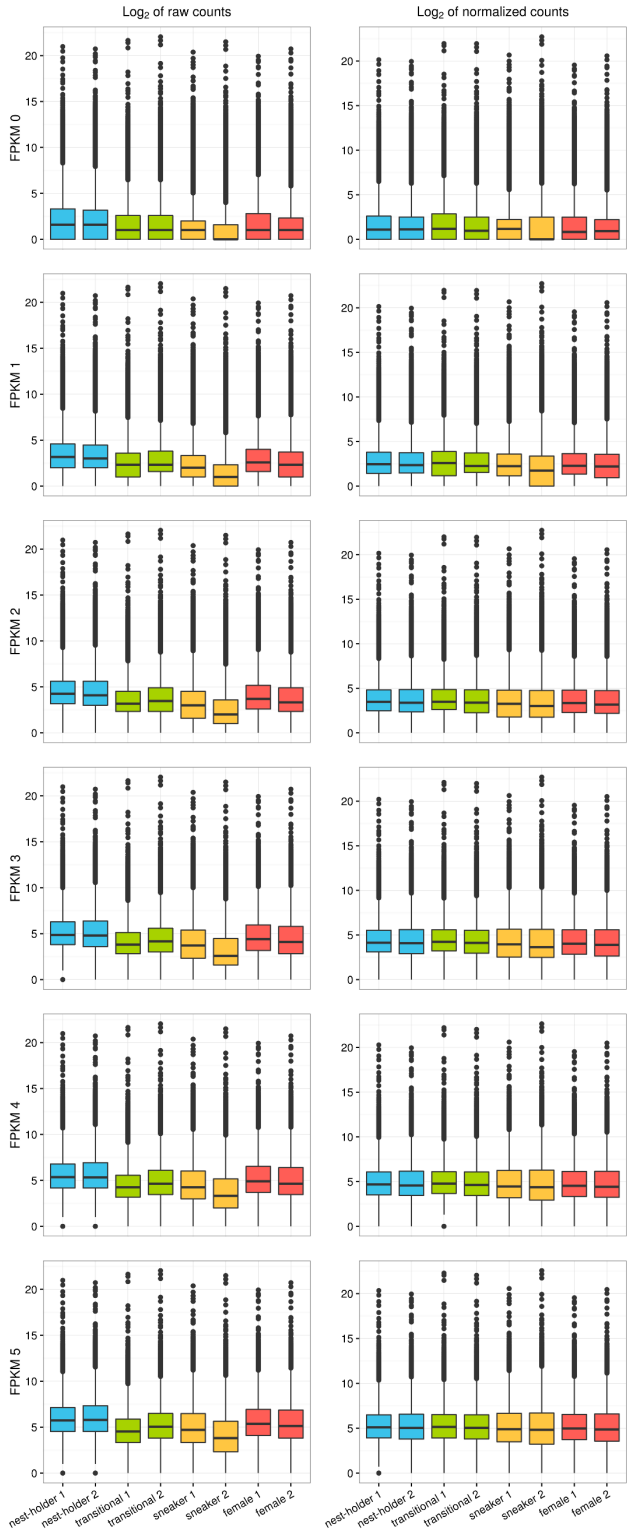


Fig. S2 – Boxplots of the shifted logarithm of raw (left column) and normalized counts (right column), for each FPKM threshold (rows) tested in this study to remove low-level expression noise. Further analyses proceeded with a reduced transcriptome (179,202 contigs) whose transcripts had at least 2 mapped FPKM fragments. Each sequenced sample is colour coded in agreement with the corresponding phenotype, blue for nest-holder males, green for transitional males, orange for sneaker males, and red for females.

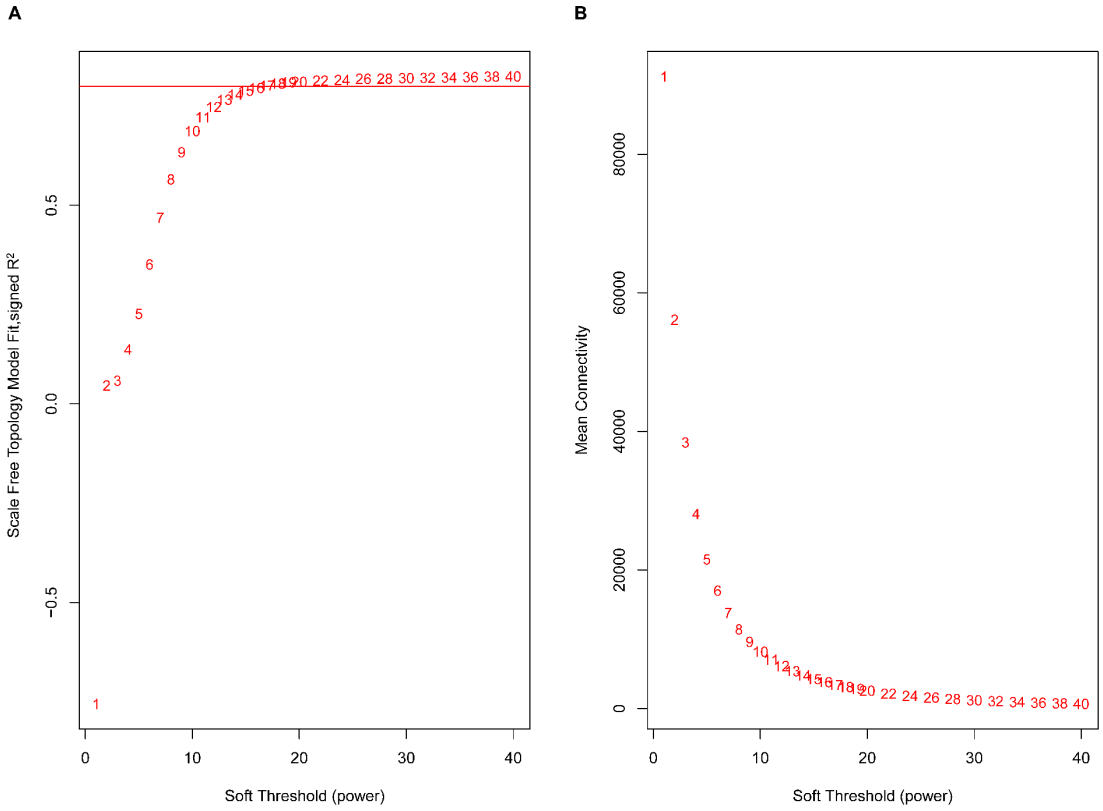


Fig. S3 – Summary of network indices, A) signed R square for scale free topology model fit, and B) mean connectivity, in function of the soft thresholding power. Approximate scale free-topology is attained around the soft-thresholding power of 17, corresponding to a mean connectivity of 3,406.6 transcripts.

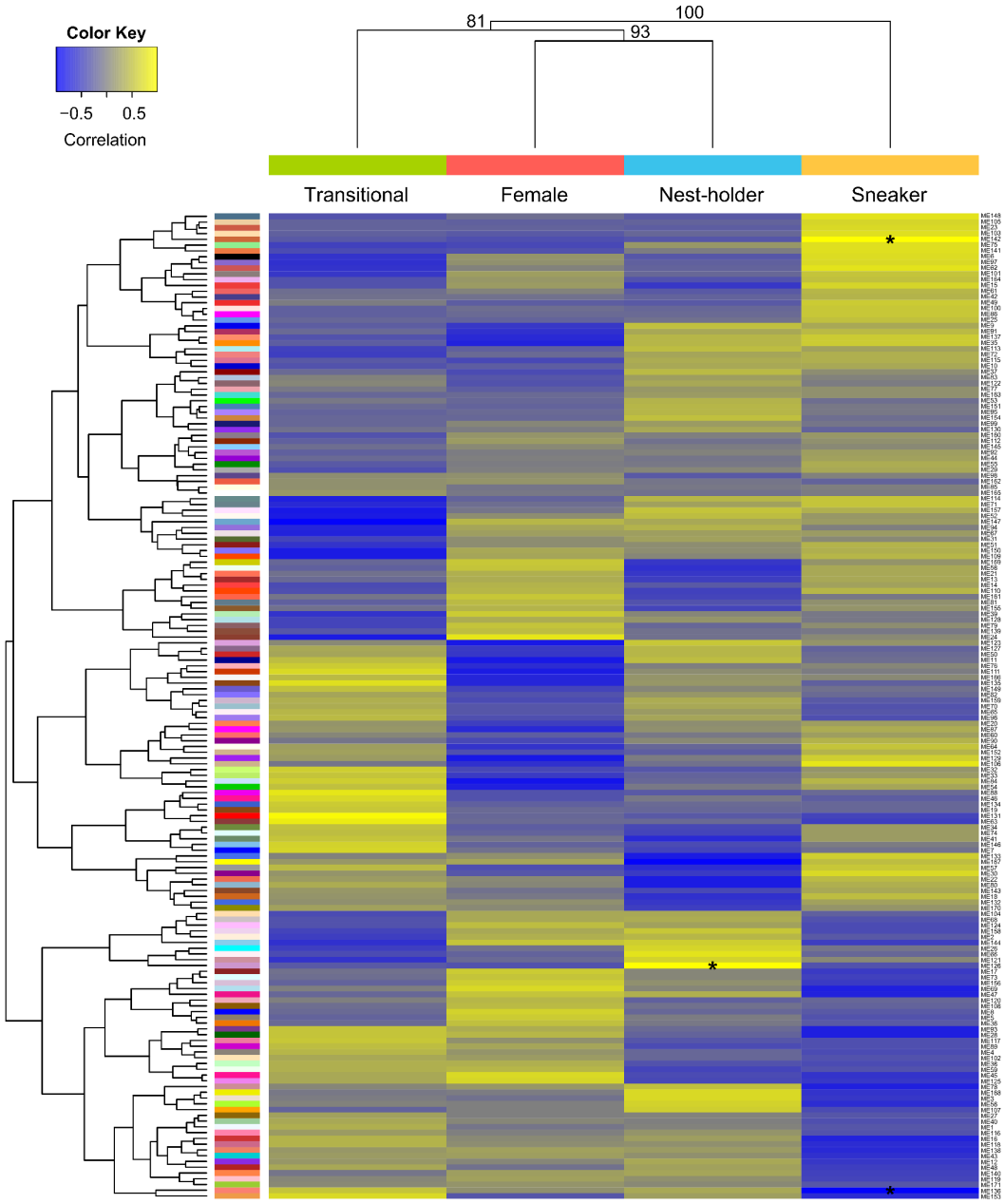


Fig. S4 – Heatmap showing the association between identified WGCNA transcript modules (rows) and each phenotype (columns). From the 171 identified modules, three modules remained significantly associated with a specific phenotype, after correcting the *P*-value for multiple comparisons (*). Intensity of colour indicates the strength of the correlation coefficient, ranging between negative values (in blue) and positive (in yellow). Similarity between phenotypes with hierarchical clustering can be seen above the heatmap with respective bootstrap values.

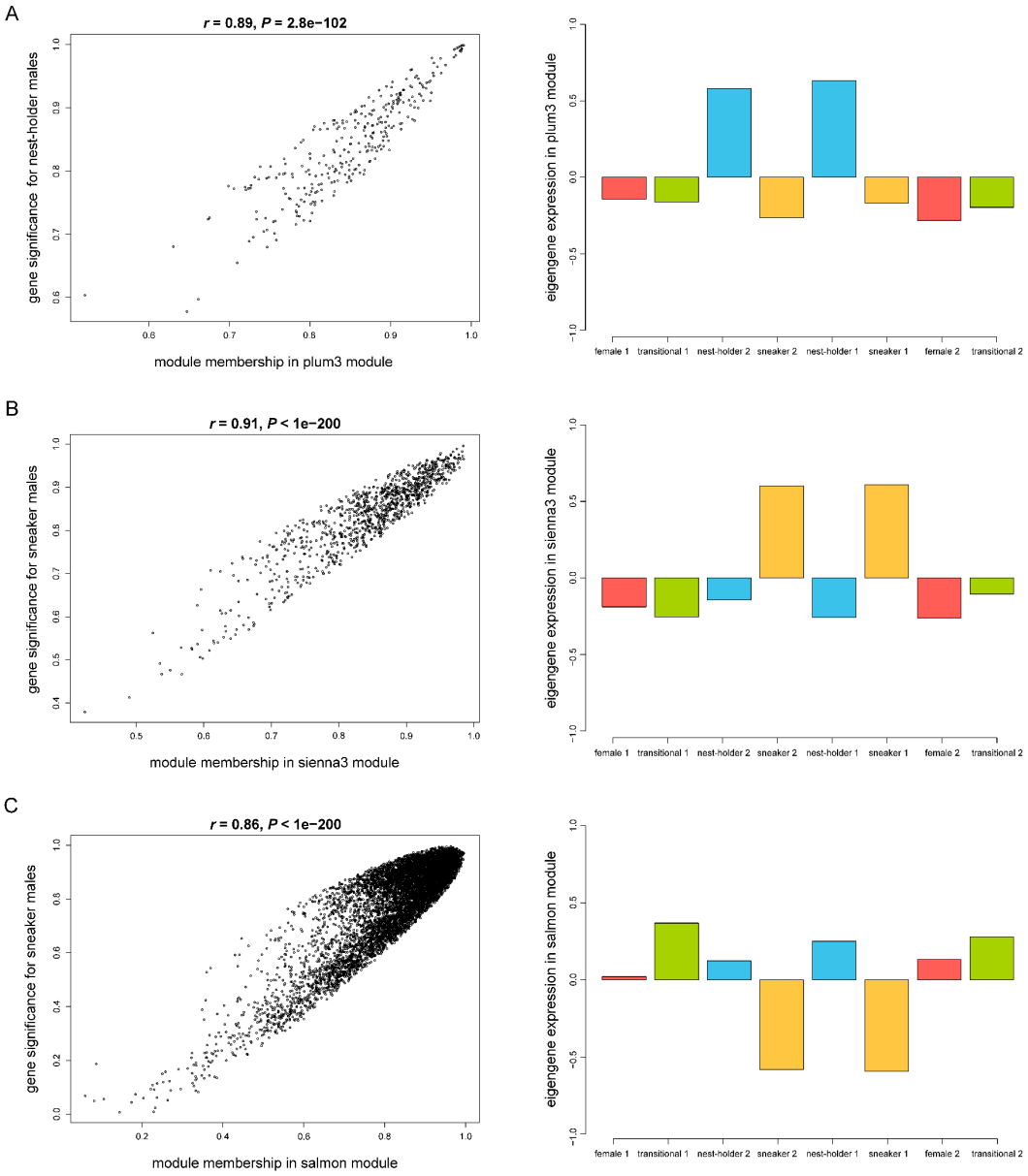


Fig. S5 – Plots representing the correlation between gene significance for the phenotype and module membership (left), with respective correlations and *P*-values, and expression values of the module eigengene for each sample sequenced (right), for A) plum3 module in nest-holder males, B) sienna3 and C) salmon modules in sneaker males. A significant positive correlation between transcript significance for the phenotype and module membership can be seen in all three modules, corroborating the importance of these modules for both males. Bars are colour coded according to phenotype so that blue represents nest-holder males, green transitional males, orange sneaker males and red females.

IV.3. Supplementary Tables

Table S1 – Sample-wise numbers of raw and filtered reads are shown. Raw data was stored in NCBI under BioProject PRJNA329073.

Sequencing	Sample name	No. reads	Quality filter	
			No. reads	Mean size (bp)
454 SE	Spavo_multi-tissue	642,012	640,760	320.41
454 SE	nest-holder_454	165,683	149,537	385.55
Illumina PE	nest-holder 1	36,867,409	31,153,046	92.87
Illumina PE	nest-holder 2	41,550,814	35,459,553	93.12
454 SE	female_454	170,426	158,772	398.14
Illumina PE	female 1	29,758,929	25,397,988	93.08
Illumina PE	female 2	33,359,499	27,445,776	92.60
454 SE	sneaker_454	169,751	157,130	386.89
Illumina PE	sneaker 1	24,001,326	20,728,000	93.78
Illumina PE	sneaker 2	22,880,412	18,889,379	92.98
454 SE	transitional_454	180,252	163,302	384.16
Illumina PE	transitional 1	23,652,592	20,584,361	93.21
Illumina PE	transitional 2	35,733,143	30,319,852	92.94

Table S2 – Summary of transcriptome and mapped read details when considering different FPKM thresholds for the assembled transcripts.

		FPKM=0	FPKM=1	FPKM=2	FPKM=3	FPKM=4	FPKM=5
Transcriptome	Trinity “genes”	467,304	255,904	142,805	98,546	75,346	61,584
	Trinity transcripts	577,532	312,948	179,202	124,479	95,301	77,818
	% GC	43.80	43.80	44.54	45.27	45.86	46.34
	N50 length	1,165	1,320	1,646	1,826	1,947	2,031
	N50 contig	91,053	61,056	36,270	25,078	19,136	15,570
	Median contig length	422	636	808	878	926	957
	Average contig length	754.36	952.52	1,166.28	1,271.07	1,340.69	1,387.49
	Maximum contig length	29,422	29,422	29,401	29,401	29,401	29,401
	Assembled bases	4.36E+08	2.98E+08	2.09E+08	1.58E+08	1.28E+08	1.08E+08
Mapped reads	Average no. mapped pairs of reads (%)	79.70	79.09	77.80	76.55	75.50	74.62
	Standard error	0.37	0.37	0.41	0.50	0.59	0.68

Table S3 – Principal Component Analysis (PCA) loadings for the first three components of each sequenced sample.

Sample	Comp. 1	Comp. 2	Comp. 3
nest-holder 1	0.2822	-0.1539	-0.5509
nest-holder 2	0.1928	-0.2314	-0.5733
transitional 1	0.3106	0.4931	0.1392
transitional 2	0.2786	0.5067	0.1629
sneaker 1	-0.5731	0.0490	-0.0428
sneaker 2	-0.5723	0.0508	-0.0396
female 1	0.1616	-0.4648	0.3879
female 2	0.1647	-0.4494	0.4100

Table S4 – List of differentially expressed contigs between nest-holder males and females of *Salaria pavo*. Positive log2FoldChange indicates an up-regulation of the transcript for females, whereas a negative log2FoldChange indicates an up-regulation of the transcript for nest-holder males.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp3506_c0_seq82	*fryl*	protein furry homolog-like	0.00E+00	14.8666	1.85E-21
comp6311_c0_seq37	*adam23*	ADAM metallopeptidase domain 23	0.00E+00	13.3302	4.79E-12
comp4000_c4_seq5	*atp1a3b*	ATPase, Na+/K+ transporting, alpha 3 polypeptide-like	0.00E+00	-13.9933	3.21E-11
comp22092_c0_seq2	–	–	–	-7.6673	5.32E-10
comp23805_c0_seq5	*celf1*	cugbp, Elav-like family member 1-like	0.00E+00	-12.5157	2.12E-08
comp292_c6_seq16	*map4k4*	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	11.9678	2.12E-08
comp29531_c0_seq12	*Ttll4*	tubulin tyrosine ligase-like family, member 4-like	0.00E+00	12.3342	2.12E-08
comp211_c0_seq5	*ralgps1*	ras-specific guanine nucleotide-releasing factor RalGPS1-like	2.80E-160	12.1273	2.17E-08
comp700_c4_seq17	–	–	–	11.8654	5.01E-08
comp7003_c0_seq3	*fam20c*	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	-12.0094	6.56E-08
comp21424_c0_seq1	*pde7b*	cAMP-specific 3',5'-cyclic phosphodiesterase 7B-like	0.00E+00	6.4400	9.78E-08
comp19383_c0_seq14	*asa2*	Ras GTPase-activating protein 2-like	0.00E+00	-12.3033	3.37E-07
comp5839_c0_seq35	*nfasc*	neurofascin-like	0.00E+00	12.6722	1.96E-06
comp3506_c0_seq318	*fryl*	protein furry homolog-like	2.29E-114	-7.0223	3.40E-06
comp7069_c0_seq4	*st7*	suppression of tumorigenicity 7 protein homolog	0.00E+00	-13.2985	5.26E-06
comp6074_c0_seq1	–	–	–	-6.5744	7.89E-06
comp22879_c1_seq3	*glyatl1*	glycine N-acyltransferase-like	4.79E-76	-12.1959	8.51E-06
comp3480_c0_seq3	–	–	–	-13.0559	9.65E-06
comp26171_c0_seq2	*fmnl2*	formin-like protein 2-like	7.64E-155	11.3044	9.85E-06
comp1842_c1_seq19	*dst*	dystonin	0.00E+00	-3.1903	2.01E-05
comp16233_c0_seq2	*tdrd3*	tudor domain-containing protein 3-like	0.00E+00	-11.4672	3.26E-05
comp7272_c6_seq10	*dnmt3a*	DNA (cytosine-5-)-methyltransferase 3 alpha-like	0.00E+00	12.1913	4.21E-05
comp27209_c0_seq2	–	–	–	-11.4411	5.40E-05
comp93579_c0_seq2	–	–	–	11.2446	5.84E-05
comp5839_c0_seq41	*nfasc*	neurofascin-like	0.00E+00	12.7069	6.17E-05
comp25676_c0_seq3	*satb1*	DNA-binding protein SATB1-like	0.00E+00	5.1761	7.16E-05
comp72_c10_seq25	–	–	–	4.0880	7.87E-05
comp2075_c4_seq61	–	–	–	11.5580	9.08E-05
comp82826_c0_seq2	*tdrd15*	tudor domain-containing protein 15-like	0.00E+00	-12.3165	1.08E-04
comp6182_c0_seq1	*rasgrf1*	Ras-specific guanine nucleotide-releasing factor 1-like	0.00E+00	5.5078	1.16E-04
comp15688_c1_seq7	*cacna1i*	calcium channel, voltage-dependent, T type, alpha 1I subunit	0.00E+00	11.5375	1.31E-04
comp1842_c1_seq17	*dst*	dystonin	0.00E+00	-3.8836	1.31E-04
comp22780_c0_seq1	*ankrd26*	ankyrin repeat domain 26-like	3.17E-137	-11.8591	1.31E-04
comp23344_c0_seq2	–	–	–	11.3618	1.31E-04
comp26642_c0_seq8	*tmem62*	transmembrane protein 62-like	0.00E+00	10.8018	1.31E-04
comp39182_c0_seq36	*erbb4*	receptor tyrosine-protein kinase erbB-4-like	0.00E+00	-11.1664	1.31E-04
comp61048_c0_seq2	–	–	–	-11.9313	1.31E-04
comp6389_c1_seq2	*tle2*	transducin-like enhancer protein 4-like	0.00E+00	-11.4657	1.48E-04
comp12858_c0_seq53	*plxna2*	plexin A2-like	0.00E+00	5.7177	1.54E-04
comp14699_c0_seq4	*srgap3*	SLIT-ROBO Rho GTPase activating protein 3	0.00E+00	11.9878	1.54E-04
comp24806_c0_seq3	–	–	–	-11.4672	1.72E-04
comp94250_c0_seq1	–	–	–	11.1228	1.93E-04
comp83787_c0_seq2	–	–	–	-11.4440	2.27E-04
comp7483_c2_seq2	–	–	–	11.4988	2.53E-04
comp62390_c0_seq2	*scinla*	scinderin like a, also known as gelsolin-like	0.00E+00	-11.5304	2.78E-04

comp24803_c0_seq5	*bahcc1*	BAH and coiled-coil domain-containing protein 1-like	1.31E-127	10.9186	2.85E-04
comp11609_c0_seq15	*phlpp1*	PH domain leucine-rich repeat-containing protein phosphatase 1	0.00E+00	-5.4167	3.17E-04
comp13731_c2_seq198	*csmd1*	CUB and sushi domain-containing protein 1	0.00E+00	-11.1241	3.25E-04
comp28508_c1_seq1	–	–	–	-11.4007	3.25E-04
comp4820_c1_seq3	*mapt*	microtubule-associated protein tau-like	3.86E-10	-11.4133	3.25E-04
comp38328_c0_seq2	*s100a1*	protein S100-A1-like	4.81E-50	4.3628	3.74E-04
comp45523_c0_seq1	*pms1*	PMS1 protein homolog 1-like	0.00E+00	-11.1223	4.47E-04
comp7178_c5_seq1	*prickle2*	prickle homolog 2	0.00E+00	-3.4279	4.52E-04
comp7069_c0_seq2	*rnf6*	ring finger protein (C3H2C3 type) 6	2.69E-10	-13.3601	5.19E-04
comp3852_c0_seq13	*arhgef9*	Cdc42 guanine nucleotide exchange factor (GEF) 9	0.00E+00	6.4464	6.24E-04
comp18811_c0_seq1	*map3k9*	mitogen-activated protein kinase kinase kinase 9-like	0.00E+00	2.9256	7.18E-04
comp18295_c0_seq1	*dock9*	dedicator of cytokinesis protein 9-like	0.00E+00	-3.6058	8.89E-04
comp2068_c1_seq5	–	–	–	15.2464	9.05E-04
comp9181_c0_seq15	*fam171b*	protein FAM171B-like	0.00E+00	13.9489	9.05E-04
comp23805_c0_seq11	*celf1*	cugbp, Elav-like family member 1-like	1.80E-155	6.1711	9.36E-04
comp3556_c0_seq1	–	–	–	-4.2637	9.36E-04
comp5682_c0_seq7	*kif5c*	kinesin family member 5C-like	0.00E+00	6.4743	9.45E-04
comp36345_c0_seq2	–	–	–	-5.9919	9.46E-04
comp11119_c0_seq24	*cdkl5*	cyclin-dependent kinase-like 5-like	0.00E+00	12.3067	0.0010
comp15454_c0_seq8	*tln2*	talin-2-like	9.48E-98	11.1108	0.0011
comp23805_c0_seq2	*celf1*	cugbp, Elav-like family member 1-like	0.00E+00	7.4024	0.0012
comp5839_c0_seq25	*nfasc*	neurofascin-like	0.00E+00	12.0495	0.0012
comp16358_c1_seq5	*esyt2b*	extended synaptotagmin-like protein 2b-like	0.00E+00	3.9053	0.0012
comp30660_c0_seq3	*wdr20*	WD repeat-containing protein 20-like	0.00E+00	-10.8819	0.0012
comp11603_c0_seq1	*cers2*	ceramide synthase 2-like	1.25E-162	-11.6873	0.0012
comp24708_c0_seq4	*dlgap4*	discs, large (Drosophila) homolog-associated protein 4-like	0.00E+00	-4.0107	0.0012
comp131394_c0_seq1	*frmd3*	FERM domain-containing protein 3	1.11E-38	-11.1144	0.0012
comp2387_c2_seq29	*chd9*	chromodomain helicase DNA binding protein 9-like	0.00E+00	-11.7260	0.0012
comp30709_c0_seq2	*zc3h7b*	zinc finger CCCH domain-containing protein 7B	0.00E+00	-4.6875	0.0012
comp6434_c0_seq11	–	–	–	11.1716	0.0013
comp13708_c0_seq2	*stard9*	StAR-related lipid transfer (START) domain containing 9	2.49E-147	-4.2074	0.0013
comp2075_c4_seq32	*cacna1h*	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	9.38E-147	-2.8726	0.0013
comp13368_c1_seq10	*dnah10*	dynein, axonemal, heavy chain 10-like	0.00E+00	4.2735	0.0013
comp1794_c1_seq5	*sv2c*	synaptic vesicle glycoprotein 2C-like	2.49E-100	-11.5229	0.0013
comp9957_c1_seq34	*mtm1*	myotubularin 1	9.34E-55	10.9854	0.0016
comp24708_c0_seq13	*dlgap4*	discs, large (Drosophila) homolog-associated protein 4-like	6.24E-174	-6.0646	0.0017
comp14932_c2_seq93	*trpm3*	transient receptor potential cation channel, subfamily M, member 3	0.00E+00	-11.9046	0.0018
comp19166_c0_seq8	*c-fos*	proto-oncogene c-Fos-like	3.01E-172	11.5336	0.0018
comp33_c0_seq5	–	–	–	2.7643	0.0018
comp21751_c0_seq105	*trrap*	transformation/transcription domain-associated protein	0.00E+00	7.9839	0.0019
comp10687_c0_seq2	*kt3k*	ketosamine-3-kinase-like	5.48E-154	11.0484	0.0019
comp9163_c1_seq50	–	–	–	11.4834	0.0020
comp9691_c0_seq4	*c2cd5*	C2 calcium-dependent domain containing 5	0.00E+00	-12.1291	0.0021
comp20190_c0_seq3	*fam21c*	WASH complex subunit FAM21C-like	2.15E-78	-11.1126	0.0022
comp15419_c1_seq1	*shank3*	SH3 and multiple ankyrin repeat domains 3	0.00E+00	-4.3112	0.0024
comp3744_c0_seq26	–	–	–	-10.8872	0.0025
comp4110_c3_seq3	*ncor2*	nuclear receptor corepressor 2-like	0.00E+00	2.5752	0.0025
comp4084_c1_seq3	*hnrnpul1*	heterogeneous nuclear ribonucleoprotein U-like 1	0.00E+00	-12.8492	0.0027
comp1521_c4_seq19	*magi1*	membrane associated guanylate kinase, WW and PDZ domain containing 1	0.00E+00	-4.1384	0.0028
comp20824_c0_seq1	*aldh2*	aldehyde dehydrogenase 2	0.00E+00	-2.7696	0.0037
comp57997_c0_seq8	*dla*	delta-like protein A-like	0.00E+00	3.7968	0.0042
comp20893_c0_seq1	*traf2*	Tnf receptor-associated factor 2	0.00E+00	11.2804	0.0043
comp22879_c1_seq1	*mgme1*	mitochondrial genome maintenance exonuclease 1-like	0.00E+00	3.0505	0.0043
comp25998_c0_seq2	*dlgap2*	discs, large (Drosophila) homolog-associated protein 2-like	1.42E-25	-10.5805	0.0043
comp36007_c0_seq1	*nrg1*	pro-neuregulin-1, membrane-bound isoform-like	0.00E+00	5.9515	0.0043
comp2064_c0_seq7	*nr2f*	COUP transcription factor 2-like	2.59E-133	10.5405	0.0047
comp700_c4_seq5	*ptprd*	protein tyrosine phosphatase, receptor type, D-like	0.00E+00	-13.0302	0.0048
comp120052_c0_seq2	–	–	–	-10.7529	0.0058
comp28515_c0_seq2	–	–	–	3.8439	0.0061
comp94169_c0_seq1	–	–	–	-4.0127	0.0062
comp15454_c0_seq6	*tln2*	talin-2-like	3.06E-101	-11.6548	0.0074
comp8775_c0_seq14	*mpp3*	membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	1.18E-120	-4.5538	0.0078
comp898_c2_seq19	*erc1b*	ELKS/Rab6-interacting/CAST family member 1b-like	0.00E+00	2.6526	0.0079
comp52574_c0_seq11	*znf236*	zinc finger protein 236-like	0.00E+00	-10.5229	0.0081
comp17880_c0_seq6	*qrich1*	glutamine-rich protein 1-like	0.00E+00	-4.6714	0.0086
comp200_c1_seq1	*chchd3*	coiled-coil-helix-coiled-coil-helix domain containing 3	1.23E-14	-11.3049	0.0086
comp3506_c0_seq308	*fryl*	protein furry homolog-like	0.00E+00	12.7958	0.0097
comp24887_c0_seq2	*lcorl*	ligand-dependent nuclear receptor corepressor-like protein	6.68E-27	-12.3682	0.0105
comp3422_c0_seq6	*add1*	adducin 1 (alpha)	1.12E-69	-2.6263	0.0105
comp42_c0_seq1	–	–	–	2.5586	0.0105
comp14289_c0_seq3	*ankrd13c*	ankyrin repeat domain 13C-like	1.32E-122	-11.4326	0.0111
comp35117_c1_seq17	*44M2.3*	putative RNA exonuclease NEF-sp-like	7.46E-138	-4.7464	0.0115
comp26212_c0_seq3	*myo1e*	unconventional myosin IE-like	0.00E+00	-2.5785	0.0129
comp87515_c0_seq1	–	–	–	10.6435	0.0129
comp5666_c0_seq3	*srcin1*	SRC kinase signaling inhibitor 1	0.00E+00	3.7084	0.0134
comp3932_c0_seq4	*ttbk1*	tau tubulin kinase 1-like	0.00E+00	-3.3124	0.0135
comp4620_c1_seq10	*tcf7l2*	transcription factor 7-like 2-like	0.00E+00	5.1196	0.0135
comp3334_c0_seq8	*tom1l2*	target of myb1-like protein 2-like	2.29E-34	-12.3926	0.0143
comp80286_c0_seq2	*tnnt1*	troponin T, slow skeletal muscle-like	4.94E-10	-10.4710	0.0149
comp2289_c0_seq3	*anp32e*	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	9.98E-62	11.8051	0.0156

comp14205_c0_seq3	*eif4g3*	eukaryotic translation initiation factor 4 gamma, 3-like	5.53E-86	10.1289	0.0164
comp130526_c0_seq2	–	–	–	-5.4044	0.0175
comp63278_c0_seq2	–	–	–	11.3106	0.0176
comp4620_c1_seq14	*tcf7l2*	transcription factor 7-like 2-like	0.00E+00	-12.6637	0.0179
comp8780_c0_seq9	*Rap1gap2*	RAP1 GTPase activating protein 2	1.61E-141	10.6015	0.0182
comp203_c7_seq1	*ankrd34a*	ankyrin repeat domain 34a-like	0.00E+00	12.4178	0.0183
comp2319_c1_seq6	–	–	–	12.1987	0.0192
comp1124_c0_seq3	*tacc2*	transforming, acidic coiled-coil containing protein 2	0.00E+00	12.4103	0.0200
comp349_c1_seq4	–	–	–	2.3953	0.0200
comp63278_c0_seq1	–	–	–	-11.0624	0.0209
comp828_c1_seq2	–	–	–	13.6620	0.0214
comp14814_c3_seq17	*fcho2*	FCH domain only protein 2-like	1.38E-117	-11.9959	0.0220
comp49156_c0_seq2	–	–	–	-10.8726	0.0224
comp18110_c0_seq3	*pom121*	nuclear envelope pore membrane protein POM121-like	0.00E+00	-12.1754	0.0234
comp50821_c0_seq1	*rab3il1*	guanine nucleotide exchange factor for Rab-3A-like	1.99E-121	-10.8949	0.0235
comp74053_c0_seq1	–	–	–	-6.7007	0.0259
comp55445_c0_seq2	–	–	–	-4.3971	0.0259
comp3813_c0_seq6	*znf207*	zinc finger protein 207-like	1.44E-71	-12.6405	0.0298
comp7601_c0_seq3	*kcnma1*	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.00E+00	-3.8402	0.0299
comp42357_c0_seq1	–	–	–	-2.6527	0.0305
comp21751_c0_seq20	*trrap*	transformation/transcription domain-associated protein	0.00E+00	2.5354	0.0308
comp456_c4_seq4	*lgi1*	leucine-rich glioma-inactivated protein 1-like	0.00E+00	11.7607	0.0308
comp36508_c0_seq4	*fxr1*	fragile X mental retardation syndrome-related protein 1-like	6.82E-104	-10.3646	0.0312
comp53985_c0_seq1	*muc2*	mucin 2-like	5.71E-46	3.4015	0.0312
comp36143_c1_seq1	*dgcr6*	DiGeorge syndrome critical region gene 6-like	5.36E-84	-10.1699	0.0315
comp11119_c0_seq22	*cdkl5*	cyclin-dependent kinase-like 5-like	0.00E+00	11.7287	0.0342
comp124261_c0_seq1	–	–	–	-3.8692	0.0406
comp1803_c0_seq2	*ub*	ubiquitin	1.97E-52	-2.5752	0.0413
comp14425_c1_seq9	*parp14*	poly (ADP-ribose) polymerase 14-like	3.04E-44	-10.7191	0.0446
comp3137_c0_seq5	*fubp1*	far upstream element-binding protein 1-like	0.00E+00	8.5337	0.0499

Table S5 – List of differentially expressed contigs between nest-holder males and sneaker males of *Salaria pavo*. Positive log2FoldChange indicates an up-regulation of the transcript for sneaker males, whereas a negative log2FoldChange indicates an up-regulation of the transcript for nest-holder males.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp7496_c4_seq1	*lrrfip2*	leucine rich repeat (in FLII) interacting protein 2	4.82E-159	-14.9736	4.16E-14
comp3506_c0_seq134	*fryl*	protein furry homolog-like	0.00E+00	-14.5380	2.28E-12
comp765_c0_seq1	*rnf10*	ring finger protein 10-like	0.00E+00	-14.5108	2.28E-12
comp2816_c0_seq2	*lphn3*	latrophilin 3-like	0.00E+00	-14.3750	3.90E-12
comp451_c4_seq1	–	–	–	-14.3397	1.42E-11
comp1424_c1_seq3	*arhgap44*	Rho GTPase activating protein 44-like	9.68E-42	-13.9592	2.03E-10
comp292_c6_seq16	*map4k4*	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	12.5521	2.57E-09
comp7689_c1_seq2	*crebzf*	CREB/ATF bZIP transcription factor-like	2.75E-56	-13.6073	7.46E-09
comp211_c0_seq5	*ralgps1*	ras-specific guanine nucleotide-releasing factor RalGPS1-like	2.80E-160	12.4654	7.53E-09
comp3422_c0_seq6	*add1*	adducin 1 (alpha)	1.12E-69	-13.6562	7.66E-09
comp11119_c0_seq12	*cdkl5*	cyclin-dependent kinase-like 5-like	0.00E+00	-13.5046	1.31E-08
comp40969_c0_seq2	*akap13*	A kinase (PRKA) anchor protein 13-like	0.00E+00	12.9956	1.86E-08
comp700_c4_seq17	–	–	–	12.2136	2.34E-08
comp19398_c0_seq1	*setd2*	histone-lysine N-methyltransferase SETD2-like	0.00E+00	-13.7006	3.08E-08
comp20540_c0_seq1	*pi4k2a*	phosphatidylinositol 4-kinase type 2 alpha-like	0.00E+00	-13.8299	3.08E-08
comp10728_c1_seq2	–	–	–	-13.2455	9.80E-08
comp2446_c0_seq48	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	2.45E-150	-13.1205	2.04E-07
comp3813_c0_seq4	*znf207*	zinc finger protein 207-like	4.88E-71	-13.9971	2.04E-07
comp26642_c0_seq8	*tmem62*	transmembrane protein 62-like	0.00E+00	12.3258	3.28E-07
comp24068_c0_seq3	–	–	–	12.3988	3.90E-07
comp7272_c6_seq10	*dnmt3a*	DNA (cytosine-5-)-methyltransferase 3 alpha-like	0.00E+00	13.6134	6.92E-07
comp21022_c0_seq8	*bptf*	nucleosome-remodeling factor subunit BPTF-like	0.00E+00	-12.8796	1.39E-06
comp26171_c0_seq2	*fmnl2*	formin-like protein 2-like	7.64E-155	12.0703	1.39E-06
comp26241_c0_seq3	*nhs*	Nance-Horan syndrome protein-like	0.00E+00	-12.8095	1.39E-06
comp7003_c0_seq5	*fam20c*	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	-12.7554	2.35E-06
comp17250_c1_seq26	*dgcr2*	integral membrane protein DGCR2/IDD-like, also known as DiGeorge syndrome critical region gene 2	0.00E+00	12.6286	2.51E-06
comp6074_c0_seq1	–	–	–	-14.6347	2.51E-06
comp1127_c4_seq2	*iqsec2*	IQ motif and SEC7 domain-containing protein 2-like	0.00E+00	-5.9104	2.81E-06
comp16332_c0_seq2	*ophn1*	oligophrenin 1-like	0.00E+00	-12.9844	9.51E-06
comp537_c1_seq2	*cask*	peripheral plasma membrane protein CASK-like, also calcium/calmodulin-dependent serine protein kinase (MAGUK family) (cask)	0.00E+00	-4.2848	1.22E-05
comp2189_c1_seq26	*rbm6*	RNA-binding motif protein 6-like	1.07E-128	-3.6571	3.32E-05
comp6371_c1_seq11	*mical3*	protein-methionine sulfoxide oxidase mical3a-like, also known as microtubule associated monooxygenase, calponin and LIM domain containing 3	0.00E+00	-12.5449	3.64E-05
comp72_c10_seq25	–	–	–	4.7395	3.94E-05
comp20427_c0_seq2	*park*	parkin RBR E3 ubiquitin protein ligase	3.50E-114	11.8335	5.21E-05
comp6820_c0_seq1	–	–	–	-13.0952	5.26E-05
comp17443_c0_seq1	*abi2*	abl-interactor 2-like	5.69E-91	-12.4575	5.47E-05

comp57997_c0_seq8	*dla*	delta-like protein A-like	0.00E+00	4.9781	6.32E-05
comp4900_c1_seq16	–	–	–	-12.3850	6.92E-05
comp4650_c1_seq3	*gpc6*	glypican 6-like	0.00E+00	-12.3619	7.86E-05
comp7124_c0_seq7	–	–	–	-6.1709	8.43E-05
comp643_c1_seq4	–	–	–	-12.2430	1.02E-04
comp12126_c0_seq6	*sntg1*	syntrophin, gamma 1-like	0.00E+00	4.0648	1.16E-04
comp17142_c0_seq1	*myadm*	myeloid-associated differentiation marker homolog	2.09E-132	12.4019	1.30E-04
comp7149_c0_seq3	*wbp11*	WW domain-binding protein 11-like	1.48E-70	-12.3315	1.30E-04
comp2446_c0_seq38	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	-12.3279	1.50E-04
comp13993_c0_seq128	*prune2*	prune homolog 2	0.00E+00	-12.1923	1.54E-04
comp10086_c0_seq5	*fbp*	f-type lectin 3 (fucolectins)	1.17E-137	-14.4131	1.54E-04
comp1794_c1_seq4	*sv2c*	synaptic vesicle glycoprotein 2C-like	1.74E-99	4.9144	1.67E-04
comp17129_c0_seq1	–	–	–	-12.2625	1.82E-04
comp4110_c3_seq3	*ncor2*	nuclear receptor corepressor 2-like	0.00E+00	3.0561	2.04E-04
comp12377_c0_seq8	–	–	–	-12.2098	2.22E-04
comp18401_c1_seq2	*ubash3b*	ubiquitin associated and SH3 domain containing protein B-like	0.00E+00	-12.7075	2.24E-04
comp36007_c0_seq1	*nrg1*	pro-neuregulin-1, membrane-bound isoform-like	0.00E+00	7.1800	2.32E-04
comp7477_c0_seq1	*otof*	otoferlin-like	0.00E+00	-4.0050	2.32E-04
comp15688_c1_seq7	*cacna1i*	calcium channel, voltage-dependent, T type, alpha 1I subunit	0.00E+00	11.7355	2.49E-04
comp35039_c0_seq2	*dapk2*	death-associated protein kinase 2-like	5.18E-145	-12.2044	2.49E-04
comp14699_c0_seq7	*srgap3*	SLIT-ROBO Rho GTPase activating protein 3	2.27E-73	4.9558	2.53E-04
comp3506_c0_seq319	*fryl*	protein furry homolog-like	1.21E-100	-4.3743	3.02E-04
comp45141_c0_seq2	–	–	–	-12.1669	3.24E-04
comp11119_c0_seq24	*cdkl5*	cyclin-dependent kinase-like 5-like	0.00E+00	12.7698	3.42E-04
comp5839_c0_seq35	*nfasc*	neurofascin-like	0.00E+00	12.0445	3.58E-04
comp32242_c0_seq2	*znf592*	zinc finger protein 592-like	0.00E+00	-12.2291	3.94E-04
comp45781_c0_seq1	*thbs3a*	thrombospondin 3a-like	9.39E-135	5.5386	4.42E-04
comp11489_c0_seq8	*erc2*	ELKS/RAB6-interacting/CAST family member 2	0.00E+00	-4.7063	5.15E-04
comp12858_c0_seq53	*plxna2*	plexin A2-like	0.00E+00	5.6239	5.15E-04
comp13488_c2_seq2	*rdh14*	retinol dehydrogenase 14-like	4.95E-162	-12.2732	5.15E-04
comp24803_c0_seq5	*bahcc1*	BAH and coiled-coil domain-containing protein 1-like	1.31E-127	11.1879	5.15E-04
comp9222_c1_seq1	*fez2*	fasciculation and elongation protein zeta-2-like (Zygin II)	4.02E-150	-12.7815	6.12E-04
comp2852_c1_seq4	*ndufaf7*	NADH dehydrogenase [ubiquinone] complex I, assembly factor 7-like	0.00E+00	-5.0198	7.54E-04
comp55390_c0_seq1	–	–	–	-12.6544	8.47E-04
comp7807_c0_seq7	*akap9*	A kinase (PRKA) anchor protein 9-like	0.00E+00	-13.8096	9.36E-04
comp5668_c0_seq12	*syngap1*	synaptic ras/Rap GTPase-activating protein SynGAP-like	0.00E+00	-12.8494	9.41E-04
comp3140_c0_seq1	–	–	–	-12.1058	0.0011
comp2075_c4_seq48	*cacna1h*	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	0.00E+00	-12.4074	0.0013
comp3258_c0_seq2	–	–	–	-4.8804	0.0013
comp22092_c0_seq1	–	–	–	-12.1431	0.0014
comp5080_c1_seq13	*spef2*	sperm flagellar protein 2-like	1.20E-98	-11.9926	0.0015
comp8737_c0_seq13	*trim3*	tripartite motif-containing protein 3-like	0.00E+00	-12.5349	0.0017
comp9452_c0_seq1	*h1f0*	histone H1.0-B-like	4.99E-37	3.0212	0.0017
comp2964_c0_seq1	*eml1*	echinoderm microtubule-associated protein-like 1-like	0.00E+00	3.4319	0.0019
comp13368_c1_seq10	*dnah10*	dynein, axonemal, heavy chain 10-like	0.00E+00	4.5154	0.0020
comp10612_c0_seq19	*get4*	golgi to ER traffic protein 4 homolog	5.73E-130	-12.0168	0.0024
comp20054_c0_seq3	*tha1*	threonine aldolase 1	8.85E-161	-11.8003	0.0025
comp2068_c1_seq5	–	–	–	14.3699	0.0025
comp8725_c1_seq2	*gtpbp1*	GTP binding protein 1-like	0.00E+00	-11.7655	0.0025
comp24068_c0_seq2	–	–	–	-11.8305	0.0026
comp291_c2_seq7	*cndp2*	cytosolic non-specific dipeptidase-like	0.00E+00	-11.9289	0.0026
comp30496_c0_seq1	–	–	–	-11.8967	0.0028
comp5180_c0_seq3	–	–	–	-3.2186	0.0028
comp15373_c0_seq2	*agrn*	agrin-like	0.00E+00	-12.5726	0.0029
comp6080_c3_seq26	*arhgef11*	Rho guanine nucleotide exchange factor (GEF) 11-like	0.00E+00	-12.3414	0.0029
comp1555_c2_seq1	*snx27*	sorting nexin family member 27-like	0.00E+00	-11.7962	0.0031
comp93579_c0_seq2	–	–	–	10.2457	0.0032
comp16106_c0_seq1	*vps29*	vacuolar protein sorting-associated protein 29-like	7.30E-123	-3.4612	0.0035
comp1637_c4_seq15	*map3k9*	mitogen-activated protein kinase kinase kinase 9-like	2.20E-69	-11.7960	0.0037
comp11606_c0_seq4	*hmcn2*	hemicentin-2-like	3.73E-72	-3.9076	0.0038
comp19166_c0_seq8	*c-fos*	proto-oncogene c-Fos-like	3.01E-172	11.6864	0.0038
comp19489_c0_seq1	*tnrc6b*	trinucleotide repeat containing 6b-like	3.36E-125	-3.5628	0.0038
comp3506_c0_seq308	*fryl*	protein furry homolog-like	0.00E+00	13.8237	0.0038
comp5318_c2_seq29	*ddx3x*	DEAD (Asp-Glu-Ala-Asp) box helicase 3, also known as ATP-dependent RNA helicase DDX3X-like	0.00E+00	-4.2671	0.0038
comp5389_c0_seq2	*ylpm1*	YLP motif-containing protein 1-like	0.00E+00	-14.0177	0.0038
comp5648_c0_seq4	–	–	–	-11.9067	0.0038
comp6412_c0_seq3	*galnt18*	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 18-like	0.00E+00	-12.7201	0.0038
comp7003_c0_seq3	*fam20c*	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	2.9223	0.0039
comp16520_c0_seq1	*chd6*	chromodomain helicase DNA binding protein 6-like	1.11E-78	-11.9536	0.0041
comp55090_c0_seq1	–	–	–	-11.6506	0.0041
comp9181_c0_seq15	*fam171b*	protein FAM171B-like	0.00E+00	12.6563	0.0043
comp18599_c0_seq3	*ttn*	titin-like	2.30E-113	-4.7327	0.0046
comp18811_c0_seq1	*map3k9*	mitogen-activated protein kinase kinase kinase 9-like	0.00E+00	2.8302	0.0048
comp261418_c0_seq1	–	–	–	11.1497	0.0048
comp8404_c0_seq4	*limch1*	LIM and calponin homology domains-containing protein 1-like	1.71E-89	-4.7003	0.0048
comp9163_c1_seq50	–	–	–	11.4615	0.0063
comp1157_c1_seq11	*magi2*	membrane associated guanylate kinase, WW and PDZ domain containing 2	0.00E+00	3.0366	0.0065
comp1511_c1_seq9	*ncam1*	neural cell adhesion molecule 1-like	5.13E-15	-5.2040	0.0065

comp22089_c1_seq1	*gk*	glycerol kinase-like	0.00E+00	2.9373	0.0067
comp111738_c0_seq1	*nme2*	NME/NM23 nucleoside diphosphate kinase 2 (ndpkb)	2.45E-101	2.9793	0.0070
comp42933_c0_seq3	*tmem150c*	transmembrane protein 150C-like	9.74E-70	12.0288	0.0070
comp31969_c0_seq3	*ttll3*	tubulin tyrosine ligase-like family, member 3-like	0.00E+00	-5.0131	0.0070
comp23534_c1_seq9	*grip2*	glutamate receptor-interacting protein 2-like	0.00E+00	-12.0093	0.0077
comp28515_c0_seq2	–	–	–	4.1456	0.0077
comp13385_c1_seq10	*tjp1*	tight junction protein ZO-1-like	9.43E-10	-4.7969	0.0078
comp6051_c1_seq7	*aldh2*	aldehyde dehydrogenase, mitochondrial	2.87E-177	-4.2942	0.0080
comp9491_c0_seq10	*dip2ba*	DIP2 disco-interacting protein 2 homolog Ba-like	0.00E+00	-4.8977	0.0086
comp3567_c5_seq14	–	–	–	-5.9403	0.0087
comp7879_c1_seq24	*rere*	arginine-glutamic acid dipeptide repeats protein-like	0.00E+00	-11.5417	0.0087
comp8897_c0_seq1	*znf384*	zinc finger protein 384-like	4.56E-112	-4.3590	0.0087
comp9222_c1_seq9	*fez2*	fasciculation and elongation protein zeta-2-like (Zygin II)	1.81E-163	3.7036	0.0087
comp130808_c0_seq1	–	–	–	-11.5363	0.0088
comp23344_c0_seq2	–	–	–	10.0983	0.0088
comp1507_c0_seq1	*rap1gds1*	RAP1, GTP-GDP dissociation stimulator 1-like	0.00E+00	-2.2488	0.0089
comp11213_c0_seq1	*senp3*	SUMO1/sentrin/SMT3 specific peptidase 3-like	4.98E-97	-12.6714	0.0095
comp21751_c0_seq105	*trrap*	transformation/transcription domain-associated protein	0.00E+00	8.0236	0.0097
comp35860_c0_seq3	*mcm7*	DNA replication licensing factor mcm7-like, also known as minichromosome maintenance complex component 7	0.00E+00	3.0262	0.0097
comp14699_c0_seq4	*srgap3*	SLIT-ROBO Rho GTPase activating protein 3	0.00E+00	10.6731	0.0097
comp27209_c0_seq2	–	–	–	-11.4411	0.0097
comp9243_c0_seq1	*brd8*	bromodomain-containing protein 8-like	0.00E+00	-2.7928	0.0100
comp20893_c0_seq1	*traf2*	Tnf receptor-associated factor 2	0.00E+00	11.0051	0.0104
comp3650_c0_seq1	*crk*	crk-like protein-like	1.39E-147	-12.1703	0.0108
comp52164_c0_seq1	*poli*	DNA polymerase iota-like	0.00E+00	-12.1927	0.0108
comp22780_c0_seq1	*ankrd26*	ankyrin repeat domain 26-like	3.17E-137	-11.8591	0.0111
comp3137_c0_seq5	*fubp1*	far upstream element-binding protein 1-like	0.00E+00	9.6006	0.0111
comp112_c4_seq1	*trappc8*	trafficking protein particle complex subunit 8-like	0.00E+00	-3.4713	0.0112
comp26336_c0_seq2	*ccnt1*	cyclin T1-like	6.57E-57	11.9888	0.0119
comp41090_c0_seq2	*plcd4*	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like	0.00E+00	-11.5206	0.0119
comp270_c0_seq3	*atp1b4*	ATPase, Na+/K+ transporting, beta 4-like	0.00E+00	-6.9841	0.0119
comp10674_c0_seq4	–	–	–	-11.4681	0.0119
comp2446_c0_seq36	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	-11.5293	0.0120
comp15424_c1_seq2	*smc6*	structural maintenance of chromosomes protein 6-like	0.00E+00	-4.9637	0.0120
comp33057_c0_seq2	*klf11a*	Krueppel-like factor 11-like	7.70E-29	11.0314	0.0120
comp9538_c0_seq4	*mtmr5*	myotubularin-related protein 5-like	0.00E+00	-2.9171	0.0120
comp22295_c0_seq9	*dos*	protein dos-like	3.60E-141	-11.9639	0.0123
comp20893_c0_seq2	*traf2*	Tnf receptor-associated factor 2	6.93E-133	-11.7651	0.0127
comp68_c2_seq5	*rhoa*	transforming protein RhoA-like	6.68E-136	6.8641	0.0127
comp72278_c0_seq1	–	–	–	-11.8584	0.0127
comp23073_c0_seq1	*macf1*	microtubule-actin cross-linking factor 1-like	0.00E+00	-12.2974	0.0132
comp6263_c1_seq3	*ei24*	etoposide-induced protein 2.4 homolog	1.60E-162	-11.5280	0.0132
comp5080_c1_seq9	*spef2*	sperm flagellar protein 2-like	1.60E-179	-11.3628	0.0132
comp13813_c0_seq2	*sox9*	SRY (sex determining region Y)-box 9	2.92E-47	3.6384	0.0133
comp40823_c0_seq1	*pvalb*	parvalbumin beta-like	8.74E-56	5.2554	0.0133
comp19606_c0_seq2	*tdg*	G/T mismatch-specific thymine DNA glycosylase-like	9.66E-158	-11.5703	0.0137
comp3852_c0_seq13	*arhgef9*	Cdc42 guanine nucleotide exchange factor (GEF) 9	0.00E+00	4.9436	0.0140
comp8598_c0_seq1	*ercc6*	excision repair cross-complementation group 6	0.00E+00	-3.7802	0.0140
comp49590_c0_seq1	*urgcp*	up-regulator of cell proliferation-like	8.86E-166	-11.7091	0.0146
comp10558_c0_seq46	*kcnt1*	potassium channel, subfamily T, member 1	0.00E+00	-11.5535	0.0152
comp17468_c1_seq2	*smap1*	stromal membrane-associated protein 1-like	1.29E-101	3.2184	0.0152
comp9983_c0_seq3	*nt5c2*	5'-nucleotidase, cytosolic II	0.00E+00	-2.3756	0.0152
comp13993_c0_seq149	*prune2*	prune homolog 2	2.06E-147	3.2891	0.0157
comp7103_c0_seq3	–	–	–	4.0568	0.0165
comp14642_c0_seq7	*ncoa2*	nuclear receptor coactivator 2-like	1.59E-74	-11.4820	0.0172
comp3017_c0_seq1	*bin1*	myc box-dependent-interacting protein 1-like	1.55E-160	-4.5569	0.0175
comp13748_c0_seq1	*znf831*	zinc finger protein 831	2.13E-107	-2.4494	0.0175
comp456_c4_seq4	*lgi1*	leucine-rich glioma-inactivated protein 1-like	0.00E+00	11.8231	0.0179
comp24806_c0_seq3	–	–	–	-11.4672	0.0184
comp288_c5_seq6	–	–	–	-11.4536	0.0184
comp96121_c0_seq3	*ptx3*	pentraxin-related protein PTX3-like	0.00E+00	4.5355	0.0184
comp38566_c0_seq1	–	–	–	-11.5062	0.0189
comp22472_c0_seq2	–	–	–	-11.9081	0.0189
comp2054_c6_seq75	–	–	–	11.4517	0.0191
comp25278_c0_seq43	*hspg2*	heparan sulfate proteoglycan 2	0.00E+00	-11.5168	0.0191
comp26691_c0_seq2	–	–	–	-11.1838	0.0191
comp23426_c0_seq2	*nfrkb*	nuclear factor related to kappa-B-binding protein-like	0.00E+00	-12.7277	0.0192
comp17106_c0_seq3	*cntn3*	contactin 3	0.00E+00	-11.3405	0.0194
comp25676_c0_seq3	*satb1*	DNA-binding protein SATB1-like	0.00E+00	3.8032	0.0194
comp1848_c2_seq2	*uqcrc2*	cytochrome b-c1 complex subunit mitochondrial-like, also known as ubiquinol-cytochrome c reductase core protein II	4.00E-178	2.6291	0.0203
comp5839_c0_seq25	*nfasc*	neurofascin-like	0.00E+00	11.6216	0.0212
comp15933_c0_seq1	–	–	–	-4.8149	0.0212
comp1904_c1_seq2	*kif21a*	kinesin family member 21A-like	0.00E+00	-11.7087	0.0220
comp62390_c0_seq2	*scinla*	scinderin like a, also known as gelsolin-like	0.00E+00	-11.5304	0.0221
comp2808_c1_seq1	*rusc2*	iporin-like, also known as RUN and SH3 domain containing 2	2.17E-168	-3.1464	0.0228
comp28868_c0_seq2	*slc35e2b*	solute carrier family 35, member E2B	6.41E-66	11.2568	0.0235
comp23805_c0_seq11	*celf1*	cugbp, Elav-like family member 1-like	1.80E-155	5.3935	0.0238

comp7601_c0_seq3	*kcnma1*	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.00E+00	-11.3154	0.0239
comp11892_c0_seq2	*spns1*	spinster homolog 1	0.00E+00	-11.3107	0.0239
comp15198_c1_seq5	*slc44a5*	solute carrier family 44, member 5, also known as choline transporter-like protein 5-B-like	0.00E+00	-4.8485	0.0239
comp18739_c0_seq1	*adpgk*	ADP-dependent glucokinase-like	0.00E+00	2.6219	0.0239
comp2075_c4_seq61	–	–	–	10.4280	0.0241
comp24708_c0_seq13	*dlgap4*	discs, large (Drosophila) homolog-associated protein 4-like	6.24E-174	-11.5399	0.0249
comp17848_c0_seq2	*btd*	biotinidase-like	0.00E+00	-11.7924	0.0257
comp40929_c0_seq7	*sgcg*	sarcoglycan, gamma-like	2.16E-69	10.7062	0.0267
comp40132_c1_seq1	*efna2*	ephrin-A2-like	7.38E-93	3.1325	0.0270
comp4820_c1_seq3	*mapt*	microtubule-associated protein tau-like	3.86E-10	-11.4133	0.0271
comp4845_c0_seq2	–	–	–	-11.4517	0.0271
comp3506_c0_seq82	*fryl*	protein furry homolog-like	0.00E+00	9.3878	0.0273
comp203_c7_seq1	*ankrd34a*	ankyrin repeat domain 34a-like	0.00E+00	12.3617	0.0277
comp10511_c0_seq3	*cog8*	conserved oligomeric golgi complex subunit 8-like	0.00E+00	2.5960	0.0277
comp16711_c0_seq3	*axin2*	axin 2-like	0.00E+00	3.9224	0.0285
comp35332_c1_seq2	–	–	–	-11.2398	0.0285
comp1788_c0_seq17	*tenm1*	teneurin transmembrane protein 1	0.00E+00	2.3035	0.0288
comp36018_c0_seq7	*exd2*	exonuclease 3'-5' domain-containing protein 2-like	0.00E+00	12.2308	0.0293
comp7891_c0_seq2	*grhpr*	glyoxylate reductase/hydroxypyruvate reductase-like	0.00E+00	-11.4012	0.0293
comp31731_c0_seq19	*znf395*	zinc finger protein 395-like	1.80E-101	3.7665	0.0303
comp32985_c0_seq15	*als2*	alsin-like	0.00E+00	-11.9027	0.0304
comp37870_c0_seq3	*ramp2*	receptor (G protein-coupled) activity modifying protein 2-like	7.59E-74	-11.4634	0.0306
comp17550_c0_seq6	*rabl6*	RAB, member RAS oncogene family-like 6-like	4.84E-172	2.9489	0.0325
comp598_c2_seq3	*hecw2*	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2-like	2.92E-94	11.7177	0.0332
comp30524_c0_seq3	*gpn3*	GPN-loop GTPase 3-like	2.23E-105	10.6112	0.0338
comp15561_c0_seq1	*mch*	melanin concentrating hormone	1.77E-31	2.7855	0.0360
comp34557_c0_seq4	*aifm3*	apoptosis-inducing factor, mitochondrion-associated, 3	0.00E+00	-11.2505	0.0360
comp1796_c1_seq12	*clcn3*	chloride channel, voltage-sensitive 3	0.00E+00	3.6725	0.0362
comp56533_c0_seq2	–	–	–	10.4280	0.0362
comp15858_c0_seq1	–	–	–	-11.7979	0.0372
comp69178_c0_seq1	–	–	–	-11.1841	0.0375
comp11609_c0_seq15	*phlpp1*	PH domain leucine-rich repeat-containing protein phosphatase 1	0.00E+00	-3.7950	0.0378
comp17731_c0_seq2	*lonp2*	lon protease homolog 2, peroxisomal-like	0.00E+00	-11.4972	0.0378
comp11637_c0_seq6	*unc13a*	unc-13 homolog A-like	9.47E-150	-11.0236	0.0380
comp12457_c0_seq4	*cdk12*	cyclin-dependent kinase 12	0.00E+00	-11.8351	0.0380
comp1424_c1_seq9	*arhgap44*	Rho GTPase activating protein 44-like	5.08E-26	-1.9609	0.0393
comp4770_c1_seq1	*ap3d1*	adaptor-related protein complex 3, delta 1 subunit-like	0.00E+00	6.2896	0.0403
comp21500_c0_seq11	*mbtd1*	mbt domain containing protein 1-like	2.66E-32	10.5502	0.0405
comp2446_c0_seq5	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	6.13E-55	-14.0571	0.0415
comp766_c1_seq18	–	–	–	-6.0280	0.0415
comp4084_c1_seq3	*hnrnpul1*	heterogeneous nuclear ribonucleoprotein U-like 1	0.00E+00	-12.8492	0.0415
comp63417_c0_seq1	*aplf*	aprataxin and PNK-like factor-like	2.75E-82	4.0575	0.0418
comp9309_c0_seq2	*mark3*	MAP/microtubule affinity-regulating kinase 3-like	0.00E+00	-4.1286	0.0429
comp1070_c0_seq17	*scn2a*	sodium channel protein type 2 subunit alpha-like	0.00E+00	12.3387	0.0445
comp3905_c1_seq1	*cabin1*	calcineurin-binding protein 1-like	0.00E+00	-13.9093	0.0445
comp1521_c4_seq15	*magi1*	membrane associated guanylate kinase, WW and PDZ domain containing 1	0.00E+00	-2.5499	0.0447
comp15933_c0_seq2	–	–	–	7.6673	0.0454
comp13385_c1_seq6	*tjp1*	tight junction protein ZO-1-like	0.00E+00	2.4853	0.0461
comp54051_c0_seq1	*Actc1*	actin, alpha, cardiac muscle 1	5.72E-57	4.7878	0.0461
comp10677_c0_seq3	*rasgrp1*	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	3.32E-53	5.4178	0.0474
comp16474_c0_seq6	*ankrd28a*	ankyrin repeat domain 28a-like	0.00E+00	12.1241	0.0475
comp17498_c0_seq2	*ano7*	anoctamin 7	0.00E+00	-11.9861	0.0493
comp2134_c0_seq4	*fbxl16*	F-box and leucine-rich repeat protein 16-like	6.92E-115	-3.8821	0.0493
comp9691_c0_seq4	*c2cd5*	C2 calcium-dependent domain containing 5	0.00E+00	-12.1291	0.0493
comp22109_c0_seq1	–	–	–	-11.6542	0.0500
comp30776_c0_seq1	*faah2a*	fatty-acid amide hydrolase 2a-like	0.00E+00	3.3863	0.0500

Table S6 – List of differentially expressed contigs between nest-holder males and transitional males of *Salaria pavo*. Positive log2FoldChange indicates an up-regulation of the transcript for transitional males, whereas a negative log2FoldChange indicates an up-regulation of the transcript for nest-holder males.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp3407_c0_seq1	–	–	–	-13.4462	3.00E-10
comp16358_c1_seq2	*esyt2b*	extended synaptotagmin-like protein 2b-like	0.00E+00	-13.2803	6.41E-10
comp9075_c0_seq2	*tox4*	TOX high mobility group box family member 4-like	0.00E+00	12.9510	1.84E-09
comp3506_c0_seq82	*fryl*	protein furry homolog-like	0.00E+00	12.1661	1.94E-09
comp14527_c0_seq2	*rbbp6*	e3 ubiquitin-protein ligase RBBP6-like	3.44E-178	-7.3099	3.22E-09
comp700_c4_seq17	–	–	–	12.1554	2.15E-08
comp292_c6_seq16	*map4k4*	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	12.0688	3.29E-08
comp444_c0_seq25	*atp2b2*	ATPase, Ca++ transporting, plasma membrane 2	1.07E-49	-13.0080	5.36E-08
comp9075_c0_seq8	*tox2*	TOX high mobility group box family member 2-like	2.64E-109	-12.6159	1.00E-07
comp6074_c0_seq1	–	–	–	-14.6347	1.11E-07
comp3063_c0_seq2	–	–	–	-13.1566	4.96E-07

comp16173_c0_seq2	*cstf1*	cleavage stimulation factor subunit 1-like	0.00E+00	-12.4024	5.03E-07
comp25676_c0_seq3	*satb1*	DNA-binding protein SATB1-like	0.00E+00	6.0130	1.23E-06
comp2822_c1_seq9	*epb41*	erythrocyte membrane protein band 4.1	7.40E-48	-12.8868	1.52E-06
comp7272_c6_seq10	*dnmt3a*	DNA (cytosine-5-)-methyltransferase 3 alpha-like	0.00E+00	13.1110	2.37E-06
comp21801_c0_seq3	–	–	–	-6.9872	2.93E-06
comp72_c10_seq25	–	–	–	4.7078	3.72E-06
comp26306_c1_seq2	–	–	–	-12.2816	5.23E-06
comp29531_c0_seq12	*Till4*	tubulin tyrosine ligase-like family, member 4-like	0.00E+00	11.4921	6.45E-06
comp700_c4_seq15	*ptprd*	protein tyrosine phosphatase, receptor type, D-like	1.90E-72	-12.2765	6.76E-06
comp13104_c1_seq1	*nedd4l*	e3 ubiquitin-protein ligase nedd4-like	0.00E+00	-4.7428	8.29E-06
comp32242_c0_seq2	*znf592*	zinc finger protein 592-like	0.00E+00	-12.2291	1.14E-05
comp44962_c0_seq1	*rtel1*	regulator of telomere elongation helicase 1-like	0.00E+00	-12.0557	1.36E-05
comp11_c35_seq60	–	–	–	-12.5387	1.37E-05
comp15571_c0_seq2	*tsn8*	tetraspanin 8-like	3.56E-154	-11.9109	1.52E-05
comp26171_c0_seq2	*fmnl2*	formin-like protein 2-like	7.64E-155	11.2523	1.81E-05
comp10687_c0_seq2	*kt3k*	ketosamine-3-kinase-like	5.48E-154	12.3854	3.04E-05
comp14205_c0_seq3	*eif4g3*	eukaryotic translation initiation factor 4 gamma, 3-like	5.53E-86	11.6167	3.50E-05
comp26642_c0_seq8	*tmem62*	transmembrane protein 62-like	0.00E+00	11.2080	3.59E-05
comp11609_c0_seq15	*phlpp1*	PH domain leucine-rich repeat-containing protein phosphatase 1	0.00E+00	-12.2622	5.05E-05
comp93579_c0_seq2	–	–	–	11.3090	5.71E-05
comp896_c3_seq34	*capza2*	f-actin-capping protein subunit alpha-2-like	6.27E-146	-6.8074	8.01E-05
comp3852_c0_seq13	*arhgef9*	Cdc42 guanine nucleotide exchange factor (GEF) 9	0.00E+00	7.0994	1.05E-04
comp13731_c2_seq203	*csmd1*	CUB and sushi domain-containing protein 1	0.00E+00	-4.1991	1.13E-04
comp70843_c0_seq2	–	–	–	-11.6965	1.36E-04
comp9075_c0_seq4	*tox4*	TOX high mobility group box family member 4-like	0.00E+00	-4.2269	1.43E-04
comp94250_c0_seq1	–	–	–	11.2897	1.43E-04
comp14530_c0_seq1	*acad8*	acyl-CoA dehydrogenase family, member 8	0.00E+00	-4.5699	1.47E-04
comp264_c1_seq10	*cacnb4*	calcium channel, voltage-dependent, L type, beta 4 subunit-like	0.00E+00	-6.3805	2.06E-04
comp14699_c0_seq4	*srgap3*	SLIT-ROBO Rho GTPase activating protein 3	0.00E+00	11.8970	2.22E-04
comp5388_c0_seq4	*rap1gap*	RAP1 GTPase activating protein-like	2.68E-67	-3.5316	2.88E-04
comp21976_c1_seq1	–	–	–	-11.4613	3.26E-04
comp13368_c1_seq10	*dnah10*	dynein, axonemal, heavy chain 10-like	0.00E+00	4.7802	3.72E-04
comp29094_c0_seq1	*morc3*	MORC family CW-type zinc finger protein 3	0.00E+00	-4.4135	3.72E-04
comp30679_c0_seq1	–	–	–	-12.2463	3.72E-04
comp1107_c9_seq1	*ube2j1*	ubiquitin-conjugating enzyme E2, J1	2.44E-133	-12.5877	4.21E-04
comp2149_c0_seq6	*FAM13A*	family with sequence similarity 13, member A	1.12E-77	-11.5930	4.21E-04
comp5839_c0_seq35	*nfasc*	neurofascin-like	0.00E+00	11.6387	4.21E-04
comp16937_c0_seq2	*pak4*	p21 protein (Cdc42/Rac)-activated kinase 4, also known as serine/threonine-protein kinase PAK 7-like	1.12E-119	-11.4052	4.25E-04
comp3506_c0_seq37	*fryl*	protein furry homolog-like	0.00E+00	-13.9655	4.25E-04
comp6360_c0_seq31	*frmpd1*	FERM and PDZ domain-containing protein 1-like	0.00E+00	-3.4761	4.25E-04
comp7151_c1_seq2	*p4htm*	transmembrane prolyl 4-hydroxylase-like	8.46E-98	-11.3241	4.94E-04
comp3258_c0_seq8	–	–	–	-13.0983	5.07E-04
comp16003_c0_seq1	*slc39a6*	solute carrier family 39 (zinc transporter), member 6	0.00E+00	-6.0087	5.07E-04
comp12001_c1_seq4	*asl*	argininosuccinate lyase-like	4.21E-109	4.7922	5.35E-04
comp1792_c0_seq17	*sept8a*	septin 8a-like	0.00E+00	-3.9672	5.35E-04
comp4110_c3_seq3	*ncor2*	nuclear receptor corepressor 2-like	0.00E+00	2.8395	7.58E-04
comp5080_c1_seq9	*spef2*	sperm flagellar protein 2-like	1.60E-179	-11.3628	7.78E-04
comp12926_c0_seq7	–	–	–	-5.4647	8.16E-04
comp24803_c0_seq5	*bahcc1*	BAH and coiled-coil domain-containing protein 1-like	1.31E-127	10.6830	8.16E-04
comp23344_c0_seq2	–	–	–	10.9861	8.58E-04
comp37529_c0_seq1	–	–	–	-11.4517	8.61E-04
comp41090_c0_seq1	*plcd4*	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like	0.00E+00	-12.0797	8.61E-04
comp20893_c0_seq2	*traf2*	Tnf receptor-associated factor 2	6.93E-133	-11.7651	9.89E-04
comp9163_c1_seq50	–	–	–	11.7351	0.0011
comp2064_c0_seq7	*nr2f*	COUP transcription factor 2-like	2.59E-133	10.9766	0.0013
comp93799_c0_seq2	–	–	–	-11.4681	0.0013
comp16713_c0_seq5	*zcchc11*	terminal uridylyltransferase 4-like, alson known as zinc finger, CCHC domain containing 11	4.96E-31	-11.4830	0.0013
comp11430_c0_seq1	*acox3*	acyl-CoA oxidase 3, pristanoyl	0.00E+00	-11.2329	0.0015
comp1794_c1_seq4	*sv2c*	synaptic vesicle glycoprotein 2C-like	1.74E-99	4.4256	0.0017
comp2130_c1_seq16	*numa1*	nuclear mitotic apparatus protein 1-like	0.00E+00	-3.2729	0.0017
comp292_c6_seq6	*map4k4*	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	-5.1555	0.0017
comp12811_c0_seq1	*cybb*	cytochrome b-245, beta polypeptide	0.00E+00	-11.6667	0.0017
comp7342_c0_seq23	*tbl1xr1*	F-box-like/WD repeat-containing protein TBL1XR1-like	0.00E+00	-11.3410	0.0019
comp16838_c0_seq6	*map4k5*	mitogen-activated protein kinase kinase kinase kinase 5-like	0.00E+00	-2.6976	0.0019
comp2215_c0_seq18	*rims1*	regulating synaptic membrane exocytosis 1-like	0.00E+00	-11.6222	0.0020
comp23227_c0_seq2	*lig3*	DNA ligase 3-like	0.00E+00	-11.2090	0.0020
comp1124_c0_seq3	*tacc2*	transforming, acidic coiled-coil containing protein 2	0.00E+00	14.0918	0.0022
comp19267_c1_seq4	*kmt2c*	histone-lysine N-methyltransferase 2C-like	0.00E+00	3.1835	0.0022
comp2054_c6_seq75	–	–	–	11.7181	0.0022
comp44426_c0_seq3	*slc26a10*	solute carrier family 26, member 10-like	0.00E+00	-12.2547	0.0022
comp456_c4_seq4	*lgi1*	leucine-rich glioma-inactivated protein 1-like	0.00E+00	13.0021	0.0022
comp43041_c0_seq1	–	–	–	-4.1239	0.0023
comp3258_c0_seq29	–	–	–	-3.8035	0.0023
comp23255_c0_seq2	*bfar*	bifunctional apoptosis regulator	0.00E+00	-13.3294	0.0024
comp4845_c0_seq2	–	–	–	-11.4517	0.0024
comp40969_c0_seq2	*akap13*	A kinase (PRKA) anchor protein 13-like	0.00E+00	10.3140	0.0024
comp3506_c0_seq134	*fryl*	protein furry homolog-like	0.00E+00	-2.6541	0.0024
comp7891_c0_seq2	*grhpr*	glyoxylate reductase/hydroxypyruvate reductase-like	0.00E+00	-11.4012	0.0025

comp61831_c0_seq2	*gpr56*	G protein-coupled receptor 56-like	3.22E-119	-11.5982	0.0026
comp34557_c0_seq4	*aifm3*	apoptosis-inducing factor, mitochondrion-associated, 3	0.00E+00	-11.2505	0.0031
comp5979_c0_seq2	*sephs2*	selenide, water dikinase 2-like	5.71E-178	-11.2004	0.0031
comp3277_c1_seq42	*bahcc1*	BAH and coiled-coil domain-containing protein 1-like	0.00E+00	-4.2960	0.0034
comp18835_c0_seq4	*c1r*	complement component 1, r subcomponent	0.00E+00	-2.7856	0.0034
comp203_c7_seq1	*ankrd34a*	ankyrin repeat domain 34a-like	0.00E+00	13.2473	0.0034
comp55445_c0_seq2	–	–	–	-11.2426	0.0035
comp3506_c0_seq308	*fryl*	protein furry homolog-like	0.00E+00	13.1415	0.0036
comp17550_c0_seq6	*rabl6*	RAB, member RAS oncogene family-like 6-like	4.84E-172	3.3959	0.0036
comp1351_c0_seq2	*myef2*	myelin expression factor 2-like	4.82E-154	-3.5624	0.0036
comp2446_c0_seq36	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	-5.9702	0.0037
comp2075_c4_seq18	*cacna1h*	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	0.00E+00	2.9523	0.0038
comp35117_c1_seq17	*44M2.3*	putative RNA exonuclease NEF-sp-like	7.46E-138	-11.2216	0.0039
comp6622_c0_seq1	*snx4*	sorting nexin 4-like	0.00E+00	-11.1357	0.0039
comp9836_c0_seq4	*chd6*	chromodomain helicase-DNA-binding protein 6-like	0.00E+00	3.9394	0.0039
comp23548_c0_seq4	*atxn10*	ataxin 10-like	0.00E+00	-11.4735	0.0042
comp22690_c0_seq9	*hectd3*	HECT domain containing 3	0.00E+00	-5.4372	0.0043
comp15419_c1_seq1	*shank3*	SH3 and multiple ankyrin repeat domains 3	0.00E+00	-4.4138	0.0045
comp110833_c0_seq2	–	–	–	-11.3096	0.0048
comp18575_c0_seq2	*rgs3*	regulator of G-protein signaling 3-like	3.28E-49	-10.9573	0.0049
comp15046_c0_seq3	*sipa1l2*	signal-induced proliferation-associated 1-like protein 2-like	0.00E+00	-3.6208	0.0053
comp1351_c0_seq3	*myef2*	myelin expression factor 2-like	5.41E-158	-8.3733	0.0058
comp16424_c0_seq2	*rapgef4*	Rap guanine nucleotide exchange factor (GEF) 4-like	0.00E+00	-2.9459	0.0058
comp20194_c0_seq16	*ipo8*	importin 8-like	6.30E-114	10.8845	0.0060
comp26059_c1_seq118	*csmd3*	CUB and Sushi multiple domains 3-like	0.00E+00	-11.1467	0.0063
comp2068_c1_seq5	–	–	–	13.5870	0.0063
comp4437_c1_seq9	*cacna2d2*	calcium channel, voltage-dependent, alpha 2/delta subunit 2	0.00E+00	-11.7758	0.0063
comp131139_c1_seq1	*siglec10*	sialic acid-binding Ig-like lectin 10-like	0.00E+00	-11.0178	0.0064
comp476_c0_seq23	*nlrc3*	NLR family, CARD domain containing 3-like	2.84E-15	-11.1405	0.0068
comp1794_c1_seq5	*sv2c*	synaptic vesicle glycoprotein 2C-like	2.49E-100	-11.5229	0.0069
comp18465_c0_seq4	*cpsf3*	cleavage and polyadenylation specificity factor subunit 3-like	0.00E+00	-10.8033	0.0072
comp5839_c0_seq41	*nfasc*	neurofascin-like	0.00E+00	11.2903	0.0077
comp68_c2_seq6	*rhoa*	transforming protein RhoA-like	1.15E-139	-11.5451	0.0077
comp5666_c0_seq3	*srcin1*	SRC kinase signaling inhibitor 1	0.00E+00	3.9306	0.0078
comp1591_c3_seq2	*aste1*	asteroid homolog 1-like	0.00E+00	-13.0176	0.0082
comp32506_c0_seq1	*asb6*	ankyrin repeat and SOCS box protein 6-like	0.00E+00	-10.7787	0.0082
comp4084_c1_seq3	*hnrnpul1*	heterogeneous nuclear ribonucleoprotein U-like 1	0.00E+00	-12.8492	0.0083
comp12126_c0_seq6	*sntg1*	syntrophin, gamma 1-like	0.00E+00	3.0036	0.0085
comp36629_c0_seq1	–	–	–	-11.3860	0.0087
comp591_c0_seq1	*lhpp*	phospholysine phosphohistidine inorganic pyrophosphate phosphatase-like	4.39E-139	-12.0025	0.0091
comp5250_c0_seq2	*usp46*	ubiquitin carboxyl-terminal hydrolase 46-like	0.00E+00	-3.1886	0.0092
comp23065_c0_seq4	*tbc1d22b*	TBC1 domain family, member 22B-like	0.00E+00	-11.5830	0.0095
comp10677_c0_seq3	*rasgrp1*	RAS guanyl releasing protein 1 (calcium and DAG-regulated)	3.32E-53	5.5369	0.0098
comp292_c6_seq11	*map4k4*	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	-12.0974	0.0098
comp14233_c1_seq3	*crat*	arnitine O-acetyltransferase-like	0.00E+00	-10.7795	0.0100
comp6182_c0_seq1	*rasgrf1*	Ras-specific guanine nucleotide-releasing factor 1-like	0.00E+00	4.5944	0.0100
comp8826_c0_seq3	*b3gat1*	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1-like	1.55E-28	-10.7741	0.0100
comp24063_c1_seq1	*pde9a*	high affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A	0.00E+00	-5.0606	0.0102
comp58690_c0_seq1	*ergic3*	endoplasmic reticulum-Golgi intermediate compartment protein 3-like (ERGIC and golgi 3)	0.00E+00	-11.3194	0.0102
comp2446_c0_seq5	*cacna1a*	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	6.13E-55	-14.0571	0.0106
comp44046_c0_seq1	*gnrhr*	gonadotropin-releasing hormone receptor	4.02E-49	-4.7492	0.0109
comp80627_c0_seq2	–	–	–	-11.5856	0.0112
comp24604_c0_seq3	*neto1*	neuropilin (NRP) and tolloid (TLL)-like 1	0.00E+00	-11.4722	0.0114
comp20054_c0_seq2	*tha1*	threonine aldolase 1	0.00E+00	-10.6780	0.0115
comp19894_c0_seq2	–	–	–	-11.1150	0.0121
comp6374_c0_seq2	*zfyve1*	zinc finger FYVE domain-containing protein 1-like	0.00E+00	-3.3867	0.0124
comp112_c4_seq75	–	–	–	-12.0320	0.0131
comp22544_c0_seq1	*kiaa0907*	UPF0469 protein KIAA0907 homolog	5.56E-130	-3.3354	0.0131
comp9163_c1_seq48	–	–	–	-11.0605	0.0138
comp7143_c1_seq5	*lrrc7*	leucine-rich repeat-containing protein 7-like	0.00E+00	-4.8332	0.0140
comp44326_c0_seq2	–	–	–	-11.7335	0.0141
comp3442_c0_seq8	*sema6dl*	semaphorin 6D-like	0.00E+00	-12.2678	0.0148
comp2645_c4_seq148	*cacna1d*	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-11.3918	0.0156
comp15833_c0_seq2	*cerk*	ceramide kinase-like	0.00E+00	3.3421	0.0157
comp8820_c0_seq5	*rictor*	RPTOR independent companion of MTOR, complex 2	0.00E+00	-12.6840	0.0163
comp11526_c0_seq15	*gab1*	GRB2-associated-binding protein 1	3.28E-86	-3.1349	0.0165
comp44541_c0_seq1	–	–	–	-11.2972	0.0165
comp17250_c1_seq2	*rock2*	rho-associated protein kinase 2-like	0.00E+00	-3.2892	0.0167
comp5560_c2_seq13	*ighm*	immunoglobulin heavy constant mu	1.69E-165	-2.4818	0.0174
comp2645_c4_seq139	*cacna1d*	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-6.1771	0.0179
comp16358_c1_seq5	*esyt2b*	extended synaptotagmin-like protein 2b-like	0.00E+00	3.4322	0.0179
comp11_c35_seq7	*cdk17*	cyclin-dependent kinase 17-like	0.00E+00	-2.8317	0.0184
comp8947_c1_seq1	*tgm3*	protein-glutamine gamma-glutamyltransferase E-like	0.00E+00	-12.0498	0.0190
comp20896_c1_seq2	–	–	–	4.1026	0.0191
comp39472_c1_seq3	*zbtb22*	zinc finger and BTB domain-containing protein 22-like	5.79E-147	4.3904	0.0212
comp40032_c0_seq2	*exosc7*	exosome complex component RRP42-like	0.00E+00	-10.6394	0.0212
comp4978_c2_seq2	*f13a*	coagulation factor XIII A chain-like	0.00E+00	-11.9602	0.0212
comp2075_c4_seq61	–	–	–	9.9514	0.0213
comp33_c0_seq4	–	–	–	-2.4934	0.0224

comp21177_c0_seq3	*hagh*	hydroxyacylglutathione hydrolase-like	3.01E-80	-11.1956	0.0238
comp6537_c0_seq2	*mtpap*	mitochondrial poly(A) polymerase-like	0.00E+00	-11.1751	0.0238
comp18742_c0_seq3	*tmem154*	transmembrane protein 154-like	1.70E-29	-4.9849	0.0239
comp3402_c3_seq18	*mef2d*	myocyte-specific enhancer factor 2D homolog	4.93E-90	-10.7282	0.0241
comp46259_c0_seq1	*dgat1*	diacylglycerol O-acyltransferase 1-like	0.00E+00	-11.3941	0.0255
comp29958_c0_seq3	*kctd3*	BTB/POZ domain-containing protein KCTD3-like, also known as potassium channel tetramerization domain containing 3	0.00E+00	-11.1321	0.0256
comp108218_c0_seq1	–	–	–	-10.8268	0.0258
comp3754_c1_seq3	*ppfibp1*	liprin beta 1-like, also known as PTPRF interacting protein, binding protein 1	0.00E+00	-11.7856	0.0259
comp40929_c0_seq7	*sgcg*	sarcoglycan, gamma-like	2.16E-69	10.2523	0.0259
comp7496_c4_seq3	*lrrfip2*	leucine rich repeat (in FLII) interacting protein 2	4.22E-167	-10.6272	0.0259
comp6269_c0_seq2	*rbfox1*	RNA binding protein, fox-1 homolog 1	4.76E-16	-3.0523	0.0269
comp27673_c1_seq1	*yrdc*	yrdC domain-containing protein, mitochondrial-like	5.60E-71	-11.1757	0.0270
comp27920_c0_seq2	–	–	–	-10.9400	0.0271
comp68_c2_seq5	*rhoa*	transforming protein RhoA-like	6.68E-136	6.4188	0.0272
comp13493_c0_seq11	*ndufs1*	NADH dehydrogenase (ubiquinone) Fe-S protein 1	0.00E+00	-12.1106	0.0273
comp33_c0_seq6	–	–	–	-2.8644	0.0273
comp16012_c0_seq1	*ctnnd1*	catenin (cadherin-associated protein), delta 1-like	0.00E+00	-11.9121	0.0277
comp20660_c0_seq37	*mdga2*	MAM domain-containing glycosylphosphatidylinositol anchor protein 2	2.62E-57	-10.7258	0.0286
comp24037_c0_seq2	–	–	–	-2.7438	0.0295
comp5839_c0_seq25	*nfasc*	neurofascin-like	0.00E+00	11.0952	0.0296
comp17707_c1_seq4	*col12a1*	collagen, type XII, alpha 1-like	6.43E-65	-11.2487	0.0298
comp971_c7_seq1	*coa6*	cytochrome c oxidase assembly factor 6 homolog	2.51E-34	-10.6465	0.0298
comp5560_c2_seq18	*ighm*	immunoglobulin heavy constant mu	1.54E-98	-11.1303	0.0311
comp9012_c0_seq4	*lrrc7*	leucine-rich repeat-containing protein 7	2.42E-24	-10.6898	0.0313
comp223_c0_seq2	*GTH-alpha*	gonadotropin alpha subunit, also known as glycoprotein hormones, alpha polypeptide (cga)	4.74E-48	-2.7966	0.0318
comp23114_c0_seq5	*ift172*	intraflagellar transport protein 172 homolog	0.00E+00	-4.4725	0.0324
comp8108_c0_seq2	–	–	–	-14.1408	0.0327
comp24708_c0_seq4	*dlgap4*	discs, large (Drosophila) homolog-associated protein 4-like	0.00E+00	-3.3902	0.0342
comp11119_c0_seq22	*cdkl5*	cyclin-dependent kinase-like 5-like	0.00E+00	11.6904	0.0342
comp15966_c0_seq1	*tshb*	thyroid stimulating hormone, beta subunit	7.31E-39	-2.4884	0.0350
comp4770_c1_seq1	*ap3d1*	adaptor-related protein complex 3, delta 1 subunit-like	0.00E+00	6.2809	0.0351
comp61213_c0_seq2	*itpkc*	inositol-trisphosphate 3-kinase C-like	1.15E-170	-10.7071	0.0352
comp27_c2_seq3	*atp1a1*	ATPase, Na+/K+ transporting, alpha 1-like	1.54E-114	-11.4486	0.0367
comp8780_c0_seq9	*Rap1gap2*	RAP1 GTPase activating protein 2	1.61E-141	10.4497	0.0367
comp18827_c1_seq3	–	–	–	10.1577	0.0374
comp11052_c0_seq2	*rpz3*	rapunzel 3	0.00E+00	-3.2051	0.0392
comp52653_c0_seq1	*spata18*	Spermatogenesis associated 18, also known as mitochondria-eating protein-like	0.00E+00	-11.4454	0.0396
comp10036_c0_seq1	*lipin2*	phosphatidate phosphatase lpin2-like	0.00E+00	-3.2030	0.0397
comp18952_c0_seq2	*golga2*	golgin subfamily A member 2-like	0.00E+00	-2.8688	0.0397
comp3567_c5_seq6	–	–	–	-2.3416	0.0397
comp39557_c0_seq2	–	–	–	4.7217	0.0397
comp4627_c0_seq15	*map2*	microtubule-associated protein 2-like	8.97E-106	4.4555	0.0397
comp54866_c0_seq4	*lpin2*	phosphatidate phosphatase LPIN2-like	3.09E-166	-10.8900	0.0397
comp11489_c0_seq8	*erc2*	ELKS/RAB6-interacting/CAST family member 2	0.00E+00	-3.1226	0.0398
comp4939_c0_seq1	*grl1*	gamma-aminobutyric acid receptor-associated protein-like 1-like	1.27E-23	-10.4586	0.0402
comp20893_c0_seq1	*traf2*	Tnf receptor-associated factor 2	0.00E+00	10.5887	0.0410
comp67298_c0_seq3	*arap2*	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2-like	0.00E+00	-11.5259	0.0419
comp73909_c0_seq1	*map3k19*	mitogen-activated protein kinase kinase kinase 19	0.00E+00	-3.2939	0.0419
comp983_c0_seq1	*ddx17*	DEAD (Asp-Glu-Ala-Asp) box helicase 17	0.00E+00	-13.5430	0.0419
comp30391_c0_seq1	*mx*	interferon-induced GTP-binding protein Mx-like	0.00E+00	-2.5360	0.0427
comp3599_c0_seq3	*adcyap1*	adenylate cyclase activating polypeptide 1 (pituitary), glucagon family neuropeptides	6.89E-83	-12.1629	0.0440
comp7477_c0_seq1	*otof*	otoferlin-like	0.00E+00	-2.4283	0.0449
comp22347_c0_seq1	*ngb*	neuroglobin-like	1.37E-88	-11.0946	0.0453
comp7303_c2_seq7	*myt1a*	myelin transcription factor 1-like	0.00E+00	-3.3719	0.0465
comp19616_c1_seq5	*fli1*	retroviral integration site protein Fli-1 homolog	1.05E-97	-11.1728	0.0478
comp8407_c0_seq2	*svop*	synaptic vesicle 2-related protein-like	0.00E+00	-2.9701	0.0479
comp8139_c3_seq5	*uba1*	ubiquitin-like modifier-activating enzyme 1-like	0.00E+00	-11.4667	0.0479
comp15858_c0_seq1	–	–	–	-4.7966	0.0487
comp9181_c0_seq15	*fam171b*	protein FAM171B-like	0.00E+00	11.3665	0.0487
comp14814_c3_seq17	*fcho2*	FCH domain only protein 2-like	1.38E-117	-11.9959	0.0491
comp20190_c0_seq3	*fam21c*	WASH complex subunit FAM21C-like	2.15E-78	-5.5535	0.0491
comp35839_c1_seq1	*limk1*	LIM domain kinase 1-like	0.00E+00	-2.7457	0.0491
comp69795_c0_seq3	*or52k1*	olfactory receptor 52K1-like	1.12E-76	10.3111	0.0491
comp6345_c0_seq4	*angl4*	angiopoietin-related protein 4-like	1.02E-177	-2.2476	0.0492
comp19020_c0_seq3	*prss12*	protease, serine, 12-like	2.79E-47	-11.0242	0.0498

Table S7 – List of differentially expressed contigs between females and sneaker males of *Salaria pavo*. Positive log2FoldChangE indicates an up-regulation of the transcript for sneaker males, whereas a negative log2FoldChange indicates an up-regulation of the transcript for females.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp7003_c0_seq3	fam20c	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	14.9317	7.31E-19
comp2816_c0_seq2	lphn3	latrophilin 3-like	0.00E+00	-14.6001	3.09E-13

comp1424_c1_seq3	arhgap44	Rho GTPase activating protein 44-like	9.68E-42	-13.9867	9.98E-11
comp19398_c0_seq1	setd2	histone-lysine N-methyltransferase SETD2-like	0.00E+00	-14.2173	6.95E-10
comp26241_c0_seq3	nhs	Nance-Horan syndrome protein-like	0.00E+00	-13.8213	6.95E-10
comp7496_c4_seq1	lrrfip2	leucine rich repeat (in FLII) interacting protein 2	4.82E-159	-13.6696	6.95E-10
comp20540_c0_seq1	pi4k2a	phosphatidylinositol 4-kinase type 2 alpha-like	0.00E+00	-14.3276	8.72E-10
comp3506_c0_seq134	fryl	protein furry homolog-like	0.00E+00	-13.5706	2.27E-09
comp22092_c0_seq2	–	–	–	7.5082	3.88E-09
comp3813_c0_seq4	znf207	zinc finger protein 207-like	4.88E-71	-14.6265	3.88E-09
comp4000_c4_seq5	atp1a3b	ATPase, Na+/K+ transporting, alpha 3 polypeptide-like	0.00E+00	13.2736	3.88E-09
comp7003_c0_seq5	fam20c	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	-13.5231	4.91E-09
comp21022_c0_seq8	bptf	nucleosome-remodeling factor subunit BPTF-like	0.00E+00	-13.4891	1.09E-08
comp7069_c0_seq4	st7	suppression of tumorigenicity 7 protein homolog	0.00E+00	14.7997	1.15E-08
comp7689_c1_seq2	crebzf	CREB/ATF bZIP transcription factor-like	2.75E-56	-13.4060	1.15E-08
comp33_c0_seq5	–	–	–	-5.0987	2.99E-08
comp3506_c0_seq82	fryl	protein furry homolog-like	0.00E+00	-5.4788	3.44E-08
comp643_c1_seq4	–	–	–	-13.3263	5.12E-08
comp16233_c0_seq2	tdrd3	tudor domain-containing protein 3-like	0.00E+00	12.6146	1.04E-07
comp765_c0_seq1	rnf10	ring finger protein 10-like	0.00E+00	-13.0451	1.04E-07
comp451_c4_seq1	–	–	–	-13.0435	1.15E-07
comp13731_c2_seq198	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	12.7946	1.27E-07
comp1127_c4_seq2	iqsec2	IQ motif and SEC7 domain-containing protein 2-like	0.00E+00	-6.3656	2.18E-07
comp6820_c0_seq1	–	–	–	-14.0763	2.18E-07
comp6311_c0_seq37	adam23	ADAM metallopeptidase domain 23	0.00E+00	-13.3302	2.40E-07
comp13993_c0_seq128	prune2	prune homolog 2	0.00E+00	-13.1151	2.42E-07
comp39182_c0_seq36	erbb4	receptor tyrosine-protein kinase erbB-4-like	0.00E+00	12.2113	3.73E-07
comp34961_c0_seq9	fhdc1	FH2 domain-containing protein 1-like	0.00E+00	12.3195	6.12E-07
comp2446_c0_seq48	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	2.45E-150	-12.7844	7.89E-07
comp33_c0_seq4	–	–	–	-3.8962	1.31E-06
comp24068_c0_seq3	–	–	–	12.3988	1.83E-06
comp22092_c0_seq1	–	–	–	-13.4092	2.00E-06
comp11119_c0_seq12	cdkl5	cyclin-dependent kinase-like 5-like	0.00E+00	-12.6577	2.25E-06
comp23805_c0_seq5	celf1	cugbp, Elav-like family member 1-like	0.00E+00	11.5900	3.67E-06
comp22879_c1_seq3	glyatl1	glycine N-acyltransferase-like	4.79E-76	12.5575	4.56E-06
comp30660_c0_seq3	wdr20	WD repeat-containing protein 20-like	0.00E+00	11.9547	5.24E-06
comp2319_c1_seq4	–	–	–	12.3825	6.43E-06
comp10558_c0_seq46	kcnt1	potassium channel, subfamily T, member 1	0.00E+00	-13.3105	6.65E-06
comp17250_c1_seq26	dgcr2	integral membrane protein DGCR2/IDD-like, also known as DiGeorge syndrome critical region gene 2	0.00E+00	12.6286	8.74E-06
comp3480_c0_seq3	–	–	–	12.8897	9.10E-06
comp7069_c0_seq2	rnf6	ring finger protein (C3H2C3 type) 6	2.69E-10	14.8343	9.48E-06
comp21424_c0_seq1	pde7b	cAMP-specific 3',5'-cyclic phosphodiesterase 7B-like	0.00E+00	-12.8506	1.05E-05
comp3506_c0_seq319	fryl	protein furry homolog-like	1.21E-100	-4.7357	1.21E-05
comp10728_c1_seq2	–	–	–	-12.3869	1.37E-05
comp2446_c0_seq36	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	-12.9237	1.61E-05
comp5080_c1_seq9	spef2	sperm flagellar protein 2-like	1.60E-179	-12.6888	1.61E-05
comp6389_c1_seq2	tle2	transducin-like enhancer protein 4-like	0.00E+00	12.1093	2.49E-05
comp3258_c0_seq2	–	–	–	-5.5421	2.68E-05
comp16767_c1_seq2	–	–	–	-4.2790	2.86E-05
comp11489_c0_seq8	erc2	ELKS/RAB6-interacting/CAST family member 2	0.00E+00	-5.0735	3.49E-05
comp10355_c0_seq3	phf2	lysine-specific demethylase phf2-like (PHD finger protein 2)	0.00E+00	-12.9958	4.20E-05
comp26840_c0_seq1	slc6a	sodium-dependent serotonin transporter-like, also known as solute carrier family 6 (neurotransmitter transporter)	0.00E+00	4.4001	4.48E-05
comp4900_c1_seq16	–	–	–	-12.2349	7.11E-05
comp16520_c0_seq1	chd6	chromodomain helicase DNA binding protein 6-like	1.11E-78	-12.7471	9.19E-05
comp61048_c0_seq2	–	–	–	12.3396	1.14E-04
comp3266_c0_seq3	fam184a	protein FAM184A-like	0.00E+00	3.9419	1.15E-04
comp2442_c0_seq1	–	–	–	3.2604	1.28E-04
comp4650_c1_seq3	gpc6	glypican 6-like	0.00E+00	-12.1228	1.47E-04
comp10086_c0_seq5	fbp	f-type lectin 3 (fucolectins)	1.17E-137	-14.2504	1.50E-04
comp29531_c0_seq12	Ttll4	tubulin tyrosine ligase-like family, member 4-like	0.00E+00	-12.3342	1.54E-04
comp31969_c0_seq3	ttll3	tubulin tyrosine ligase-like family, member 3-like	0.00E+00	-5.7092	1.54E-04
comp16106_c0_seq1	vps29	vacuolar protein sorting-associated protein 29-like	7.30E-123	-3.8944	1.55E-04
comp33_c0_seq6	–	–	–	-3.9463	1.67E-04
comp7149_c0_seq3	wbp11	WW domain-binding protein 11-like	1.48E-70	-12.1198	1.83E-04
comp8977_c0_seq2	isoc2	isochorismatase domain-containing protein 2, mitochondrial-like	1.20E-122	3.7529	1.90E-04
comp2931_c1_seq2	gabrg1	gamma-aminobutyric acid (GABA) receptor subunit gamma-1-like	0.00E+00	-12.2248	2.46E-04
comp8737_c0_seq13	trim3	tripartite motif-containing protein 3-like	0.00E+00	-12.8803	3.06E-04
comp10558_c0_seq40	kcnt1	potassium channel, subfamily T, member 1	0.00E+00	-12.1533	3.07E-04
comp11637_c0_seq6	unc13a	unc-13 homolog A-like	9.47E-150	-12.0469	3.85E-04
comp17129_c0_seq1	–	–	–	-11.9854	3.85E-04
comp5180_c0_seq3	–	–	–	-3.4800	4.38E-04
comp16332_c0_seq2	ophn1	oligophrenin 1-like	0.00E+00	-12.1099	4.51E-04
comp42_c0_seq1	–	–	–	-3.5814	4.51E-04
comp1511_c1_seq9	ncam1	neural cell adhesion molecule 1-like	5.13E-15	-5.6049	4.89E-04
comp12377_c0_seq8	–	–	–	-11.9270	4.91E-04
comp55090_c0_seq1	–	–	–	-12.0286	4.93E-04
comp36143_c1_seq1	dgcr6	DiGeorge syndrome critical region gene 6-like	5.36E-84	11.4846	5.79E-04
comp3650_c0_seq1	crk	crk-like protein-like	1.39E-147	-12.9601	5.87E-04
comp2645_c4_seq148	cacna1d	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-13.0498	7.90E-04
comp22472_c0_seq2	–	–	–	-12.8105	8.11E-04

comp7807_c0_seq7	akap9	A kinase (PRKA) anchor protein 9-like	0.00E+00	-13.7196	8.59E-04
comp5109_c0_seq2	–	–	–	11.3999	8.60E-04
comp3753_c2_seq5	ptpre	receptor-type tyrosine-protein phosphatase E-like	5.80E-150	-12.4167	9.33E-04
comp24068_c0_seq2	–	–	–	-11.9429	9.41E-04
comp9222_c1_seq1	fez2	fasciculation and elongation protein zeta-2-like (Zygin II)	4.02E-150	-12.5581	9.41E-04
comp32242_c0_seq2	znf592	zinc finger protein 592-like	0.00E+00	-11.8849	0.0011
comp24708_c0_seq4	dlgap4	discs, large (Drosophila) homolog-associated protein 4-like	0.00E+00	4.3258	0.0011
comp7151_c1_seq2	p4htm	transmembrane prolyl 4-hydroxylase-like	8.46E-98	-3.4177	0.0011
comp7367_c1_seq5	ptprk	protein tyrosine phosphatase, receptor type, K-like	0.00E+00	-2.4425	0.0011
comp20054_c0_seq3	tha1	threonine aldolase 1	8.85E-161	-11.8603	0.0012
comp18110_c0_seq3	pom121	nuclear envelope pore membrane protein POM121-like	0.00E+00	13.3535	0.0012
comp40969_c0_seq2	akap13	A kinase (PRKA) anchor protein 13-like	0.00E+00	4.4132	0.0012
comp52574_c0_seq11	znf236	zinc finger protein 236-like	0.00E+00	10.9997	0.0012
comp3334_c0_seq8	tom1l2	target of myb1-like protein 2-like	2.29E-34	13.9189	0.0012
comp6412_c0_seq3	galnt18	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 18-like	0.00E+00	-12.9367	0.0014
comp3813_c0_seq6	znf207	zinc finger protein 207-like	1.44E-71	14.2722	0.0014
comp20780_c0_seq7	plekho1a	pleckstrin homology domain containing, family O member 1a-like	6.96E-133	-11.8595	0.0014
comp6371_c1_seq11	mical3	protein-methionine sulfoxide oxidase mical3a-like, also known as microtubule associated monooxygenase, calponin and LIM domain containing 3	0.00E+00	-11.6750	0.0015
comp27984_c0_seq1	calm	calmodulin-like	5.42E-15	3.6467	0.0015
comp17106_c0_seq3	cntn3	contactin 3	0.00E+00	-11.9272	0.0015
comp2852_c1_seq4	ndufaf7	NADH dehydrogenase [ubiquinone] complex I, assembly factor 7	0.00E+00	-4.7895	0.0015
comp80286_c0_seq2	tnnt1	troponin T, slow skeletal muscle-like	4.94E-10	11.2613	0.0015
comp123221_c0_seq2	mgp	matrix Gla protein	5.67E-49	11.4260	0.0016
comp131394_c0_seq1	frmd3	FERM domain-containing protein 3	1.11E-38	11.1926	0.0016
comp92463_c0_seq2	rsph10b	radial spoke head 10 homolog B-like	0.00E+00	11.3290	0.0017
comp41090_c0_seq2	plcd4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like	0.00E+00	-11.8890	0.0017
comp18295_c0_seq1	dock9	dedicator of cytokinesis protein 9-like	0.00E+00	3.4529	0.0017
comp17848_c0_seq2	btd	biotinidase-like	0.00E+00	-12.5643	0.0017
comp1934_c0_seq2	–	–	–	-12.1012	0.0018
comp7178_c5_seq1	prickle2	prickle homolog 2	0.00E+00	3.4697	0.0018
comp6263_c1_seq3	ei24	etoposide-induced protein 2.4 homolog	1.60E-162	-11.9328	0.0019
comp13748_c0_seq1	znf831	zinc finger protein 831	2.13E-107	-2.7507	0.0019
comp21414_c0_seq1	hsd17b12a	estradiol 17-beta-dehydrogenase 12-A, also known as hydroxysteroid (17-beta) dehydrogenase 12a	1.75E-91	11.2613	0.0021
comp35039_c0_seq2	dapk2	death-associated protein kinase 2-like	5.18E-145	-11.5949	0.0022
comp45141_c0_seq2	–	–	–	-11.6148	0.0022
comp24887_c0_seq2	lcorl	ligand-dependent nuclear receptor corepressor-like protein	6.68E-27	12.7878	0.0024
comp1842_c1_seq19	dst	dystonin	0.00E+00	2.6051	0.0027
comp1521_c4_seq15	magi1	membrane associated guanylate kinase, WW and PDZ domain containing 1	0.00E+00	-3.0522	0.0029
comp3905_c1_seq20	cabin1	calcineurin-binding protein 1-like	0.00E+00	12.8070	0.0031
comp3852_c0_seq10	arhgef9	Cdc42 guanine nucleotide exchange factor (GEF) 9	0.00E+00	-3.1897	0.0035
comp28508_c1_seq1	–	–	–	10.5133	0.0035
comp36508_c0_seq4	fxr1	fragile X mental retardation syndrome-related protein 1-like	6.82E-104	11.2923	0.0035
comp20824_c0_seq1	aldh2	aldehyde dehydrogenase 2	0.00E+00	2.8878	0.0039
comp15933_c0_seq1	–	–	–	-5.1862	0.0040
comp19489_c0_seq1	tnrc6b	trinucleotide repeat containing 6b-like	3.36E-125	-3.4755	0.0040
comp94169_c0_seq1	–	–	–	4.2323	0.0040
comp5080_c1_seq13	spef2	sperm flagellar protein 2-like	1.20E-98	-11.6227	0.0040
comp15373_c0_seq2	agrn	agrin-like	0.00E+00	-12.3141	0.0042
comp3506_c0_seq318	fryl	protein furry homolog-like	2.29E-114	5.7364	0.0042
comp26691_c0_seq2	–	–	–	-11.4371	0.0044
comp34138_c1_seq1	–	–	–	-11.7347	0.0046
comp2189_c1_seq26	rbm6	RNA-binding motif protein 6-like	1.07E-128	-2.6496	0.0046
comp4829_c1_seq3	wdr54	WD repeat domain 54	0.00E+00	-11.4827	0.0048
comp14932_c2_seq93	trpm3	transient receptor potential cation channel, subfamily M, member 3	0.00E+00	12.0036	0.0048
comp9523_c0_seq11	vps13c	vacuolar protein sorting 13 homolog C-like	0.00E+00	12.2193	0.0050
comp8725_c1_seq2	gtpbp1	GTP binding protein 1-like	0.00E+00	-11.4467	0.0054
comp35117_c1_seq17	44M2.3	putative RNA exonuclease NEF-sp-like	7.46E-138	4.9587	0.0054
comp11606_c0_seq4	hmcn2	hemicentin-2-like	3.73E-72	-3.7118	0.0054
comp1571_c0_seq4	kcnd3	potassium voltage-gated channel, Shal-related subfamily, member 3	0.00E+00	-3.1992	0.0055
comp24280_c1_seq7	zmym4	zinc finger, MYM-type 4-like	0.00E+00	-4.9204	0.0056
comp7477_c0_seq1	otof	otoferlin-like	0.00E+00	-3.3604	0.0057
comp13731_c2_seq140	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	-3.0954	0.0059
comp34557_c0_seq4	aifm3	apoptosis-inducing factor, mitochondrion-associated, 3	0.00E+00	-11.6662	0.0060
comp38328_c0_seq2	s100a1	protein S100-A1-like	4.81E-50	-4.6237	0.0060
comp22879_c1_seq1	mgme1	mitochondrial genome maintenance exonuclease 1-like	0.00E+00	-3.8711	0.0069
comp5520_c0_seq5	–	–	–	-12.7320	0.0069
comp55390_c0_seq1	–	–	–	-11.8455	0.0070
comp56902_c0_seq2	–	–	–	11.4360	0.0071
comp595_c10_seq8	ints3	integrator complex subunit 3-like	0.00E+00	-13.4631	0.0071
comp45523_c0_seq1	pms1	PMS1 protein homolog 1-like	0.00E+00	10.4400	0.0072
comp3346_c0_seq1	–	–	–	4.6489	0.0072
comp1424_c1_seq9	arhgap44	Rho GTPase activating protein 44-like	5.08E-26	-2.2209	0.0076
comp130526_c0_seq2	–	–	–	5.7173	0.0077
comp15331_c0_seq5	gulp1	GULP, engulfment adaptor PTB domain containing protein 1-like	0.00E+00	-12.1701	0.0081
comp2075_c4_seq32	cacna1h	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	9.38E-147	2.7448	0.0088
comp33263_c0_seq3	gdap2	ganglioside induced differentiation associated protein 2-like	0.00E+00	-12.1063	0.0088
comp130808_c0_seq1	–	–	–	-11.4033	0.0088

comp56560_c0_seq2	–	–	–	11.3098	0.0092
comp31470_c1_seq2	–	–	–	-11.6198	0.0094
comp33_c0_seq3	–	–	–	-3.9268	0.0094
comp5389_c0_seq2	ylpm1	YLP motif-containing protein 1-like	0.00E+00	-13.4010	0.0098
comp5682_c0_seq7	kif5c	kinesin family member 5C-like	0.00E+00	-12.3466	0.0099
comp17880_c0_seq6	qrich1	glutamine-rich protein 1-like	0.00E+00	4.7883	0.0101
comp22109_c0_seq1	–	–	–	-12.1219	0.0107
comp7739_c0_seq28	tcf4	transcription factor 4-like	0.00E+00	-11.8924	0.0112
comp23426_c0_seq2	nfrkb	nuclear factor related to kappa-B-binding protein-like	0.00E+00	-12.8123	0.0115
comp42933_c0_seq3	tmem150c	transmembrane protein 150C-like	9.74E-70	12.0288	0.0115
comp275_c0_seq6	–	–	–	3.5214	0.0117
comp55736_c0_seq1	–	–	–	-11.3992	0.0121
comp17540_c1_seq4	pdcd4	programmed cell death protein 4-like	0.00E+00	-11.3433	0.0126
comp3060_c2_seq4	–	–	–	11.1974	0.0128
comp25278_c0_seq43	hspg2	heparan sulfate proteoglycan 2	0.00E+00	-11.5171	0.0130
comp34961_c0_seq8	fhdc1	FH2 domain-containing protein 1-like	0.00E+00	-3.2716	0.0130
comp23703_c0_seq8	limch1	LIM and calponin homology domains-containing protein 1-like	5.11E-71	11.5410	0.0134
comp23974_c0_seq1	–	–	–	3.5699	0.0140
comp14289_c0_seq3	ankrd13c	ankyrin repeat domain 13C-like	1.32E-122	11.1643	0.0142
comp12087_c0_seq2	–	–	–	-12.0687	0.0153
comp24200_c0_seq4	sipa1l3	signal-induced proliferation-associated 1 like 3-like	0.00E+00	-2.5273	0.0164
comp19865_c0_seq3	chat	choline O-acetyltransferase-like	0.00E+00	-11.2628	0.0172
comp4648_c0_seq1	actb	beta-actin	6.35E-32	2.6909	0.0172
comp13097_c0_seq6	pogz	pogo transposable element with ZNF domain	6.20E-102	-4.6990	0.0172
comp6080_c3_seq26	arhgef11	Rho guanine nucleotide exchange factor (GEF) 11-like	0.00E+00	-11.6324	0.0172
comp35531_c0_seq3	nt5m	5'(3')-deoxyribonucleotidase, mitochondrial-like	1.78E-96	11.3376	0.0176
comp200_c1_seq1	chchd3	coiled-coil-helix-coiled-coil-helix domain containing 3	1.23E-14	11.3586	0.0177
comp1842_c1_seq17	dst	dystonin	0.00E+00	2.8545	0.0182
comp20190_c0_seq3	fam21c	WASH complex subunit FAM21C-like	2.15E-78	10.6629	0.0183
comp74112_c0_seq1	–	–	–	11.6268	0.0183
comp11815_c0_seq1	–	–	–	2.6434	0.0184
comp6404_c0_seq5	ces1	liver carboxylesterase 1-like	9.54E-57	-11.3901	0.0185
comp12835_c0_seq7	ccser2	coiled-coil serine-rich protein 2-like	0.00E+00	-13.1723	0.0186
comp11892_c0_seq2	spns1	spinster homolog 1	0.00E+00	-11.2544	0.0187
comp6103_c0_seq3	gramd1a	GRAM domain-containing protein 1A-like	0.00E+00	-3.0617	0.0187
comp18599_c0_seq3	ttn	titin-like	2.30E-113	-4.2610	0.0190
comp26336_c0_seq2	ccnt1	cyclin T1-like	6.57E-57	11.9888	0.0194
comp28356_c0_seq1	–	–	–	-11.2299	0.0201
comp74053_c0_seq1	–	–	–	6.7364	0.0201
comp35332_c1_seq2	–	–	–	-11.2664	0.0202
comp93725_c0_seq1	–	–	–	3.2033	0.0202
comp2892_c3_seq2	top2b	DNA topoisomerase 2-beta-like	0.00E+00	-13.0217	0.0204
comp3422_c0_seq6	add1	adducin 1 (alpha)	1.12E-69	-11.0299	0.0204
comp40696_c0_seq3	asmt	acetylserotonin O-methyltransferase-like	0.00E+00	3.1779	0.0204
comp55445_c0_seq2	–	–	–	4.6714	0.0205
comp15419_c1_seq14	shank3	SH3 and multiple ankyrin repeat domains 3	5.78E-38	-11.3543	0.0222
comp23728_c0_seq1	–	–	–	2.6465	0.0228
comp17669_c1_seq6	epc1	enhancer of polycomb homolog 1-like	0.00E+00	11.0158	0.0229
comp24681_c0_seq3	orc1	origin recognition complex, subunit 1-like	1.69E-125	11.0314	0.0229
comp30709_c0_seq2	zc3h7b	zinc finger CCCH domain-containing protein 7B	0.00E+00	4.1393	0.0229
comp733_c0_seq26	scn8a	sodium channel, voltage gated, type VIII, alpha-like	0.00E+00	-3.6338	0.0229
comp5648_c0_seq4	–	–	–	-11.2536	0.0232
comp8913_c0_seq3	grid2ip	delphilin-like, also known as glutamate receptor, ionotropic, delta 2-interacting protein 1	0.00E+00	-11.2179	0.0238
comp17731_c0_seq2	lonp2	lon protease homolog 2, peroxisomal-like	0.00E+00	-11.5501	0.0243
comp270_c0_seq3	atp1b4	ATPase, Na+/K+ transporting, beta 4-like	0.00E+00	-6.3608	0.0243
comp12893_c0_seq1	spag9	sperm associated antigen 9, also known as c-Jun-amino-terminal kinase-interacting protein 4-like	0.00E+00	-11.2788	0.0244
comp63190_c0_seq1	pck1	phosphoenolpyruvate carboxykinase 1, cytosolic	0.00E+00	2.8739	0.0245
comp8897_c0_seq1	znf384	zinc finger protein 384-like	4.56E-112	-3.9163	0.0248
comp40969_c0_seq3	akap13	A kinase (PRKA) anchor protein 13-like	0.00E+00	-11.2903	0.0250
comp24613_c0_seq11	fam58a	cyclin-related protein FAM58A-like	7.58E-46	10.9478	0.0256
comp30496_c0_seq1	–	–	–	-11.1435	0.0261
comp32878_c0_seq1	gbe1	Glucan (1,4-alpha-), branching enzyme 1	0.00E+00	2.3714	0.0263
comp13385_c1_seq10	tjp1	tight junction protein ZO-1-like	9.43E-10	-4.3710	0.0268
comp10674_c0_seq4	–	–	–	-11.0742	0.0272
comp112_c4_seq1	trappc8	trafficking protein particle complex subunit 8-like	0.00E+00	-3.1780	0.0272
comp14303_c0_seq2	Ift57	intraflagellar transport protein 57 homolog	0.00E+00	12.8456	0.0272
comp33344_c0_seq2	–	–	–	5.3507	0.0272
comp36345_c0_seq2	–	–	–	5.1394	0.0272
comp30696_c0_seq3	slc8b1	sodium/potassium/calcium exchanger 6, mitochondrial-like	0.00E+00	12.1689	0.0272
comp44227_c0_seq3	ssh2	slingshot protein phosphatase 2-like	0.00E+00	11.2044	0.0272
comp64448_c0_seq1	cass4	Cas scaffolding protein family member 4-like	0.00E+00	-11.2270	0.0274
comp3850_c1_seq2	ktn1	kinectin-like	0.00E+00	-1.8423	0.0278
comp82826_c0_seq2	tdrd15	tudor domain-containing protein 15-like	0.00E+00	11.1030	0.0278
comp27890_c0_seq10	plekha7	pleckstrin homology domain containing, family A member 7-like	9.01E-105	-11.2832	0.0280
comp17443_c0_seq1	abi2	abl-interactor 2-like	5.69E-91	-10.9824	0.0283
comp291_c2_seq7	cndp2	cytosolic non-specific dipeptidase-like	0.00E+00	-11.1347	0.0284
comp3140_c0_seq1	–	–	–	-11.0701	0.0284
comp11918_c0_seq1	–	–	–	3.0871	0.0286

comp23073_c0_seq1	macf1	microtubule-actin cross-linking factor 1-like	0.00E+00	-11.8416	0.0286
comp18089_c0_seq6	smurf1	SMAD specific E3 ubiquitin protein ligase 1-like	0.00E+00	6.4767	0.0288
comp3905_c1_seq1	cabin1	calcineurin-binding protein 1-like	0.00E+00	-13.9880	0.0301
comp8404_c0_seq4	limch1	LIM and calponin homology domains-containing protein 1-like	1.71E-89	-4.1353	0.0314
comp29372_c0_seq2	vil1	villin 1-like	2.50E-55	3.5185	0.0320
comp27517_c0_seq2	cep112	centrosomal protein 112	3.16E-117	-3.0860	0.0331
comp47214_c0_seq1	akr1b10	aldo-keto reductase family 1 member B10-like	4.88E-95	3.5051	0.0339
comp36283_c0_seq2	syde2	synapse defective 1, Rho GTPase, homolog 2-like	0.00E+00	-11.9458	0.0341
comp8598_c0_seq1	ercc6	excision repair cross-complementation group 6	0.00E+00	-3.4666	0.0341
comp9348_c1_seq3	c2cd2l	C2 domain-containing protein 2-like	1.24E-159	-11.3467	0.0341
comp2446_c0_seq5	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	6.13E-55	-13.9807	0.0354
comp50134_c0_seq3	usp53	inactive ubiquitin carboxyl-terminal hydrolase 53-like	0.00E+00	-11.2299	0.0354
comp25951_c1_seq1	ano8	anoctamin 8-like	0.00E+00	-11.1139	0.0354
comp51225_c0_seq1	gnrhr2	gonadotropin-releasing hormone II receptor	0.00E+00	-3.9570	0.0354
comp9983_c0_seq3	nt5c2	5'-nucleotidase, cytosolic II	0.00E+00	-2.1092	0.0354
comp9764_c0_seq3	samd14	sterile alpha motif domain-containing protein 14-like	2.61E-88	10.7194	0.0382
comp39558_c1_seq3	fam120b	constitutive coactivator of peroxisome proliferator-activated receptor gamma-like; also known as fam120b-like	1.89E-147	11.8574	0.0388
comp896_c3_seq41	capza2	f-actin-capping protein subunit alpha-2-like	6.76E-82	11.9561	0.0388
comp83787_c0_seq2	–	–	–	10.2320	0.0396
comp3567_c5_seq14	–	–	–	-5.2536	0.0397
comp10355_c0_seq11	phf2	lysine-specific demethylase phf2-like (PHD finger protein 2)	0.00E+00	-12.9285	0.0417
comp59739_c0_seq1	faah	fatty-acid amide hydrolase 1	8.39E-58	2.7457	0.0431
comp1555_c2_seq1	snx27	sorting nexin family member 27-like	0.00E+00	-10.9298	0.0432
comp15901_c0_seq6	lrch1	leucine-rich repeats and calponin homology (CH) domain containing protein 1-like	4.22E-122	-12.8603	0.0441
comp12780_c0_seq1	lancl3	LanC antibiotic synthetase component C-like 3-like	0.00E+00	-11.9156	0.0441
comp19075_c0_seq2	stox2	storkhead box protein 2-like	0.00E+00	-10.8778	0.0441
comp78394_c0_seq2	cnksr3	connector enhancer of kinase suppressor of ras 3-like	5.48E-126	10.7611	0.0443
comp43915_c0_seq1	pmp22	peripheral myelin protein 22	4.69E-63	3.2536	0.0444
comp60948_c0_seq2	map1b	microtubule-associated protein 1B	3.89E-29	2.6398	0.0444
comp108218_c0_seq1	–	–	–	-11.2113	0.0451
comp36018_c0_seq7	exd2	exonuclease 3'-5' domain-containing protein 2-like	0.00E+00	12.2308	0.0451
comp2075_c4_seq48	cacna1h	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	0.00E+00	-11.1763	0.0455
comp44117_c0_seq1	ck3	CC chemokine CK3	3.30E-38	3.3171	0.0455
comp53830_c0_seq1	ptp4a2	protein tyrosine phosphatase type IVA, member 2	1.81E-49	2.6919	0.0455
comp10086_c0_seq6	fbp	f-type lectin 3 (fucolectins)	6.94E-137	4.6328	0.0466
comp44046_c0_seq1	gnrhr	gonadotropin-releasing hormone receptor	4.02E-49	-3.8779	0.0466
comp7353_c0_seq2	gas8	growth arrest-specific protein 8-like	2.22E-40	-11.2196	0.0470
comp1507_c0_seq1	rap1gds1	RAP1, GTP-GDP dissociation stimulator 1-like	0.00E+00	-1.8282	0.0482
comp17883_c0_seq4	ubr3	ubiquitin protein ligase E3 component n-recognin 3-like	0.00E+00	-12.1595	0.0482
comp7009_c3_seq39	mrc1	macrophage mannose receptor 1-like	0.00E+00	3.3957	0.0482
comp94250_c0_seq1	–	–	–	-11.1228	0.0487
comp14425_c1_seq9	parp14	poly (ADP-ribose) polymerase 14-like	3.04E-44	10.6112	0.0497
comp24109_c0_seq1	calm	calmodulin-like	2.35E-07	3.0948	0.0497
comp8983_c0_seq2	vps11	vacuolar protein sorting-associated 11 homolog	0.00E+00	-11.0305	0.0497

Table S8 – List of differentially expressed contigs between females and transitional males of *Salaria pavo*. Positive log2FoldChange indicates an up-regulation of the transcript for transitional males, whereas a negative log2FoldChange indicates an up-regulation of the transcript for females.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp3407_c0_seq1	–	–	–	-13.6088	3.04E-11
comp4000_c4_seq5	atp1a3b	ATPase, Na+/K+ transporting, alpha 3 polypeptide-like	0.00E+00	13.8430	6.86E-11
comp33_c0_seq5	–	–	–	-5.4055	1.38E-10
comp33_c0_seq4	–	–	–	-4.5876	1.40E-09
comp7151_c1_seq2	p4htm	transmembrane prolyl 4-hydroxylase-like	8.46E-98	-13.2999	1.40E-09
comp9075_c0_seq8	tox2	TOX high mobility group box family member 2-like	2.64E-109	-13.0917	1.40E-09
comp14527_c0_seq2	rbbp6	e3 ubiquitin-protein ligase RBBP6-like	3.44E-178	-7.1951	2.10E-09
comp16173_c0_seq2	cstf1	cleavage stimulation factor subunit 1-like	0.00E+00	-13.0365	2.10E-09
comp444_c0_seq25	atp2b2	ATPase, Ca++ transporting, plasma membrane 2	1.07E-49	-13.3176	2.14E-09
comp33_c0_seq6	–	–	–	-5.2844	5.86E-09
comp25998_c0_seq2	dlgap2	discs, large (Drosophila) homolog-associated protein 2-like	1.42E-25	12.9219	2.05E-08
comp27209_c0_seq2	–	–	–	12.7105	2.05E-08
comp13731_c2_seq198	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	12.8209	2.55E-08
comp15571_c0_seq2	tsn8	tetraspanin 8-like	3.56E-154	-12.6303	6.17E-08
comp9075_c0_seq2	tox4	TOX high mobility group box family member 4-like	0.00E+00	7.6905	6.89E-08
comp12926_c0_seq7	–	–	–	-6.7734	7.50E-08
comp42_c0_seq1	–	–	–	-4.5752	8.50E-08
comp21976_c1_seq1	–	–	–	-12.6334	1.09E-07
comp7496_c4_seq3	lrrfip2	leucine rich repeat (in FLII) interacting protein 2	4.22E-167	-13.0781	1.09E-07
comp13731_c2_seq203	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	-5.0340	1.18E-07
comp5080_c1_seq9	spef2	sperm flagellar protein 2-like	1.60E-179	-12.6888	2.15E-07
comp7069_c0_seq4	st7	suppression of tumorigenicity 7 protein homolog	0.00E+00	13.7219	4.88E-07
comp23805_c0_seq5	celf1	cugbp, Elav-like family member 1-like	0.00E+00	12.0158	5.25E-07

comp39182_c0_seq36	erbb4	receptor tyrosine-protein kinase erbB-4-like	0.00E+00	12.1092	5.33E-07
comp3258_c0_seq8	–	–	–	-15.1553	5.48E-07
comp12811_c0_seq1	cybb	cytochrome b-245, beta polypeptide	0.00E+00	-13.2248	9.57E-07
comp16937_c0_seq2	pak4	p21 protein (Cdc42/Rac)-activated kinase 4, also known as serine/threonine-protein kinase PAK 7-like	1.12E-119	-12.3126	1.03E-06
comp2130_c1_seq16	numa1	nuclear mitotic apparatus protein 1-like	0.00E+00	-4.3918	1.08E-06
comp2446_c0_seq36	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	-7.3646	2.16E-06
comp26059_c1_seq118	csmd3	CUB and Sushi multiple domains 3-like	0.00E+00	-12.7662	3.51E-06
comp16358_c1_seq2	esyt2b	extended synaptotagmin-like protein 2b-like	0.00E+00	-11.8313	5.53E-06
comp2387_c2_seq29	chd9	chromodomain helicase DNA binding protein 9-like	0.00E+00	13.2415	5.53E-06
comp26306_c1_seq2	–	–	–	-12.0851	6.85E-06
comp44962_c0_seq1	rtel1	regulator of telomere elongation helicase 1-like	0.00E+00	-12.0168	6.85E-06
comp19383_c0_seq14	asa2	Ras GTPase-activating protein 2-like	0.00E+00	11.7082	1.03E-05
comp28508_c1_seq1	–	–	–	12.1233	1.06E-05
comp3480_c0_seq3	–	–	–	13.0331	1.06E-05
comp3063_c0_seq2	–	–	–	-12.3898	1.13E-05
comp6311_c0_seq37	adam23	ADAM metallopeptidase domain 23	0.00E+00	-4.5770	1.27E-05
comp45523_c0_seq1	pms1	PMS1 protein homolog 1-like	0.00E+00	11.7887	1.32E-05
comp9075_c0_seq4	tox4	TOX high mobility group box family member 4-like	0.00E+00	-4.4365	1.43E-05
comp21801_c0_seq3	–	–	–	-6.5477	1.65E-05
comp11430_c0_seq1	acox3	acyl-CoA oxidase 3, pristanoyl	0.00E+00	-11.9844	1.71E-05
comp6360_c0_seq31	frmpd1	FERM and PDZ domain-containing protein 1-like	0.00E+00	-3.7878	1.71E-05
comp23548_c0_seq4	atxn10	ataxin 10-like	0.00E+00	-12.7643	1.72E-05
comp7003_c0_seq3	fam20c	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	11.5249	1.72E-05
comp70843_c0_seq2	–	–	–	-11.8227	2.54E-05
comp11_c35_seq60	–	–	–	-12.2280	2.59E-05
comp264_c1_seq10	cacnb4	calcium channel, voltage-dependent, L type, beta 4 subunit-like	0.00E+00	-6.6149	2.59E-05
comp32242_c0_seq2	znf592	zinc finger protein 592-like	0.00E+00	-11.8849	2.59E-05
comp22544_c0_seq1	kiaa0907	UPF0469 protein KIAA0907 homolog	5.56E-130	-4.3054	3.32E-05
comp2645_c4_seq148	cacna1d	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-13.0498	4.09E-05
comp3258_c0_seq29	–	–	–	-4.4332	4.12E-05
comp2149_c0_seq1	cyp19a2	brain aromatase cytochrome P450	0.00E+00	-3.0938	4.40E-05
comp16003_c0_seq1	slc39a6	solute carrier family 39 (zinc transporter), member 6	0.00E+00	-6.3288	7.58E-05
comp33_c0_seq3	–	–	–	-5.3372	8.16E-05
comp11328_c0_seq1	arid1a	AT-rich interactive domain-containing protein 1A-like	0.00E+00	-13.7057	8.50E-05
comp4437_c1_seq9	cacna2d2	calcium channel, voltage-dependent, alpha 2/delta subunit 2	0.00E+00	-12.9820	8.60E-05
comp700_c4_seq5	ptprd	protein tyrosine phosphatase, receptor type, D-like	0.00E+00	14.9197	9.70E-05
comp37529_c0_seq1	–	–	–	-11.7333	1.10E-04
comp44541_c0_seq1	–	–	–	-12.6486	1.10E-04
comp32506_c0_seq1	asb6	ankyrin repeat and SOCS box protein 6-like	0.00E+00	-11.6157	1.20E-04
comp24806_c0_seq3	–	–	–	11.4615	1.27E-04
comp28515_c0_seq2	–	–	–	-11.6695	1.29E-04
comp16233_c0_seq2	tdrd3	tudor domain-containing protein 3-like	0.00E+00	11.1362	1.35E-04
comp44046_c0_seq1	gnrhr	gonadotropin-releasing hormone receptor	4.02E-49	-5.5789	1.35E-04
comp22092_c0_seq2	–	–	–	5.6384	1.37E-04
comp24604_c0_seq3	neto1	neuropilin (NRP) and tolloid (TLL)-like 1	0.00E+00	-12.6535	1.38E-04
comp223_c0_seq2	GTH-alpha	gonadotropin alpha subunit, also known as glycoprotein hormones, alpha polypeptide (cga)	4.74E-48	-3.7540	1.41E-04
comp3556_c0_seq1	–	–	–	4.5254	1.62E-04
comp7342_c0_seq23	tbl1xr1	F-box-like/WD repeat-containing protein TBL1XR1-like	0.00E+00	-11.7196	1.62E-04
comp5388_c0_seq4	rap1gap	RAP1 GTPase activating protein-like	2.68E-67	-3.4602	1.73E-04
comp18575_c0_seq2	rgs3	regulator of G-protein signaling 3-like	3.28E-49	-11.5661	1.83E-04
comp7655_c0_seq1	pianp	PILR alpha-associated neural protein-like	2.94E-88	5.6309	1.89E-04
comp52574_c0_seq11	znf236	zinc finger protein 236-like	0.00E+00	11.3470	2.34E-04
comp6374_c0_seq2	zfyve1	zinc finger FYVE domain-containing protein 1-like	0.00E+00	-4.0008	2.45E-04
comp23227_c0_seq2	lig3	DNA ligase 3-like	0.00E+00	-11.4988	2.52E-04
comp3744_c0_seq26	–	–	–	11.4670	2.71E-04
comp44426_c0_seq3	slc26a10	solute carrier family 26, member 10-like	0.00E+00	-12.7347	2.71E-04
comp32670_c0_seq9	cacna1b	calcium channel, voltage-dependent, N type, alpha 1B subunit-like	0.00E+00	11.6167	2.81E-04
comp34557_c0_seq4	aifm3	apoptosis-inducing factor, mitochondrion-associated, 3	0.00E+00	-11.6662	2.84E-04
comp452_c0_seq1	–	–	–	-2.9905	3.01E-04
comp5250_c0_seq2	usp46	ubiquitin carboxyl-terminal hydrolase 46-like	0.00E+00	-3.6709	3.21E-04
comp26424_c0_seq1	gtf3c5	general transcription factor IIIC, polypeptide 5-like	0.00E+00	-11.6092	3.86E-04
comp14233_c1_seq3	crat	arnitine O-acetyltransferase-like	0.00E+00	-11.4509	4.12E-04
comp7069_c0_seq2	rnf6	ring finger protein (C3H2C3 type) 6	2.69E-10	13.2444	4.34E-04
comp4939_c0_seq1	grl1	gamma-aminobutyric acid receptor-associated protein-like 1-like	1.27E-23	-11.4943	4.65E-04
comp56560_c0_seq2	–	–	–	12.4729	4.99E-04
comp30660_c0_seq3	wdr20	WD repeat-containing protein 20-like	0.00E+00	10.9618	6.14E-04
comp8826_c0_seq3	b3gat1	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1-like	1.55E-28	-11.3226	6.26E-04
comp6512_c0_seq2	capn1	calpain-1 catalytic subunit-like	1.86E-113	-12.7363	6.72E-04
comp1107_c9_seq1	ube2j1	ubiquitin-conjugating enzyme E2, J1	2.44E-133	-12.1967	7.03E-04
comp82826_c0_seq2	tdrd15	tudor domain-containing protein 15-like	0.00E+00	11.6966	7.62E-04
comp53985_c0_seq1	muc2	mucin 2-like	5.71E-46	-5.6000	7.84E-04
comp11052_c0_seq2	rpz3	rapunzel 3	0.00E+00	-3.8027	8.33E-04
comp700_c4_seq15	ptprd	protein tyrosine phosphatase, receptor type, D-like	1.90E-72	-11.2583	8.33E-04
comp3932_c0_seq4	ttbk1	tau tubulin kinase 1-like	0.00E+00	3.8356	8.48E-04
comp3329_c0_seq7	–	–	–	-3.4555	9.33E-04
comp4273_c0_seq2	cdkl5	cyclin-dependent kinase-like 5-like	3.46E-132	-5.8397	9.33E-04
comp36143_c1_seq1	dgcr6	DiGeorge syndrome critical region gene 6-like	5.36E-84	11.0450	0.0010
comp20054_c0_seq2	tha1	threonine aldolase 1	0.00E+00	-11.1188	0.0011

comp4627_c0_seq9	map2	microtubule-associated protein 2-like	0.00E+00	-13.2249	0.0011
comp30679_c0_seq1	–	–	–	-11.7254	0.0012
comp79831_c0_seq1	muc5b	mucin 5B-like	6.87E-62	-11.6864	0.0012
comp1764_c2_seq3	rnf123	e3 ubiquitin-protein ligase RNF123-like	0.00E+00	-11.3372	0.0012
comp17880_c0_seq6	qrich1	glutamine-rich protein 1-like	0.00E+00	5.1830	0.0012
comp18952_c0_seq2	golga2	golgin subfamily A member 2-like	0.00E+00	-3.3823	0.0012
comp4820_c1_seq3	mapt	microtubule-associated protein tau-like	3.86E-10	11.0969	0.0012
comp13104_c1_seq1	nedd4l	e3 ubiquitin-protein ligase nedd4-like	0.00E+00	-3.8130	0.0012
comp120052_c0_seq2	–	–	–	11.1493	0.0013
comp7601_c0_seq3	kcnma1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.00E+00	4.4408	0.0013
comp17669_c1_seq6	epc1	enhancer of polycomb homolog 1-like	0.00E+00	12.4855	0.0014
comp18110_c0_seq3	pom121	nuclear envelope pore membrane protein POM121-like	0.00E+00	13.4892	0.0014
comp61048_c0_seq2	–	–	–	11.2241	0.0014
comp13731_c2_seq185	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	3.5964	0.0014
comp10036_c0_seq1	lipin2	phosphatidate phosphatase lpin2-like	0.00E+00	-3.7520	0.0015
comp42859_c0_seq2	–	–	–	-11.3959	0.0015
comp2149_c0_seq6	FAM13A	family with sequence similarity 13, member A	1.12E-77	-11.1513	0.0015
comp87564_c0_seq6	ulk4	serine/threonine-protein kinase ULK4-like	0.00E+00	-11.5043	0.0016
comp11104_c1_seq4	rims1	regulating synaptic membrane exocytosis 1-like	2.79E-172	-3.7614	0.0017
comp38328_c0_seq1	s100a1	protein S100-A1-like	2.58E-54	-6.2619	0.0017
comp7178_c5_seq1	prickle2	prickle homolog 2	0.00E+00	3.1999	0.0017
comp94178_c0_seq1	–	–	–	11.2665	0.0018
comp11918_c0_seq1	–	–	–	3.1964	0.0019
comp12653_c1_seq2	srgap2	SLIT-ROBO Rho GTPase-activating protein 2-like	0.00E+00	4.0599	0.0019
comp34961_c0_seq9	fhdc1	FH2 domain-containing protein 1-like	0.00E+00	10.7971	0.0020
comp61048_c0_seq1	–	–	–	-12.1144	0.0020
comp42490_c0_seq3	adamts14	A disintegrin and metalloproteinase (ADAM) with thrombospondin motifs 14-like	0.00E+00	-11.1347	0.0022
comp7143_c1_seq5	lrrc7	leucine-rich repeat-containing protein 7-like	0.00E+00	-5.1001	0.0023
comp3754_c1_seq3	ppfibp1	liprin beta 1-like, also known as PTPRF interacting protein, binding protein 1	0.00E+00	-12.5296	0.0023
comp15688_c1_seq7	cacna1i	calcium channel, voltage-dependent, T type, alpha 1I subunit	0.00E+00	-11.5375	0.0023
comp131139_c1_seq1	siglec10	sialic acid-binding Ig-like lectin 10-like	0.00E+00	-11.1317	0.0023
comp18495_c0_seq2	ilk	integrin-linked protein kinase-like	0.00E+00	-12.8606	0.0024
comp21424_c0_seq1	pde7b	cAMP-specific 3',5'-cyclic phosphodiesterase 7B-like	0.00E+00	-3.5660	0.0024
comp13708_c0_seq2	stard9	StAR-related lipid transfer (START) domain containing 9	2.49E-147	4.0613	0.0025
comp23805_c0_seq2	celf1	cugbp, Elav-like family member 1-like	0.00E+00	-12.2280	0.0025
comp3266_c0_seq3	fam184a	protein FAM184A-like	0.00E+00	3.3860	0.0025
comp765_c0_seq1	rnf10	ring finger protein 10-like	0.00E+00	2.5252	0.0026
comp22347_c0_seq1	ngb	neuroglobin-like	1.37E-88	-11.9321	0.0027
comp3753_c2_seq7	ptprf	receptor-type tyrosine-protein phosphatase F-like	1.60E-62	-12.6131	0.0028
comp1591_c3_seq2	aste1	asteroid homolog 1-like	0.00E+00	-13.2689	0.0028
comp6269_c0_seq2	rbfox1	RNA binding protein, fox-1 homolog 1	4.76E-16	-3.3722	0.0030
comp3506_c0_seq318	fryl	protein furry homolog-like	2.29E-114	5.7869	0.0030
comp24887_c0_seq2	lcorl	ligand-dependent nuclear receptor corepressor-like protein	6.68E-27	12.8369	0.0032
comp31969_c0_seq3	ttll3	tubulin tyrosine ligase-like family, member 3-like	0.00E+00	-3.8624	0.0033
comp99983_c0_seq3	–	–	–	11.1167	0.0033
comp11489_c0_seq8	erc2	ELKS/RAB6-interacting/CAST family member 2	0.00E+00	-3.4898	0.0035
comp452_c0_seq2	–	–	–	-2.9869	0.0038
comp2944_c0_seq2	–	–	–	13.8763	0.0040
comp38328_c0_seq2	s100a1	protein S100-A1-like	4.81E-50	-3.8937	0.0040
comp7483_c2_seq2	–	–	–	-11.4988	0.0040
comp108218_c0_seq1	–	–	–	-11.2113	0.0041
comp3506_c0_seq82	fryl	protein furry homolog-like	0.00E+00	-2.7005	0.0042
comp6622_c0_seq1	snx4	sorting nexin 4-like	0.00E+00	-10.9932	0.0044
comp5979_c0_seq2	sephs2	selenide, water dikinase 2-like	5.71E-178	-10.9800	0.0045
comp112_c4_seq75	–	–	–	-12.2891	0.0045
comp22742_c0_seq2	wdtc1	WD and tetratricopeptide repeats protein 1-like	0.00E+00	-12.0375	0.0045
comp1376_c0_seq34	ank2	ankyrin 2	0.00E+00	-2.6995	0.0045
comp2229_c1_seq4	–	–	–	-2.7687	0.0046
comp3277_c1_seq42	bahcc1	BAH and coiled-coil domain-containing protein 1-like	0.00E+00	-4.0192	0.0051
comp1792_c0_seq17	sept8a	septin 8a-like	0.00E+00	-3.3818	0.0051
comp2645_c4_seq139	cacna1d	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-6.5150	0.0053
comp44326_c0_seq2	–	–	–	-11.9446	0.0054
comp26621_c0_seq3	rbm41	RNA binding motif protein 41-like	3.53E-178	-6.2509	0.0057
comp7891_c0_seq2	grhpr	glyoxylate reductase/hydroxypyruvate reductase-like	0.00E+00	-11.0295	0.0060
comp18835_c0_seq4	c1r	complement component 1, r subcomponent	0.00E+00	-2.5613	0.0069
comp22879_c1_seq3	glyatl1	glycine N-acyltransferase-like	4.79E-76	10.6626	0.0071
comp87988_c0_seq1	–	–	–	10.8453	0.0071
comp1124_c0_seq4	tacc2	transforming, acidic coiled-coil containing protein 2	0.00E+00	-14.2071	0.0072
comp3506_c0_seq37	fryl	protein furry homolog-like	0.00E+00	-12.4531	0.0073
comp23255_c0_seq2	bfar	bifunctional apoptosis regulator	0.00E+00	-12.5324	0.0076
comp15966_c0_seq1	tshb	thyroid stimulating hormone, beta subunit	7.31E-39	-2.6369	0.0081
comp2446_c0_seq5	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	6.13E-55	-13.9807	0.0081
comp413_c0_seq7	ttyh1	tweety homolog 1-like	1.39E-127	-4.1663	0.0085
comp73909_c0_seq1	map3k19	mitogen-activated protein kinase kinase kinase 19	0.00E+00	-3.5535	0.0085
comp476_c0_seq23	nlrc3	NLR family, CARD domain containing 3-like	2.84E-15	-10.9344	0.0086
comp54264_c0_seq1	–	–	–	-10.9012	0.0099
comp1771_c0_seq3	atg16l1	autophagy-related protein 16-like 1-like	4.51E-35	4.3223	0.0103
comp18465_c0_seq4	cpsf3	cleavage and polyadenylation specificity factor subunit 3-like	0.00E+00	-10.6752	0.0103

comp27_c2_seq3	atp1a1	ATPase, Na+/K+ transporting, alpha 1-like	1.54E-114	-11.7711	0.0103
comp7754_c2_seq18	–	–	–	10.7009	0.0103
comp16355_c0_seq4	tenc1	tensin-like C1 domain-containing phosphatase-like	2.02E-19	-3.2209	0.0103
comp2986_c0_seq1	–	–	–	-2.4279	0.0103
comp4978_c2_seq2	f13a	coagulation factor XIII A chain-like	0.00E+00	-12.0892	0.0103
comp3402_c3_seq18	mef2d	myocyte-specific enhancer factor 2D homolog	4.93E-90	-10.8216	0.0119
comp22492_c0_seq8	dtnbp1a	dystrobrevin-binding protein 1 (dysbindin-1)	2.56E-31	-4.0946	0.0119
comp52653_c0_seq1	spata18	Spermatogenesis associated 18, also known as mitochondria-eating protein-like	0.00E+00	-11.7597	0.0123
comp13493_c0_seq11	ndufs1	NADH dehydrogenase (ubiquinone) Fe-S protein 1	0.00E+00	-12.3425	0.0124
comp21177_c0_seq3	hagh	hydroxyacylglutathione hydrolase-like	3.01E-80	-11.2752	0.0124
comp153_c4_seq16	gria2	glutamate receptor 2-like	3.09E-32	-3.3814	0.0129
comp32717_c0_seq2	vrk1	serine/threonine-protein kinase VRK1-like, also known as vaccinia related kinase 1	0.00E+00	10.8743	0.0129
comp51225_c0_seq1	gnrhr2	gonadotropin-releasing hormone II receptor	0.00E+00	-4.0421	0.0139
comp6360_c0_seq19	frmpd1	FERM and PDZ domain-containing protein 1-like	0.00E+00	-7.8926	0.0139
comp10523_c0_seq4	scn8a	sodium channel, voltage gated, type VIII, alpha-like	6.55E-116	-4.1800	0.0140
comp25000_c0_seq1	–	–	–	-4.1639	0.0140
comp8947_c1_seq1	tgm3	protein-glutamine gamma-glutamyltransferase E-like	0.00E+00	-11.9980	0.0142
comp4845_c0_seq2	–	–	–	-10.8411	0.0143
comp62390_c0_seq2	scinla	scinderin like a, also known as gelsolin-like	0.00E+00	10.6078	0.0144
comp16838_c0_seq6	map4k5	mitogen-activated protein kinase kinase kinase kinase 5-like	0.00E+00	-2.2974	0.0148
comp15454_c0_seq8	tln2	talin-2-like	9.48E-98	-11.1108	0.0149
comp8820_c0_seq5	rictor	RPTOR independent companion of MTOR, complex 2	0.00E+00	-12.5186	0.0150
comp80854_c1_seq1	shpk	sedoheptulokinase-like	0.00E+00	-10.7806	0.0157
comp14411_c0_seq1	–	–	–	2.7345	0.0165
comp7303_c2_seq7	myt1a	myelin transcription factor 1-like	0.00E+00	-3.5760	0.0171
comp5551_c0_seq10	ank3	ankyrin 3-like	6.43E-55	11.9718	0.0178
comp11526_c0_seq15	gab1	GRB2-associated-binding protein 1	3.28E-86	-3.0275	0.0179
comp16767_c1_seq2	–	–	–	-2.4287	0.0193
comp7250_c0_seq13	–	–	–	-2.9831	0.0193
comp11637_c0_seq4	unc13a	unc-13 homolog A-like	9.89E-136	-3.1741	0.0195
comp27673_c1_seq1	yrdc	yrdC domain-containing protein, mitochondrial-like	5.60E-71	-11.1860	0.0195
comp17707_c1_seq4	col12a1	collagen, type XII, alpha 1-like	6.43E-65	-11.2353	0.0196
comp24708_c0_seq13	dlgap4	discs, large (Drosophila) homolog-associated protein 4-like	6.24E-174	5.5064	0.0202
comp754_c3_seq2	–	–	–	10.6893	0.0202
comp83787_c0_seq2	–	–	–	10.4002	0.0202
comp22671_c0_seq6	mrto4	mRNA turnover protein 4 homolog	5.44E-124	-10.7552	0.0202
comp9306_c0_seq4	ube2q2	ubiquitin-conjugating enzyme E2Q family member 2-like	0.00E+00	-12.2346	0.0202
comp30696_c0_seq3	slc8b1	sodium/potassium/calcium exchanger 6, mitochondrial-like	0.00E+00	11.9544	0.0205
comp17998_c0_seq9	rimbp2	RIMS-binding protein 2-like	5.11E-64	-3.9181	0.0218
comp27716_c0_seq1	tiam2	T-cell lymphoma invasion and metastasis-inducing protein 2	0.00E+00	-2.7692	0.0219
comp264_c1_seq16	smim19	small integral membrane protein 19 homolog	3.77E-58	-3.4265	0.0222
comp12845_c3_seq4	lrrk2	leucine-rich repeat kinase 2	0.00E+00	2.6029	0.0225
comp16012_c0_seq1	ctnnd1	catenin (cadherin-associated protein), delta 1-like	0.00E+00	-11.8392	0.0225
comp22690_c0_seq9	hectd3	HECT domain containing 3	0.00E+00	-4.9091	0.0232
comp70851_c0_seq2	–	–	–	10.5727	0.0236
comp28188_c2_seq20	tcf12	transcription factor 12-like	5.70E-125	-10.9790	0.0238
comp68_c2_seq1	rhoa	transforming protein RhoA-like	1.93E-126	-11.8424	0.0245
comp23516_c0_seq14	atp11b	probable phospholipid-transporting ATPase IF-like	0.00E+00	-10.8094	0.0250
comp92463_c0_seq2	rsph10b	radial spoke head 10 homolog B-like	0.00E+00	10.3070	0.0250
comp105744_c0_seq2	–	–	–	5.9599	0.0255
comp2550_c2_seq2	mmp16	matrix metalloproteinase 16-like	0.00E+00	-14.1156	0.0269
comp2319_c1_seq4	–	–	–	10.0827	0.0273
comp20590_c1_seq15	dot1l	DOT1-like histone H3K79 methyltransferase	8.35E-138	2.9383	0.0279
comp40032_c0_seq2	exosc7	exosome complex component RRP42-like	0.00E+00	-10.5211	0.0282
comp8108_c0_seq2	–	–	–	-14.0131	0.0299
comp7916_c1_seq4	tra2a	transformer 2 alpha homolog	5.00E-07	-10.7488	0.0306
comp23703_c0_seq8	limch1	LIM and calponin homology domains-containing protein 1-like	5.11E-71	10.8348	0.0322
comp6389_c1_seq2	tle2	transducin-like enhancer protein 4-like	0.00E+00	9.9514	0.0322
comp1383_c1_seq15	–	–	–	-10.8362	0.0322
comp49502_c0_seq2	mvb12a	multivesicular body subunit 12A-like	1.38E-45	-10.5299	0.0341
comp1631_c0_seq13	macf1	microtubule-actin cross-linking factor 1	0.00E+00	-3.0051	0.0349
comp15665_c0_seq2	cops8	COP9 signalosome complex subunit 8-like	1.07E-100	-12.2016	0.0349
comp17420_c0_seq3	pcdh1gc5	protocadherin 1 gamma C5-like	3.92E-88	-5.1860	0.0349
comp19894_c0_seq2	–	–	–	-10.6392	0.0349
comp2822_c1_seq9	epb41	erythrocyte membrane protein band 4.1	7.40E-48	-10.5331	0.0349
comp983_c0_seq1	ddx17	DEAD (Asp-Glu-Ala-Asp) box helicase 17	0.00E+00	-13.5489	0.0349
comp35840_c0_seq1	raf1	Raf-1 proto-oncogene, serine/threonine kinase	0.00E+00	-2.7000	0.0351
comp100639_c0_seq1	sh3d19	SH3 domain containing 19	9.85E-40	-10.7642	0.0354
comp51745_c0_seq2	usp53	inactive ubiquitin carboxyl-terminal hydrolase 53-like	0.00E+00	-2.7313	0.0357
comp15929_c0_seq8	adcy3	adenylate cyclase 3-like	0.00E+00	-2.6629	0.0363
comp591_c0_seq1	lhpp	phospholysine phosphohistidine inorganic pyrophosphate phosphatase-like	4.39E-139	-11.2361	0.0366
comp41090_c0_seq1	plcd4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like	0.00E+00	-10.6868	0.0368
comp19405_c0_seq18	pax6	paired box protein Pax-6	0.00E+00	-3.6637	0.0388
comp20660_c0_seq37	mdga2	MAM domain-containing glycosylphosphatidylinositol anchor protein 2	2.62E-57	-10.5537	0.0388
comp1635_c1_seq1	GtH-1 beta	gonadotropin 1 beta-subunit	5.96E-39	-3.1940	0.0393
comp4437_c1_seq3	cacna2d2	calcium channel, voltage-dependent, alpha 2/delta subunit 2	0.00E+00	-3.6697	0.0394
comp4707_c1_seq4	nrxn3b	neurexin 3b-like	4.03E-71	11.8453	0.0394

comp223_c0_seq1	GTH-alpha	gonadotropin alpha subunit, also known as glycoprotein hormones, alpha polypeptide (cga)	2.08E-41	-4.8219	0.0400
comp16838_c0_seq3	map4k5	mitogen-activated protein kinase kinase kinase kinase 5-like	0.00E+00	-13.1387	0.0403
comp36629_c0_seq1	–	–	–	-10.7934	0.0433
comp8977_c0_seq2	isoc2	isochorismatase domain-containing protein 2, mitochondrial-like	1.20E-122	2.7056	0.0437
comp23757_c0_seq6	apaf1	apoptotic protease activating factor 1-like	0.00E+00	-3.9482	0.0461
comp4620_c1_seq14	tcf7l2	transcription factor 7-like 2-like	0.00E+00	11.7392	0.0468
comp3506_c0_seq261	fryl	protein furry homolog-like	0.00E+00	-14.1717	0.0474
comp47774_c0_seq1	–	–	–	-11.1018	0.0475
comp12457_c0_seq7	cdk9	cyclin-dependent kinase 9	0.00E+00	-10.5825	0.0475
comp41016_c1_seq3	mmum	homocysteine S-methyltransferase 1-like	7.25E-104	-10.5798	0.0492
comp110833_c0_seq2	–	–	–	-10.5523	0.0499

Table S9 – List of differentially expressed contigs between sneaker males and transitional males of *Salaria pavo*. Positive log2FoldChange indicates an up-regulation of the transcript for transitional males, whereas a negative log2FoldChange indicates an up-regulation of the transcript for sneaker males.

Contig identifier	Gene Symbol	Gene name	E-value	log2FC	padj
comp765_c0_seq1	rnf10	ring finger protein 10-like	0.00E+00	15.5704	3.53E-17
comp7496_c4_seq1	lrrfip2	leucine rich repeat (in FLII) interacting protein 2	4.82E-159	14.7057	4.90E-14
comp16358_c1_seq2	esyt2b	extended synaptotagmin-like protein 2b-like	0.00E+00	-13.7725	1.76E-12
comp12926_c0_seq7	–	–	–	-7.7621	5.86E-11
comp2816_c0_seq2	lphn3	latrophilin 3-like	0.00E+00	13.9669	5.86E-11
comp1424_c1_seq3	arhgap44	Rho GTPase activating protein 44-like	9.68E-42	13.6824	8.10E-10
comp2446_c0_seq48	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	2.45E-150	13.7169	8.62E-10
comp1127_c4_seq2	iqsec2	IQ motif and SEC7 domain-containing protein 2-like	0.00E+00	7.3316	3.54E-09
comp26241_c0_seq3	nhs	Nance-Horan syndrome protein-like	0.00E+00	13.5171	3.77E-09
comp7496_c4_seq3	lrrfip2	leucine rich repeat (in FLII) interacting protein 2	4.22E-167	-13.5877	7.96E-09
comp11119_c0_seq12	cdkl5	cyclin-dependent kinase-like 5-like	0.00E+00	13.4136	9.82E-09
comp10728_c1_seq2	–	–	–	13.3983	1.26E-08
comp7003_c0_seq5	fam20c	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	13.2517	2.18E-08
comp9075_c0_seq8	tox2	TOX high mobility group box family member 2-like	2.64E-109	-12.4828	5.70E-08
comp16173_c0_seq2	cstf1	cleavage stimulation factor subunit 1-like	0.00E+00	-12.4350	1.26E-07
comp22092_c0_seq1	–	–	–	13.9429	1.47E-07
comp20054_c0_seq2	tha1	threonine aldolase 1	0.00E+00	-12.7013	2.18E-07
comp18465_c0_seq4	cpsf3	cleavage and polyadenylation specificity factor subunit 3-like	0.00E+00	-12.6730	4.66E-07
comp15571_c0_seq2	tsn8	tetraspanin 8-like	3.56E-154	-12.5140	9.04E-07
comp3422_c0_seq6	add1	adducin 1 (alpha)	1.12E-69	12.8481	9.64E-07
comp14527_c0_seq2	rbbp6	e3 ubiquitin-protein ligase RBBP6-like	3.44E-178	-6.5868	1.13E-06
comp19398_c0_seq1	setd2	histone-lysine N-methyltransferase SETD2-like	0.00E+00	12.9526	1.13E-06
comp3258_c0_seq8	–	–	–	-14.8936	3.18E-06
comp14233_c1_seq3	crat	arnitine O-acetyltransferase-like	0.00E+00	-12.4439	3.88E-06
comp26691_c0_seq2	–	–	–	12.8017	3.91E-06
comp7124_c0_seq7	–	–	–	6.5475	3.92E-06
comp32506_c0_seq1	asb6	ankyrin repeat and SOCS box protein 6-like	0.00E+00	-12.2308	4.10E-06
comp3813_c0_seq4	znf207	zinc finger protein 207-like	4.88E-71	13.1503	4.75E-06
comp7689_c1_seq2	crebzf	CREB/ATF bZIP transcription factor-like	2.75E-56	12.4157	4.98E-06
comp6512_c0_seq2	capn1	calpain-1 catalytic subunit-like	1.86E-113	-14.5587	5.64E-06
comp9075_c0_seq2	tox4	TOX high mobility group box family member 4-like	0.00E+00	12.9510	5.64E-06
comp9075_c0_seq4	tox4	TOX high mobility group box family member 4-like	0.00E+00	-4.6334	5.64E-06
comp55390_c0_seq1	–	–	–	13.6970	6.79E-06
comp444_c0_seq25	atp2b2	ATPase, Ca++ transporting, plasma membrane 2	1.07E-49	-11.7816	6.91E-06
comp16838_c0_seq6	map4k5	mitogen-activated protein kinase kinase kinase kinase 5-like	0.00E+00	-3.4494	7.67E-06
comp17250_c1_seq2	rock2	rho-associated protein kinase 2-like	0.00E+00	-4.4834	9.47E-06
comp25998_c0_seq2	dlgap2	discs, large (Drosophila) homolog-associated protein 2-like	1.42E-25	12.9219	9.98E-06
comp21801_c0_seq3	–	–	–	-6.5441	1.15E-05
comp27209_c0_seq2	–	–	–	12.7105	1.18E-05
comp7003_c0_seq3	fam20c	extracellular serine/threonine protein kinase FAM20C-like	0.00E+00	-3.4068	1.18E-05
comp17443_c0_seq1	abi2	abl-interactor 2-like	5.69E-91	12.4812	1.37E-05
comp12811_c0_seq1	cybb	cytochrome b-245, beta polypeptide	0.00E+00	-12.8320	1.40E-05
comp12457_c0_seq7	cdk9	cyclin-dependent kinase 9	0.00E+00	-13.0523	1.51E-05
comp41090_c0_seq1	plcd4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like	0.00E+00	-13.0536	1.56E-05
comp10558_c0_seq46	kcnt1	potassium channel, subfamily T, member 1	0.00E+00	13.0259	2.05E-05
comp11_c35_seq60	–	–	–	-12.3947	2.53E-05
comp17250_c1_seq26	dgcr2	integral membrane protein DGCR2/IDD-like, also known as DiGeorge syndrome critical region gene 2	0.00E+00	-12.6286	3.34E-05
comp4829_c1_seq3	wdr54	WD repeat domain 54	0.00E+00	12.4796	3.34E-05
comp70843_c0_seq1	–	–	–	12.7659	3.34E-05
comp8826_c0_seq3	b3gat1	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1-like	1.55E-28	-11.9098	3.34E-05
comp17106_c0_seq3	cntn3	contactin 3	0.00E+00	12.6228	3.65E-05
comp19489_c0_seq1	tnrc6b	trinucleotide repeat containing 6b-like	3.36E-125	4.3245	3.99E-05
comp18835_c0_seq4	c1r	complement component 1, r subcomponent	0.00E+00	-3.3149	4.03E-05
comp19075_c0_seq2	stox2	storkhead box protein 2-like	0.00E+00	12.4547	4.15E-05
comp40032_c0_seq2	exosc7	exosome complex component RRP42-like	0.00E+00	-11.9740	4.15E-05
comp451_c4_seq1	–	–	–	12.1527	4.40E-05

comp2446_c0_seq38	cacna1a	calcium channel, voltage-dependent, P/Q type, alpha 1A-like	0.00E+00	12.3413	4.95E-05
comp28515_c0_seq2	–	–	–	-11.9712	4.96E-05
comp6345_c0_seq4	angl4	angiopoietin-related protein 4-like	1.02E-177	-3.1791	7.04E-05
comp4650_c1_seq3	gpc6	glypican 6-like	0.00E+00	12.2153	7.46E-05
comp3258_c0_seq29	–	–	–	-4.5988	7.57E-05
comp3407_c0_seq1	–	–	–	-11.5756	7.57E-05
comp34114_c0_seq1	chl1	neural cell adhesion molecule L1-like	0.00E+00	-3.4401	7.57E-05
comp7143_c1_seq5	lrrc7	leucine-rich repeat-containing protein 7-like	0.00E+00	-5.9150	7.74E-05
comp16937_c0_seq2	pak4	p21 protein (Cdc42/Rac)-activated kinase 4, also known as serine/threonine-protein kinase PAK 7-like	1.12E-119	-11.8305	8.06E-05
comp70843_c0_seq2	–	–	–	-11.5847	8.11E-05
comp21022_c0_seq8	bptf	nucleosome-remodeling factor subunit BPTF-like	0.00E+00	12.0205	8.12E-05
comp23548_c0_seq4	atxn10	ataxin 10-like	0.00E+00	-12.4875	8.40E-05
comp18575_c0_seq2	rgs3	regulator of G-protein signaling 3-like	3.28E-49	-11.6646	1.07E-04
comp24280_c1_seq7	zmym4	zinc finger, MYM-type 4-like	0.00E+00	5.6159	1.07E-04
comp4900_c1_seq16	–	–	–	12.0744	1.11E-04
comp1351_c0_seq2	myef2	myelin expression factor 2-like	4.82E-154	-4.1802	1.15E-04
comp26621_c0_seq3	rbm41	RNA binding motif protein 41-like	3.53E-178	-7.1853	1.15E-04
comp2130_c1_seq16	numa1	nuclear mitotic apparatus protein 1-like	0.00E+00	-3.5035	1.17E-04
comp2057_c0_seq12	lrit2	leucine-rich repeat and transmembrane domain-containing protein 2-like	2.03E-135	-11.9069	1.32E-04
comp44426_c0_seq3	slc26a10	solute carrier family 26, member 10-like	0.00E+00	-13.1545	1.61E-04
comp8897_c0_seq1	znf384	zinc finger protein 384-like	4.56E-112	4.9964	1.73E-04
comp1631_c0_seq13	macf1	microtubule-actin cross-linking factor 1	0.00E+00	-4.1174	1.86E-04
comp21976_c1_seq1	–	–	–	-11.6023	1.86E-04
comp11430_c0_seq1	acox3	acyl-CoA oxidase 3, pristanoyl	0.00E+00	-11.8275	2.07E-04
comp17870_c0_seq3	ccbl2	kynurenine-oxoglutarate transaminase 3-like, also known ascysteine conjugate-beta lyase 2	0.00E+00	-3.2710	2.24E-04
comp16106_c0_seq1	vps29	vacuolar protein sorting-associated protein 29-like	7.30E-123	3.7700	2.27E-04
comp3754_c1_seq3	ppfibp1	liprin beta 1-like, also known as PTPRF interacting protein, binding protein 1	0.00E+00	-13.4605	2.57E-04
comp44962_c0_seq1	rtel1	regulator of telomere elongation helicase 1-like	0.00E+00	-11.5465	2.57E-04
comp20540_c0_seq1	pi4k2a	phosphatidylinositol 4-kinase type 2 alpha-like	0.00E+00	11.8810	2.61E-04
comp24604_c0_seq3	neto1	neuropilin (NRP) and tolloid (TLL)-like 1	0.00E+00	-12.4478	2.61E-04
comp33263_c0_seq3	gdap2	ganglioside induced differentiation associated protein 2-like	0.00E+00	13.0302	2.61E-04
comp26059_c1_seq118	csmd3	CUB and Sushi multiple domains 3-like	0.00E+00	-11.6948	2.66E-04
comp42490_c0_seq3	adamts14	A disintegrin and metalloproteinase (ADAM) with thrombospondin motifs 14	0.00E+00	-11.6864	2.77E-04
comp3852_c0_seq10	arhgef9	Cdc42 guanine nucleotide exchange factor (GEF) 9	0.00E+00	3.5966	2.91E-04
comp8725_c1_seq2	gtpbp1	GTP binding protein 1-like	0.00E+00	11.9618	2.94E-04
comp2852_c1_seq4	ndufaf7	NADH dehydrogenase [ubiquinone] complex I, assembly factor 7	0.00E+00	4.9598	3.37E-04
comp87564_c0_seq6	ulk4	serine/threonine-protein kinase ULK4-like	0.00E+00	-11.7955	3.37E-04
comp18952_c0_seq2	golga2	golgin subfamily A member 2-like	0.00E+00	-3.5679	3.38E-04
comp3506_c0_seq134	fryl	protein furry homolog-like	0.00E+00	11.8840	3.38E-04
comp6371_c1_seq11	mical3	protein-methionine sulfoxide oxidase mical3a-like, also known as microtubule associated monooxygenase, calponin and LIM domain containing 3	0.00E+00	11.8688	3.38E-04
comp23516_c0_seq14	atp11b	probable phospholipid-transporting ATPase IF-like	0.00E+00	-11.9296	3.80E-04
comp11637_c0_seq6	unc13a	unc-13 homolog A-like	9.47E-150	11.9609	3.83E-04
comp23227_c0_seq2	lig3	DNA ligase 3-like	0.00E+00	-11.4019	4.07E-04
comp12377_c0_seq8	–	–	–	11.8585	4.38E-04
comp7879_c1_seq24	rere	arginine-glutamic acid dipeptide repeats protein-like	0.00E+00	12.0276	4.38E-04
comp18599_c0_seq3	ttn	titin-like	2.30E-113	5.0615	4.75E-04
comp264_c1_seq10	cacnb4	calcium channel, voltage-dependent, L type, beta 4 subunit-like	0.00E+00	-5.9113	4.77E-04
comp10558_c0_seq40	kcnt1	potassium channel, subfamily T, member 1	0.00E+00	11.9437	5.44E-04
comp35039_c0_seq2	dapk2	death-associated protein kinase 2-like	5.18E-145	11.7996	5.78E-04
comp19405_c0_seq18	pax6	paired box protein Pax-6	0.00E+00	-4.7093	5.97E-04
comp643_c1_seq4	–	–	–	11.7683	5.97E-04
comp61831_c0_seq2	gpr56	G protein-coupled receptor 56-like	3.22E-119	-11.8604	6.67E-04
comp18742_c0_seq2	tmem154	transmembrane protein 154-like	4.18E-29	-3.7875	7.07E-04
comp7807_c0_seq7	akap9	A kinase (PRKA) anchor protein 9-like	0.00E+00	13.6580	7.26E-04
comp16003_c0_seq1	slc39a6	solute carrier family 39 (zinc transporter), member 6	0.00E+00	-6.1150	7.38E-04
comp12640_c0_seq4	fkbp5	peptidyl-prolyl cis-trans isomerase FKBP5-like	3.92E-104	-3.3734	7.53E-04
comp7303_c2_seq7	myt1a	myelin transcription factor 1-like	0.00E+00	-4.2371	7.87E-04
comp9222_c1_seq1	fez2	fasciculation and elongation protein zeta-2-like (Zygin II)	4.02E-150	12.3641	8.09E-04
comp19865_c0_seq3	chat	choline O-acetyltransferase-like	0.00E+00	11.9038	9.01E-04
comp35117_c1_seq17	44M2.3	putative RNA exonuclease NEF-sp-like	7.46E-138	-11.4340	9.01E-04
comp17707_c1_seq4	col12a1	collagen, type XII, alpha 1-like	6.43E-65	-12.1230	9.47E-04
comp7601_c0_seq3	kcnma1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.00E+00	11.9161	9.47E-04
comp6412_c0_seq3	galnt18	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 18-like	0.00E+00	12.9277	9.48E-04
comp5250_c0_seq2	usp46	ubiquitin carboxyl-terminal hydrolase 46-like	0.00E+00	-3.2729	9.58E-04
comp9559_c1_seq11	casc3	protein CASC3-like	1.81E-112	-4.0549	9.58E-04
comp9729_c0_seq1	ctgf	connective tissue growth factor-like	0.00E+00	-3.0327	0.0011
comp23255_c0_seq2	bfar	bifunctional apoptosis regulator	0.00E+00	-13.1359	0.0012
comp45141_c0_seq2	–	–	–	11.6819	0.0012
comp700_c4_seq5	ptprd	protein tyrosine phosphatase, receptor type, D-like	0.00E+00	14.9197	0.0012
comp13993_c0_seq128	prune2	prune homolog 2	0.00E+00	11.5072	0.0013
comp55445_c0_seq2	–	–	–	-11.5169	0.0013
comp22295_c0_seq9	dos	protein dos-like	3.60E-141	12.4437	0.0014
comp3258_c0_seq2	–	–	–	4.7123	0.0014
comp38399_c0_seq2	gpr4	G protein-coupled receptor 4-like	0.00E+00	-3.9246	0.0015
comp13104_c1_seq1	nedd4l	e3 ubiquitin-protein ligase nedd4-like	0.00E+00	-3.6764	0.0016

comp1637_c4_seq15	map3k9	mitogen-activated protein kinase kinase kinase 9-like	2.20E-69	11.7340	0.0016
comp42859_c0_seq2	–	–	–	-11.4300	0.0017
comp15688_c1_seq7	cacna1i	calcium channel, voltage-dependent, T type, alpha 1I subunit	0.00E+00	-11.7355	0.0017
comp93799_c0_seq2	–	–	–	-11.2320	0.0017
comp13731_c2_seq203	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	-3.3970	0.0019
comp17540_c1_seq4	pdcd4	programmed cell death protein 4-like	0.00E+00	11.7386	0.0019
comp16355_c0_seq4	tenc1	tensin-like C1 domain-containing phosphatase-like	2.02E-19	-3.7497	0.0019
comp29929_c0_seq25	atp11a	probable phospholipid-transporting ATPase IH-like	0.00E+00	-4.4178	0.0020
comp7739_c0_seq28	tcf4	transcription factor 4-like	0.00E+00	12.3196	0.0022
comp76813_c0_seq1	–	–	–	11.7355	0.0022
comp10086_c0_seq5	fbp	f-type lectin 3 (fucolectins)	1.17E-137	13.1639	0.0022
comp72278_c0_seq1	–	–	–	12.1653	0.0022
comp971_c7_seq1	coa6	cytochrome c oxidase assembly factor 6 homolog	2.51E-34	-11.2658	0.0023
comp3277_c1_seq42	bahcc1	BAH and coiled-coil domain-containing protein 1-like	0.00E+00	-4.5658	0.0023
comp6360_c0_seq31	frmpd1	FERM and PDZ domain-containing protein 1-like	0.00E+00	-2.9111	0.0023
comp16521_c0_seq2	rps27	40S ribosomal protein S27-like	2.72E-51	-4.8781	0.0025
comp22325_c0_seq2	igfbp1	insulin-like growth factor binding protein 1	8.92E-124	-2.7400	0.0025
comp49502_c0_seq2	mvb12a	multivesicular body subunit 12A-like	1.38E-45	-11.3628	0.0025
comp1934_c0_seq2	–	–	–	11.9187	0.0025
comp64880_c0_seq1	–	–	–	11.7082	0.0026
comp13731_c2_seq140	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	3.1776	0.0026
comp13385_c1_seq10	tjp1	tight junction protein ZO-1-like	9.43E-10	4.7823	0.0028
comp17129_c0_seq1	–	–	–	11.4634	0.0028
comp67298_c0_seq3	arap2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2-like	0.00E+00	-12.5881	0.0028
comp29094_c0_seq1	morc3	MORC family CW-type zinc finger protein 3	0.00E+00	-3.7681	0.0029
comp3506_c0_seq37	fryl	protein furry homolog-like	0.00E+00	-12.3546	0.0029
comp700_c4_seq15	ptprd	protein tyrosine phosphatase, receptor type, D-like	1.90E-72	-10.9112	0.0029
comp11104_c1_seq2	rims2	regulating synaptic membrane exocytosis 2-like	1.75E-128	12.2622	0.0029
comp31310_c0_seq8	plekha4	pleckstrin homology domain-containing, family A member 4-like	0.00E+00	-4.0613	0.0032
comp32834_c0_seq1	gatm	glycine amidinotransferase, mitochondrial-like	0.00E+00	-2.7374	0.0032
comp6820_c0_seq1	–	–	–	11.8757	0.0033
comp20780_c0_seq7	plekho1a	pleckstrin homology domain containing, family O member 1a-like	6.96E-133	11.5502	0.0034
comp23065_c0_seq4	tbc1d22b	TBC1 domain family, member 22B-like	0.00E+00	-11.8351	0.0034
comp3063_c0_seq2	–	–	–	-11.5074	0.0034
comp9012_c0_seq4	lrrc7	leucine-rich repeat-containing protein 7	2.42E-24	-11.2967	0.0034
comp29531_c0_seq12	Ttll4	tubulin tyrosine ligase-like family, member 4-like	0.00E+00	11.4921	0.0035
comp67399_c0_seq4	myo7a	unconventional myosin-VIIa-like	7.87E-151	-11.3937	0.0035
comp16520_c0_seq1	chd6	chromodomain helicase DNA binding protein 6-like	1.11E-78	11.7654	0.0038
comp27716_c0_seq1	tiam2	T-cell lymphoma invasion and metastasis-inducing protein 2	0.00E+00	-3.0394	0.0038
comp6374_c0_seq2	zfyve1	zinc finger FYVE domain-containing protein 1-like	0.00E+00	-3.4347	0.0039
comp11223_c0_seq2	tdrkh	tudor and KH domain-containing protein-like	0.00E+00	-3.1541	0.0043
comp26840_c0_seq1	slc6a	sodium-dependent serotonin transporter-like, also known as solute carrier family 6 (neurotransmitter transporter)	0.00E+00	-2.9964	0.0043
comp5180_c0_seq3	–	–	–	2.9940	0.0043
comp28356_c0_seq1	–	–	–	11.5520	0.0044
comp2058_c0_seq3	igl1	L-I immunoglobulin light chain	5.47E-47	-2.7238	0.0048
comp35530_c0_seq1	–	–	–	11.6787	0.0048
comp22690_c0_seq9	hectd3	HECT domain containing 3	0.00E+00	-5.2998	0.0050
comp33344_c0_seq2	–	–	–	-11.6112	0.0052
comp2822_c1_seq9	epb41	erythrocyte membrane protein band 4.1	7.40E-48	-11.1855	0.0053
comp15373_c0_seq2	agrn	agrin-like	0.00E+00	12.1388	0.0057
comp6871_c0_seq1	–	–	–	-2.6297	0.0057
comp4939_c0_seq1	grl1	gamma-aminobutyric acid receptor-associated protein-like 1-like	1.27E-23	-11.1832	0.0058
comp2149_c0_seq6	FAM13A	family with sequence similarity 13, member A	1.12E-77	-11.0703	0.0058
comp69178_c0_seq1	–	–	–	11.4510	0.0058
comp5867_c1_seq4	–	–	–	11.4720	0.0062
comp10674_c0_seq4	–	–	–	11.3597	0.0067
comp63771_c0_seq2	cdkl1	cyclin-dependent kinase-like 1-like	1.48E-160	-11.3649	0.0067
comp3506_c0_seq319	fryl	protein furry homolog-like	1.21E-100	3.6022	0.0069
comp1555_c2_seq1	snx27	sorting nexin family member 27-like	0.00E+00	11.4092	0.0070
comp15419_c1_seq14	shank3	SH3 and multiple ankyrin repeat domains 3	5.78E-38	11.4988	0.0077
comp13731_c2_seq185	csmd1	CUB and sushi domain-containing protein 1	0.00E+00	4.0130	0.0078
comp1792_c0_seq17	sept8a	septin 8a-like	0.00E+00	-3.1394	0.0080
comp24806_c0_seq3	–	–	–	11.4615	0.0081
comp6080_c3_seq26	arhgef11	Rho guanine nucleotide exchange factor (GEF) 11-like	0.00E+00	11.7985	0.0081
comp1124_c0_seq4	tacc2	transforming, acidic coiled-coil containing protein 2	0.00E+00	-14.3449	0.0083
comp15028_c0_seq11	agtpbp1	ATP/GTP binding protein 1	1.30E-162	11.3873	0.0086
comp7601_c0_seq4	kcnma1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	0.00E+00	3.0984	0.0086
comp8000_c0_seq4	plcd3a	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-3-A-like	0.00E+00	-2.5280	0.0088
comp11802_c0_seq5	foxk1	forkhead box K1	0.00E+00	11.5681	0.0088
comp15419_c1_seq2	shank3	SH3 and multiple ankyrin repeat domains 3	0.00E+00	-12.9126	0.0093
comp17848_c0_seq2	btd	biotinidase-like	0.00E+00	11.8411	0.0093
comp55090_c0_seq1	–	–	–	11.1616	0.0093
comp28271_c0_seq8	smtn	smoothelin-like	0.00E+00	-2.8529	0.0094
comp7149_c0_seq3	wbp11	WW domain-binding protein 11-like	1.48E-70	11.1447	0.0094
comp537_c1_seq2	cask	peripheral plasma membrane protein CASK-like, also calcium/calmodulin-dependent serine protein kinase (MAGUK family) (cask)	0.00E+00	2.9605	0.0096
comp1764_c2_seq3	rnf123	e3 ubiquitin-protein ligase RNF123-like	0.00E+00	-10.9055	0.0104
comp66406_c0_seq2	stox1	storkhead box protein 1-like	0.00E+00	-3.7342	0.0112

comp2964_c0_seq1	eml1	echinoderm microtubule-associated protein-like 1-like	0.00E+00	-2.6315	0.0114
comp16713_c0_seq5	zcchc11	terminal uridylyltransferase 4-like, alson known as zinc finger, CCHC domain containing 11	4.96E-31	-10.9888	0.0115
comp21177_c0_seq3	hagh	hydroxyacylglutathione hydrolase-like	3.01E-80	-11.6492	0.0115
comp24708_c0_seq4	dlgap4	discs, large (Drosophila) homolog-associated protein 4-like	0.00E+00	-3.7053	0.0115
comp71210_c0_seq2	–	–	–	11.4515	0.0117
comp32670_c0_seq9	cacna1b	calcium channel, voltage-dependent, N type, alpha 1B subunit-like	0.00E+00	11.6167	0.0118
comp47190_c0_seq1	–	–	–	11.3617	0.0118
comp131139_c1_seq1	siglec10	sialic acid-binding Ig-like lectin 10-like	0.00E+00	-10.8139	0.0119
comp22583_c0_seq5	kdm6b	lysine (K)-specific demethylase 6B	8.61E-97	11.8310	0.0122
comp10917_c0_seq6	scrib	scribbled planar cell polarity protein homolog	0.00E+00	-3.9556	0.0122
comp11052_c0_seq2	rpz3	rapunzel 3	0.00E+00	-3.5114	0.0122
comp11104_c1_seq4	rims1	regulating synaptic membrane exocytosis 1-like	2.79E-172	-3.3085	0.0122
comp13104_c1_seq3	nedd4l	e3 ubiquitin-protein ligase nedd4-like	6.32E-119	13.0915	0.0122
comp19953_c0_seq1	–	–	–	-2.3828	0.0122
comp130808_c0_seq1	–	–	–	11.1744	0.0127
comp23848_c0_seq12	zcrb1	zinc finger CCHC-type and RNA-binding motif-containing protein 1	8.72E-75	-11.5131	0.0129
comp27_c2_seq3	atp1a1	ATPase, Na+/K+ transporting, alpha 1-like	1.54E-114	-11.6629	0.0135
comp1542_c1_seq12	pcdhga2	protocadherin gamma subfamily A, 2-like	0.00E+00	2.7202	0.0138
comp983_c0_seq1	ddx17	DEAD (Asp-Glu-Ala-Asp) box helicase 17	0.00E+00	-14.3588	0.0138
comp20054_c0_seq3	tha1	threonine aldolase 1	8.85E-161	11.1385	0.0147
comp23534_c1_seq9	grip2	glutamate receptor-interacting protein 2-like	0.00E+00	11.5502	0.0147
comp16332_c0_seq2	ophn1	oligophrenin 1-like	0.00E+00	11.1684	0.0147
comp3874_c1_seq6	pfkfb4	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4	0.00E+00	-3.2809	0.0148
comp7178_c4_seq34	rimbp2	RIMS binding protein 2-like	1.44E-149	-4.0411	0.0148
comp45743_c0_seq5	kcnmb2	calcium-activated potassium channel subunit beta-2-like	4.39E-94	-4.6273	0.0148
comp2645_c4_seq139	cacna1d	calcium channel, voltage-dependent, L type, alpha 1D subunit-like	0.00E+00	-5.5334	0.0148
comp1591_c3_seq2	aste1	asteroid homolog 1-like	0.00E+00	-12.2377	0.0150
comp54264_c0_seq1	–	–	–	-11.0677	0.0155
comp211_c0_seq5	ralgps1	ras-specific guanine nucleotide-releasing factor RalGPS1-like	2.80E-160	-3.3214	0.0157
comp68897_c0_seq2	–	–	–	12.3876	0.0157
comp1511_c1_seq9	ncam1	neural cell adhesion molecule 1-like	5.13E-15	4.7127	0.0162
comp33263_c0_seq1	gdap2	ganglioside induced differentiation associated protein 2-like	0.00E+00	-2.4092	0.0162
comp21722_c1_seq3	ddr1	discoidin domain-containing receptor 1-like	0.00E+00	-11.8897	0.0166
comp80854_c1_seq1	shpk	sedoheptulokinase-like	0.00E+00	-10.7194	0.0167
comp36629_c0_seq1	–	–	–	-10.9561	0.0176
comp20028_c0_seq1	cisd3	CDGSH iron-sulfur domain-containing protein 3, mitochondrial-like	4.53E-25	-4.0968	0.0185
comp24068_c0_seq3	–	–	–	-3.3975	0.0187
comp13731_c2_seq174	csmd3	CUB and sushi domain-containing protein 3	0.00E+00	3.7083	0.0192
comp15331_c0_seq8	gulp1	GULP, engulfment adaptor PTB domain containing protein 1-like	0.00E+00	-11.5244	0.0198
comp25951_c1_seq1	ano8	anoctamin 8-like	0.00E+00	11.2080	0.0198
comp43041_c0_seq1	–	–	–	-3.6265	0.0203
comp94250_c0_seq1	–	–	–	11.2897	0.0204
comp27920_c0_seq2	–	–	–	-11.2181	0.0204
comp5318_c2_seq29	ddx3x	DEAD (Asp-Glu-Ala-Asp) box helicase 3, also known as ATP-dependent RNA helicase DDX3X-like	0.00E+00	3.6523	0.0205
comp12516_c0_seq1	–	–	–	-2.4079	0.0208
comp14530_c0_seq1	acad8	acyl-CoA dehydrogenase family, member 8	0.00E+00	-3.7038	0.0213
comp12893_c0_seq1	spag9	sperm associated antigen 9, also known as c-Jun-amino-terminal kinase-interacting protein 4-like	0.00E+00	11.1954	0.0215
comp22472_c0_seq2	–	–	–	11.7461	0.0215
comp17869_c1_seq1	mrs2	magnesium transporter MRS2 homolog	0.00E+00	-2.6838	0.0229
comp17473_c0_seq16	cacna2d3	calcium channel, voltage dependent, alpha2/delta subunit 3-like	0.00E+00	12.1675	0.0229
comp11431_c0_seq6	–	–	–	11.3186	0.0231
comp26516_c0_seq1	–	–	–	11.2565	0.0234
comp71339_c0_seq1	gmip	GEM-interacting protein-like	0.00E+00	-3.0879	0.0234
comp57987_c0_seq1	ehbp1l1	EH domain-binding protein 1-like protein 1-like	1.93E-123	-11.4101	0.0235
comp6371_c1_seq6	mical3	protein-methionine sulfoxide oxidase mical3a-like, also known as microtubule associated monooxygenase, calponin and LIM domain containing 3	0.00E+00	-2.7119	0.0235
comp11175_c0_seq6	daam2	disheveled-associated activator of morphogenesis 2-like	0.00E+00	3.3155	0.0236
comp13731_c2_seq11	csmd3	CUB and sushi domain-containing protein 3	0.00E+00	2.7768	0.0236
comp14464_c0_seq4	–	–	–	5.1582	0.0236
comp18772_c0_seq1	–	–	–	11.5473	0.0236
comp25477_c0_seq2	tmem57	macoilin-1-like, also known as transmembrane protein 57	2.61E-98	-3.1895	0.0236
comp5388_c0_seq4	rap1gap	RAP1 GTPase activating protein-like	2.68E-67	-2.5748	0.0236
comp56587_c0_seq2	–	–	–	11.2364	0.0236
comp13097_c0_seq21	pogz	pogo transposable element with ZNF domain	6.09E-87	-4.1347	0.0240
comp3576_c0_seq10	epb41l3	erythrocyte membrane protein band 4.1-like 3-like	0.00E+00	4.4384	0.0242
comp5648_c0_seq4	–	–	–	11.2529	0.0245
comp84955_c0_seq3	spf27	pre-mRNA-splicing factor SPF27-like	1.32E-80	-10.9561	0.0245
comp13708_c1_seq5	abcc8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	0.00E+00	-12.4410	0.0251
comp9559_c1_seq3	plxnb1	plexin b1-like	0.00E+00	-2.5870	0.0251
comp5389_c0_seq6	ylpm1	YLP motif-containing protein 1-like	2.36E-44	12.4160	0.0255
comp5668_c0_seq12	syngap1	synaptic ras/Rap GTPase-activating protein SynGAP-like	0.00E+00	11.4855	0.0255
comp11482_c3_seq1	–	–	–	3.0170	0.0257
comp5520_c0_seq5	–	–	–	11.9469	0.0271
comp43982_c0_seq4	filip1	filamin A interacting protein 1-like	0.00E+00	-2.4844	0.0276
comp68_c2_seq1	rhoa	transforming protein RhoA-like	1.93E-126	-11.8780	0.0276
comp6360_c0_seq19	frmpd1	FERM and PDZ domain-containing protein 1-like	0.00E+00	-6.9036	0.0280
comp13867_c0_seq2	tsnare1	t-SNARE domain-containing protein 1-like	8.99E-165	-5.8095	0.0281

comp1424_c1_seq9	arhgap44	Rho GTPase activating protein 44-like	5.08E-26	1.8892	0.0281
comp6622_c0_seq1	snx4	sorting nexin 4-like	0.00E+00	-10.8939	0.0281
comp112_c4_seq75	–	–	–	-11.4635	0.0293
comp135342_c0_seq1	–	–	–	11.2486	0.0298
comp36989_c0_seq2	abl2	Abelson tyrosine-protein kinase 2, also known as c-abl oncogene 2, non-receptor tyrosine kinase	0.00E+00	11.2306	0.0298
comp14642_c0_seq7	ncoa2	nuclear receptor coactivator 2-like	1.59E-74	11.1190	0.0299
comp21226_c0_seq2	gclm	glutamate-cysteine ligase regulatory subunit-like	2.59E-118	-5.1692	0.0299
comp26424_c0_seq1	gtf3c5	general transcription factor IIIC, polypeptide 5-like	0.00E+00	-10.7095	0.0299
comp47128_c0_seq1	ptprs	protein tyrosine phosphatase, receptor type, s-like	1.30E-93	12.8250	0.0299
comp55736_c0_seq1	–	–	–	11.0679	0.0304
comp67091_c0_seq3	ccdc37	coiled-coil domain-containing protein 37-like	4.39E-155	11.0720	0.0304
comp16012_c0_seq1	ctnnd1	catenin (cadherin-associated protein), delta 1-like	0.00E+00	-11.2137	0.0306
comp2986_c0_seq3	–	–	–	-3.8228	0.0307
comp3567_c5_seq14	–	–	–	5.2380	0.0307
comp37088_c0_seq4	ryr1	ryanodine receptor 1-like	0.00E+00	11.9303	0.0307
comp70622_c0_seq1	–	–	–	11.2306	0.0318
comp12653_c1_seq1	srgap2	SLIT-ROBO Rho GTPase-activating protein 2-like	0.00E+00	-2.3483	0.0323
comp270_c0_seq3	atp1b4	ATPase, Na+/K+ transporting, beta 4-like	0.00E+00	5.9773	0.0330
comp102583_c0_seq1	–	–	–	11.1706	0.0332
comp38669_c0_seq1	–	–	–	11.1493	0.0332
comp31470_c1_seq2	–	–	–	11.1057	0.0333
comp59089_c0_seq1	–	–	–	3.6258	0.0334
comp31354_c0_seq6	thada	thyroid adenoma-associated protein homolog	0.00E+00	-11.8275	0.0338
comp41508_c1_seq12	–	–	–	-10.9026	0.0340
comp8108_c0_seq2	–	–	–	-13.2412	0.0340
comp25789_c0_seq1	ddit4	DNA-damage-inducible transcript 4	1.54E-104	-2.3496	0.0356
comp18156_c0_seq5	dixdc1	dixin-like	0.00E+00	-2.2027	0.0361
comp15273_c0_seq3	grik2	glutamate receptor, ionotropic, kainate 2-like	4.74E-93	11.4939	0.0361
comp5979_c0_seq2	sephs2	selenide, water dikinase 2-like	5.71E-178	-10.8016	0.0361
comp8913_c0_seq3	grid2ip	delphilin-like, also known as glutamate receptor, ionotropic, delta 2-interacting protein 1	0.00E+00	11.0006	0.0361
comp24026_c0_seq4	tnrc18	trinucleotide repeat containing 18	3.84E-55	2.9825	0.0368
comp3506_c0_seq261	fryl	protein furry homolog-like	0.00E+00	-13.8457	0.0368
comp38092_c0_seq3	eogt	EGF domain-specific O-linked N-acetylglucosamine transferase-like (glycosyltransferase)	4.60E-102	-10.8529	0.0369
comp96640_c0_seq1	–	–	–	4.1681	0.0369
comp11_c35_seq3	ttll12	tubulin tyrosine ligase-like family, member 12	0.00E+00	-13.1279	0.0371
comp3753_c2_seq5	ptpre	receptor-type tyrosine-protein phosphatase E-like	5.80E-150	11.2805	0.0371
comp4820_c1_seq3	mapt	microtubule-associated protein tau-like	3.86E-10	11.0969	0.0371
comp34017_c0_seq3	rftn2	raftlin family member 2-like	0.00E+00	-4.4316	0.0376
comp35860_c0_seq7	mcm7	DNA replication licensing factor mcm7-like, also known as minichromosome maintenance complex component 7	0.00E+00	-6.4671	0.0376
comp62161_c0_seq1	–	–	–	11.0401	0.0390
comp52914_c0_seq4	gas8	growth arrest-specific protein 8-like	8.09E-59	-4.3277	0.0391
comp15561_c0_seq1	mch	melanin concentrating hormone	1.77E-31	-2.5230	0.0391
comp22347_c0_seq1	ngb	neuroglobin-like	1.37E-88	-10.5722	0.0417
comp21808_c0_seq2	wdr59	WD repeat-containing protein 59	0.00E+00	11.4652	0.0419
comp7940_c1_seq2	psip1	PC4 and SFRS1 interacting protein 1, also known as lens epithelium-derived growth factor (LEDGF/p75)	0.00E+00	-2.9516	0.0419
comp3402_c3_seq18	mef2d	myocyte-specific enhancer factor 2D homolog	4.93E-90	-10.7062	0.0423
comp66733_c0_seq1	–	–	–	11.2306	0.0429
comp94178_c0_seq1	–	–	–	11.2665	0.0429
comp15402_c0_seq2	slc7a1	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1-like	5.80E-46	-10.7128	0.0434
comp10995_c0_seq3	–	–	–	11.3690	0.0438
comp23114_c0_seq5	ift172	intraflagellar transport protein 172 homolog	0.00E+00	-4.2611	0.0438
comp26794_c0_seq23	–	–	–	11.8267	0.0438
comp7367_c1_seq5	ptprk	protein tyrosine phosphatase, receptor type, K-like	0.00E+00	1.7496	0.0438
comp20660_c0_seq37	mdga2	MAM domain-containing glycosylphosphatidylinositol anchor protein 2	2.62E-57	-10.4943	0.0438
comp292_c6_seq6	map4k4	mitogen-activated protein kinase kinase kinase kinase 4-like	0.00E+00	-4.3908	0.0438
comp3171_c1_seq3	usp9	ubiquitin specific peptidase 9	0.00E+00	2.3300	0.0438
comp35332_c1_seq2	–	–	–	10.9326	0.0438
comp52653_c0_seq1	spata18	Spermatogenesis associated 18, also known as mitochondria-eating protein-like	0.00E+00	-10.7863	0.0438
comp3177_c0_seq24	dcun1d4	DCN1, defective in cullin neddylation 1, domain containing 4	8.37E-69	-12.7532	0.0444
comp2446_c4_seq5	slc24a2	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2-like	4.30E-48	11.1493	0.0451
comp1376_c0_seq34	ank2	ankyrin 2	0.00E+00	-2.3946	0.0458
comp32967_c0_seq1	–	–	–	10.9273	0.0463
comp19280_c0_seq2	pip5k1a	phosphatidylinositol-4-phosphate 5-kinase, type I, alpha-like	0.00E+00	10.9955	0.0484
comp2986_c0_seq1	–	–	–	-2.1918	0.0496

Table S10 – Gene ontology enrichment for transcripts up-regulated in each phenotype across pairwise comparisons. Conditional enrichment was obtained with unadjusted $P < 0.01$.

(a) Nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0070588	calcium ion transmembrane transport	1.8798	8	0.0006
GO:0070838	divalent metal ion transport	2.6416	9	0.0014
GO:0033554	cellular response to stress	2.7342	8	0.0063
GO:0044242	cellular lipid catabolic process	1.1891	5	0.0069
GO:0098660	inorganic ion transmembrane transport	5.3189	12	0.0072
Cellular Component				
GO:0005891	voltage-gated calcium channel complex	1.1106	8	0.0000
GO:1902495	transmembrane transporter complex	3.2363	11	0.0004
GO:0034703	cation channel complex	2.7253	8	0.0060
GO:0034399	nuclear periphery	0.1363	2	0.0081
Molecular Function				
GO:0005245	voltage-gated calcium channel activity	1.2344	8	0.0000
GO:0015085	calcium ion transmembrane transporter activity	2.3164	8	0.0024
GO:0005261	cation channel activity	4.7471	11	0.0085

(b) Females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0070588	calcium ion transmembrane transport	2.0816	8	0.0012
GO:0043968	histone H2A acetylation	0.0631	2	0.0017
GO:0098660	inorganic ion transmembrane transport	5.8900	14	0.0024
GO:0070838	divalent metal ion transport	2.9253	9	0.0028
GO:0098655	cation transmembrane transport	5.6298	13	0.0044
GO:0007610	behaviour	1.1354	5	0.0057
GO:0016578	histone deubiquitination	0.1183	2	0.0061
GO:0034765	regulation of ion transmembrane transport	1.6637	6	0.0067
GO:0051496	positive regulation of stress fibre assembly	0.0079	1	0.0079
GO:0044763	single-organism cellular process	90.4638	106	0.0087
GO:0043436	oxoacid metabolic process	15.2968	25	0.0098
Cellular Component				

GO:1902495	transmembrane transporter complex	3.3645	14	0.0000
GO:0005891	voltage-gated calcium channel complex	1.1545	8	0.0000
GO:0034703	cation channel complex	2.8332	11	0.0001
GO:0033276	transcription factor TFTC complex	0.0425	2	0.0007
GO:0030914	STAGA complex	0.0567	2	0.0014
GO:0035267	NuA4 histone acetyltransferase complex	0.1487	2	0.0096
Molecular Function				
GO:0005245	voltage-gated calcium channel activity	1.2905	8	0.0000
GO:0004968	gonadotropin-releasing hormone receptor activity	0.0239	2	0.0002
GO:0005244	voltage-gated ion channel activity	3.5927	11	0.0010
GO:0005261	cation channel activity	4.9629	13	0.0016
GO:0022836	gated channel activity	6.5083	15	0.0024
GO:0015085	calcium ion transmembrane transporter activity	2.4217	8	0.0031
GO:0050277	sedoheptulokinase activity	0.0080	1	0.0080

(c) Sneaker males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0070925	organelle assembly	0.9903	5	0.0032
GO:0048646	anatomical structure formation involved in morphogenesis	3.2636	9	0.0057
GO:0007050	cell cycle arrest	0.3751	3	0.0063
GO:0031032	actomyosin structure organization	0.3751	3	0.0063
GO:0048563	post-embryonic organ morphogenesis	0.0075	1	0.0075
GO:0051496	positive regulation of stress fiber assembly	0.0075	1	0.0075
GO:0030029	actin filament-based process	2.8435	8	0.0080
GO:0006270	DNA replication initiation	0.1425	2	0.0088
Cellular Component				
GO:0005884	actin filament	0.2771	4	0.0002
GO:0043230	extracellular organelle	0.1348	2	0.0079
GO:0070062	extracellular vesicular exosome	0.1348	2	0.0079
GO:0042555	MCM complex	0.1348	2	0.0079
GO:0030016	myofibril	0.4118	3	0.0081

GO:0016942	insulin-like growth factor binding protein complex	0.1498	2	0.0097
Molecular Function				
GO:0004527	exonuclease activity	0.2912	3	0.0031
GO:0050277	sedoheptulokinase activity	0.0075	1	0.0075
GO:0016884	carbon-nitrogen ligase activity, with glutamine as amido-N-donor	0.1344	2	0.0078
GO:0003995	acyl-CoA dehydrogenase activity	0.1493	2	0.0096

(d) Transitional males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0030866	cortical actin cytoskeleton organization	0.1761	3	0.0007
GO:0043085	positive regulation of catalytic activity	1.3674	6	0.0025
GO:0060632	regulation of microtubule-based movement	0.0093	1	0.0092
GO:0047497	mitochondrion transport along microtubule	0.0093	1	0.0092
GO:0051654	establishment of mitochondrion localization	0.0093	1	0.0092
GO:0031116	positive regulation of microtubule polymerization	0.0093	1	0.0092
GO:0016197	endosomal transport	0.1483	2	0.0096
GO:0044763	single-organism cellular process	53.1817	65	0.0097
Cellular Component				
GO:0034399	nuclear periphery	0.0863	2	0.0033
Molecular Function				
GO:0015269	calcium-activated potassium channel activity	0.2783	4	0.0002
GO:0098772	molecular function regulator	3.9660	11	0.0021
GO:0005089	Rho guanyl-nucleotide exchange factor activity	1.0132	5	0.0036
GO:0005096	GTPase activator activity	0.7045	4	0.0055
GO:0005261	cation channel activity	2.7092	8	0.0059
GO:0071813	lipoprotein particle binding	0.0087	1	0.0087
GO:0034185	apolipoprotein binding	0.0087	1	0.0087
GO:0060589	nucleoside-triphosphatase regulator activity	0.8349	4	0.0100

[1]Number of transcripts in each category expected based on the distribution of categories among all transcripts tested. [2]Number of transcripts conferring the enrichment in each category for each phenotype

Table S11 – Gene ontology enrichment for transcripts down-regulated in each phenotype across pairwise comparisons. Conditional enrichment was obtained with unadjusted $P < 0.01$.

(a) Nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0043968	histone H2A acetylation	0.0256	2	0.0003
GO:0016578	histone deubiquitination	0.0480	2	0.0010
GO:0043967	histone H4 acetylation	0.0768	2	0.0027
GO:0007093	mitotic cell cycle checkpoint	0.0800	2	0.0029
GO:0030240	skeletal muscle thin filament assembly	0.0064	1	0.0064
GO:0043504	mitochondrial DNA repair	0.0064	1	0.0064
GO:0032484	Ral protein signal transduction	0.0096	1	0.0096
Cellular Component				
GO:0033276	transcription factor TFTC complex	0.0194	2	0.0002
GO:0030914	STAGA complex	0.0259	2	0.0003
GO:0035267	NuA4 histone acetyltransferase complex	0.0680	2	0.0021
GO:1902562	H4 histone acetyltransferase complex	0.0745	2	0.0025
GO:0005884	actin filament	0.1198	2	0.0064
GO:0005916	fascia adherens	0.0097	1	0.0097
Molecular Function				
GO:0008297	single-stranded DNA exodeoxyribonuclease activity	0.0060	1	0.0060
GO:0004529	exodeoxyribonuclease activity	0.0090	1	0.0090

(b) Females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0010970	microtubule-based transport	0.0572	2	0.0015
GO:0051051	negative regulation of transport	0.0870	2	0.0034
GO:0060632	regulation of microtubule-based movement	0.0050	1	0.0050
GO:0010869	regulation of receptor biosynthetic process	0.0050	1	0.0050
GO:0047497	mitochondrion transport along microtubule	0.0050	1	0.0050
GO:0051654	establishment of mitochondrion localization	0.0050	1	0.0050
GO:0030187	melatonin biosynthetic process	0.0050	1	0.0050
GO:0031116	positive regulation of microtubule polymerization	0.0050	1	0.0050

GO:0007050	cell cycle arrest	0.1242	2	0.0069
GO:0003094	glomerular filtration	0.0075	1	0.0074
GO:0006837	serotonin transport	0.0099	1	0.0099
Cellular Component				
GO:0071203	WASH complex	0.0599	2	0.0016
GO:0015629	actin cytoskeleton	0.7163	4	0.0055
GO:0005916	fascia adherens	0.0075	1	0.0075
Molecular Function				
GO:0003779	actin binding	1.2441	6	0.0015
GO:0004527	exonuclease activity	0.1081	2	0.0053
GO:0071813	lipoprotein particle binding	0.0055	1	0.0055
GO:0034185	apolipoprotein binding	0.0055	1	0.0055
GO:0003844	1,4-alpha-glucan branching enzyme activity	0.0055	1	0.0055
GO:0047961	glycine N-acyltransferase activity	0.0083	1	0.0083
GO:0004613	phosphoenolpyruvate carboxykinase (GTP) activity	0.0083	1	0.0083

(c) Sneaker males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0098655	cation transmembrane transport	3.5485	12	0.0002
GO:0098660	inorganic ion transmembrane transport	3.7125	12	0.0003
GO:0070588	calcium ion transmembrane transport	1.3121	7	0.0004
GO:0070838	divalent metal ion transport	1.8438	8	0.0005
GO:0034765	regulation of ion transmembrane transport	1.0486	5	0.0041
GO:0044765	single-organism transport	13.8511	24	0.0045
GO:0071805	potassium ion transmembrane transport	1.2027	5	0.0072
GO:0006813	potassium ion transport	1.7246	6	0.0078
GO:0060632	regulation of microtubule-based movement	0.0099	1	0.0099
GO:0047497	mitochondrion transport along microtubule	0.0099	1	0.0099
GO:0051654	establishment of mitochondrion localization	0.0099	1	0.0099
GO:0043504	mitochondrial DNA repair	0.0099	1	0.0099
GO:0031116	positive regulation of microtubule polymerization	0.0099	1	0.0099

Cellular Component				
GO:0005891	voltage-gated calcium channel complex	0.7367	6	0.0001
GO:1902495	transmembrane transporter complex	2.1469	9	0.0003
GO:0034703	cation channel complex	1.8079	8	0.0004
GO:0034399	nuclear periphery	0.0904	2	0.0036
GO:0016323	basolateral plasma membrane	0.1130	2	0.0056
GO:0005782	peroxisomal matrix	0.0090	1	0.0090
Molecular Function				
GO:0005261	cation channel activity	3.1168	13	0.0000
GO:0022838	substrate-specific channel activity	4.8978	16	0.0000
GO:0022803	passive transmembrane transporter activity	4.9228	16	0.0000
GO:0022836	gated channel activity	4.0874	14	0.0001
GO:0004968	gonadotropin-releasing hormone receptor activity	0.0150	2	0.0001
GO:0005245	voltage-gated calcium channel activity	0.8105	6	0.0002
GO:0015269	calcium-activated potassium channel activity	0.3202	4	0.0003
GO:0015085	calcium ion transmembrane transporter activity	1.5209	7	0.0009
GO:0022890	inorganic cation transmembrane transporter activity	4.7978	13	0.0011
GO:0005244	voltage-gated ion channel activity	2.2563	8	0.0020
GO:0015075	ion transmembrane transporter activity	7.8545	16	0.0053
GO:0022892	substrate-specific transporter activity	8.6300	17	0.0056
GO:0071813	lipoprotein particle binding	0.0100	1	0.0100
GO:0008297	single-stranded DNA exodeoxyribonuclease activity	0.0100	1	0.0100
GO:0034185	apolipoprotein binding	0.0100	1	0.0100

(d) Transitional males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0070838	divalent metal ion transport	2.4821	8	0.0036
GO:0044242	cellular lipid catabolic process	1.1173	5	0.0054
GO:0048563	post-embryonic organ morphogenesis	0.0067	1	0.0067
GO:0051496	positive regulation of stress fiber assembly	0.0067	1	0.0067

GO:0006812	cation transport	8.9181	17	0.0078
GO:0070588	calcium ion transmembrane transport	1.7662	6	0.0088
Cellular Component				
GO:0005891	voltage-gated calcium channel complex	1.0006	6	0.0005
GO:1990351	transporter complex	2.9159	9	0.0025
GO:0034702	ion channel complex	2.7317	8	0.0060
GO:0016942	insulin-like growth factor binding protein complex	0.1228	2	0.0066
GO:0071203	WASH complex	0.1473	2	0.0094
Molecular Function				
GO:0004968	gonadotropin-releasing hormone receptor activity	0.0204	2	0.0001
GO:0005245	voltage-gated calcium channel activity	1.1035	6	0.0009
GO:0072509	divalent inorganic cation transmembrane transporter activity	2.1593	7	0.0062
GO:0050277	sedoheptulokinase activity	0.0068	1	0.0068
GO:0003995	acyl-CoA dehydrogenase activity	0.1362	2	0.0081
GO:0005520	insulin-like growth factor binding	0.1430	2	0.0089

[1]Number of transcripts in each category expected based on the distribution of categories among all transcripts tested. [2]Number of transcripts conferring the enrichment in each category for each phenotype

Table S12 – List of female-biased transcripts up-regulated in sneaker males relative to nest-holder males (N = 37). Direction of expression of these transcripts in transitional males is specified, either being up-regulated or not statistically significant (n.s.).

Contig identifier	Gene symbol	Gene name	Transitional males
comp21751_c0_seq105	trrap	transformation/transcription domain-associated protein	n.s.
comp211_c0_seq5	ralgps1	ras-specific guanine nucleotide-releasing factor RalGPS1-like	n.s.
comp12858_c0_seq53	plxna2	plexin A2-like	n.s.
comp36007_c0_seq1	nrg1	pro-neuregulin-1, membrane-bound isoform-like	n.s.
comp18811_c0_seq1	map3k9	mitogen-activated protein kinase kinase kinase 9-like	n.s.
comp3137_c0_seq5	fubp1	far upstream element-binding protein 1-like	n.s.
comp57997_c0_seq8	dla	delta-like protein A-like	n.s.
comp19166_c0_seq8	c-fos	proto-oncogene c-Fos-like	n.s.
comp23805_c0_seq11	celf1	cugbp, Elav-like family member 1-like	n.s.

comp11119_c0_seq24	cdkl5	cyclin-dependent kinase-like 5-like	n.s.
comp15688_c1_seq7	cacna1i	calcium channel, voltage-dependent, T type, alpha 1I subunit	n.s.
comp28515_c0_seq2	–	–	n.s.
comp20893_c0_seq1	traf2	Tnf receptor-associated factor 2	up
comp26642_c0_seq8	tmem62	transmembrane protein 62-like	up
comp14699_c0_seq4	srgap3	SLIT-ROBO Rho GTPase activating protein 3	up
comp25676_c0_seq3	satb1	DNA-binding protein SATB1-like	up
comp5839_c0_seq25	nfasc	neurofascin-like	up
comp5839_c0_seq35	nfasc	neurofascin-like	up
comp4110_c3_seq3	ncor2	nuclear receptor corepressor 2-like	up
comp292_c6_seq16	map4k4	mitogen-activated protein kinase kinase kinase kinase 4-like	up
comp456_c4_seq4	lgi1	leucine-rich glioma-inactivated protein 1-like	up
comp3506_c0_seq308	fryl	protein furry homolog-like	up
comp3506_c0_seq82	fryl	protein furry homolog-like	up
comp26171_c0_seq2	fmnl2	formin-like protein 2-like	up
comp9181_c0_seq15	fam171b	protein FAM171B-like	up
comp7272_c6_seq10	dnmt3a	DNA (cytosine-5-)-methyltransferase 3 alpha-like	up
comp13368_c1_seq10	dnah10	dynein, axonemal, heavy chain 10-like	up
comp24803_c0_seq5	bahcc1	BAH and coiled-coil domain-containing protein 1-like	up
comp3852_c0_seq13	arhgef9	Cdc42 guanine nucleotide exchange factor (GEF) 9	up
comp203_c7_seq1	ankrd34a	ankyrin repeat domain 34a-like	up
comp2068_c1_seq5	–	–	up
comp2075_c4_seq61	–	–	up
comp23344_c0_seq2	–	–	up
comp700_c4_seq17	–	–	up
comp72_c10_seq25	–	–	up
comp9163_c1_seq50	–	–	up
comp93579_c0_seq2	–	–	up

Table S13 – Gene ontology enrichment for shared expression of female-biased transcripts. Conditional enrichment was obtained with unadjusted $P < 0.01$.

(a) Up-regulated in both females and sneaker males relative to nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0032485	regulation of Ral protein signal transduction	0.0027	1	0.0027
GO:0048856	anatomical structure development	2.4080	7	0.0065
GO:0043968	histone H2A acetylation	0.0073	1	0.0072
GO:0044767	single-organism developmental process	2.4560	7	0.0073
GO:0007399	nervous system development	0.8008	4	0.0073
GO:0033209	tumour necrosis factor-mediated signalling pathway	0.0082	1	0.0081
GO:0034612	response to tumour necrosis factor	0.0100	1	0.0099
Cellular Component				
GO:0033276	transcription factor TFTC complex	0.0049	1	0.0048
GO:0030914	STAGA complex	0.0065	1	0.0065
Molecular Function				
–				

(b) Up-regulated in females, sneakers and transitional males relative to nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0033209	tumor necrosis factor-mediated signaling pathway	0.0043	1	0.0043
GO:0034612	response to tumor necrosis factor	0.0053	1	0.0052
GO:0007250	activation of NF-kappaB-inducing kinase activity	0.0057	1	0.0057
GO:1901222	regulation of NIK/NF-kappaB signaling	0.0057	1	0.0057
GO:0051865	protein autoubiquitination	0.0062	1	0.0062
GO:0007399	nervous system development	0.4215	3	0.0072
Cellular Component				
GO:0033276	transcription factor TFTC complex	0.0049	1	0.0048
GO:0030914	STAGA complex	0.0065	1	0.0065
Molecular Function				
–				

(c) Up-regulated in females and sneakers, excluding transitional males, relative to nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0032485	regulation of Ral protein signal transduction	0.0013	1	0.0013
GO:0043968	histone H2A acetylation	0.0034	1	0.0034
GO:0016578	histone deubiquitination	0.0065	1	0.0064
Cellular Component				
GO:0033276	transcription factor TFTC complex	0.0028	1	0.0028
GO:0030914	STAGA complex	0.0038	1	0.0038
GO:0035267	NuA4 histone acetyltransferase complex	0.0099	1	0.0099
Molecular Function				
–				

[1]Number of transcripts in each category expected based on the distribution of categories among all transcripts tested. [2]Number of transcripts conferring the enrichment in each category for each comparison under analysis

Table S14 – List of nest-holder-biased transcripts either exclusive to nest-holder males (N = 19) or also up-regulated in sneaker males relative to females (Sn, N = 48). Direction of expression of these transcripts in transitional males (Tr) is specified, either being up-regulated or not statistically significant (n.s.).

Contig identifier	Gene symbol	Gene name	Sn	Tr
comp52574_c0_seq11	znf236	zinc finger protein 236-like	up	up
comp30660_c0_seq3	wdr20	WD repeat-containing protein 20-like	up	up
comp16233_c0_seq2	tdrd3	tudor domain-containing protein 3-like	up	up
comp82826_c0_seq2	tdrd15	tudor domain-containing protein 15-like	up	up
comp6389_c1_seq2	tle2	Transducing-like enhancer protein 4-like	up	up
comp7069_c0_seq4	st7	suppression of tumorigenicity 7 protein homolog	up	up
comp7069_c0_seq2	rnf6	ring finger protein (C3H2C3 type) 6	up	up
comp39182_c0_seq36	erbb4	receptor tyrosine-protein kinase erbB-4-like	up	up
comp3506_c0_seq318	fryl	protein furry homolog-like	up	up
comp7178_c5_seq1	prickle2	prickle homolog 2	up	up
comp45523_c0_seq1	pms1	PMS1 protein homolog 1-like	up	up
comp18110_c0_seq3	pom121	nuclear envelope pore membrane protein POM121-like	up	up
comp24887_c0_seq2	lcorl	Ligand-dependent nuclear receptor corepressor-like protein	up	up
comp22879_c1_seq3	glyatl1	glycine N-acyltransferase-like	up	up
comp17880_c0_seq6	qrich1	Glutamine-rich protein 1-like	up	up

comp7003_c0_seq3	fam20c	extracellular serine/threonine protein kinase FAM20C-like	up	up
comp36143_c1_seq1	dgcr6	DiGeorge syndrome critical region gene 6	up	up
comp23805_c0_seq5	celf1	cugbp, Elav-like family member 1-like	up	up
comp13731_c2_seq198	csmd1	CUB and sushi domain-containing protein 1	up	up
comp4000_c4_seq5	atp1a3b	ATPase, Na+/K+ transporting, alpha 3 polypeptide-like	up	up
comp22092_c0_seq2	–	–	up	up
comp28508_c1_seq1	–	–	up	up
comp3480_c0_seq3	–	–	up	up
comp61048_c0_seq2	–	–	up	up
comp83787_c0_seq2	–	–	up	up
comp3813_c0_seq6	znf207	zinc finger protein 207-like	up	n.s.
comp20190_c0_seq3	fam21c	WASH complex subunit FAM21C-like	up	n.s.
comp80286_c0_seq2	tnnt1	troponin T, slow skeletal muscle-like	up	n.s.
comp14932_c2_seq93	trpm3	transient receptor potential cation channel, subfamily M, member 3-like	up	n.s.
comp3334_c0_seq8	tom1l2	target of myb1–like protein 2-like	up	n.s.
comp35117_c1_seq17	44M2.3	putative RNA exonuclease NEF-sp-like	up	n.s.
comp14425_c1_seq9	parp14	poly (ADP–ribose) polymerase 14-like	up	n.s.
comp36508_c0_seq4	fxr1	fragile X mental retardation syndrome-related protein 1-like	up	n.s.
comp131394_c0_seq1	frmd3	FERM domain-containing protein 3	up	n.s.
comp1842_c1_seq17	dst	dystonin	up	n.s.
comp1842_c1_seq19	dst	dystonin	up	n.s.
comp24708_c0_seq4	dlgap4	discs, large (Drosophila) homolog-associated protein 4-like	up	n.s.
comp18295_c0_seq1	dock9	dedicator of cytokinesis protein 9-like	up	n.s.
comp200_c1_seq1	chchd3	Coiled-coil-helix-coiled-coil-helix domain containing 3	up	n.s.
comp2075_c4_seq32	cacna1h	calcium channel, voltage-dependent, T type, alpha 1H subunit-like	up	n.s.
comp14289_c0_seq3	ankrd13c	ankyrin repeat domain 13C-like	up	n.s.
comp20824_c0_seq1	aldh2	aldehyde dehydrogenase 2	up	n.s.
comp30709_c0_seq2	zc3h7b	zinc finger CCCH domain-containing protein 7B	up	n.s.
comp130526_c0_seq2	–	–	up	n.s.
comp36345_c0_seq2	–	–	up	n.s.
comp55445_c0_seq2	–	–	up	n.s.
comp74053_c0_seq1	–	–	up	n.s.
comp94169_c0_seq1	–	–	up	n.s.

comp26212_c0_seq3	myo1e	unconventional myosin IE-like	n.s.	n.s.
comp1803_c0_seq2	ub	ubiquitin	n.s.	n.s.
comp15454_c0_seq6	tln2	Talin-2-like	n.s.	n.s.
comp1794_c1_seq5	sv2c	synaptic vesicle glycoprotein 2C-like	n.s.	n.s.
comp15419_c1_seq1	shank3	SH3 and multiple ankyrin repeat domains 3	n.s.	n.s.
comp11609_c0_seq15	phlpp1	PH domain leucine-rich repeat-containing protein phosphatase 1-like	n.s.	n.s.
comp8775_c0_seq14	mpp3	membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)-like	n.s.	n.s.
comp1521_c4_seq19	magi1	membrane associated guanylate kinase, WW and PDZ domain containing 1-like	n.s.	n.s.
comp4084_c1_seq3	hnrnpul1	heterogeneous nuclear ribonucleoprotein U-like 1	n.s.	n.s.
comp50821_c0_seq1	rab3il1	guanine nucleotide exchange factor for Rab-3A-like	n.s.	n.s.
comp14814_c3_seq17	fcho2	FCH domain only protein 2-like	n.s.	n.s.
comp11603_c0_seq1	cers2	ceramide synthase 2-like	n.s.	n.s.
comp9691_c0_seq4	c2cd5	C2 calcium-dependent domain containing 5	n.s.	n.s.
comp22780_c0_seq1	ankrd26	ankyrin repeat domain 26-like	n.s.	n.s.
comp124261_c0_seq1	–	–	n.s.	n.s.
comp42357_c0_seq1	–	–	n.s.	n.s.
comp49156_c0_seq2	–	–	n.s.	n.s.
comp6074_c0_seq1	–	–	n.s.	n.s.
comp63278_c0_seq1	–	–	n.s.	n.s.

Contig identifiers indicated in bold denote contigs that were found up-regulated in nest-holder males when compared with both sneaker and transitional males

Table S15 – Gene ontology enrichment for shared expression of nest-holder-biased transcripts. Conditional enrichment was obtained with unadjusted $P < 0.01$.

(a) Up-regulated in nest-holder and sneaker males relative to females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0007050	cell cycle arrest	0.0287	2	0.0004
GO:0010869	regulation of receptor biosynthetic process	0.0011	1	0.0011
GO:2000209	regulation of anoikis	0.0040	1	0.0040
GO:0072595	maintenance of protein localization in organelle	0.0052	1	0.0052
GO:0006621	protein retention in ER lumen	0.0052	1	0.0052
GO:0051220	cytoplasmic sequestering of protein	0.0069	1	0.0069
GO:0016048	detection of temperature stimulus	0.0075	1	0.0074

GO:0050951	sensory perception of temperature stimulus	0.0075	1	0.0074
Cellular Component				
–				
Molecular Function				
GO:0047961	glycine N-acyltransferase activity	0.0024	1	0.0024

(b) Up-regulated in nest-holder and sneaker males, excluding transitional males, relative to females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0007050	cell cycle arrest	0.0167	2	0.0001
GO:0010869	regulation of receptor biosynthetic process	0.0007	1	0.0007
GO:2000209	regulation of anoikis	0.0023	1	0.0023
GO:0072595	maintenance of protein localization in organelle	0.0030	1	0.0030
GO:0006621	protein retention in ER lumen	0.0030	1	0.0030
GO:0051726	regulation of cell cycle	0.0910	2	0.0034
GO:0051220	cytoplasmic sequestering of protein	0.0040	1	0.0040
GO:0016048	detection of temperature stimulus	0.0043	1	0.0043
GO:0050951	sensory perception of temperature stimulus	0.0043	1	0.0043
GO:0045185	maintenance of protein location	0.0097	1	0.0097
GO:0043112	receptor metabolic process	0.0100	1	0.0100
GO:0051651	maintenance of location in cell	0.0100	1	0.0100
Cellular Component				
GO:0071203	WASH complex	0.0097	1	0.0097
Molecular Function				
–				

(c) Up-regulated in nest-holder, sneaker and transitional males relative to females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
–				
Cellular Component				
–				
Molecular Function				
GO:0047961	glycine N-acyltransferase activity	0.0012	1	0.0012

(d) Up-regulated only in nest-holder males relative to females

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0003094	glomerular filtration	0.0007	1	0.0007
GO:0003014	renal system process	0.0024	1	0.0024
GO:0032835	glomerulus development	0.0043	1	0.0043
Cellular Component				
GO:0005916	fascia adherens	0.0008	1	0.0008
GO:0015629	actin cytoskeleton	0.0774	2	0.0022
GO:0044291	cell-cell contact zone	0.0038	1	0.0038
Molecular Function				
–				

[1]Number of transcripts in each category expected based on the distribution of categories among all transcripts tested. [2]Number of transcripts conferring the enrichment in each category for each comparison under analysis

Table S16 – Summary of correlations between modules' eigengene (ME) and each phenotype, with respective *P*-values, including both unadjusted and adjusted for multiple comparisons.

Module number	Module colour	Size	nest-holder			transitional			sneaker			female		
			correlation	pvalue	padj	correlation	pvalue	padj	correlation	pvalue	padj	correlation	pvalue	padj
ME1	MEaliceblue	275	0.0308	0.9423	0.9781	0.3701	0.3668	0.9258	-0.3223	0.4363	0.9258	-0.0787	0.8531	0.9696
ME2	MEantiquewhite1	3,588	0.4302	0.2873	0.9258	-0.5107	0.1959	0.9140	-0.2813	0.4997	0.9303	0.3618	0.3785	0.9258
ME3	MEantiquewhite2	300	0.6784	0.0644	0.8000	-0.0266	0.9502	0.9831	-0.5581	0.1506	0.9128	-0.0937	0.8253	0.9604
ME4	MEantiquewhite4	173	-0.1751	0.6784	0.9369	0.4245	0.2946	0.9258	-0.4085	0.3150	0.9258	0.1591	0.7066	0.9385
ME5	MEbisque4	83	-0.0308	0.9423	0.9781	-0.2219	0.5974	0.9369	-0.3385	0.4121	0.9258	0.5912	0.1227	0.9099
ME6	MEblack	23,395	-0.3057	0.4615	0.9272	-0.5748	0.1361	0.9128	0.7154	0.0460	0.7863	0.1652	0.6959	0.9369
ME7	MEblue	203	-0.4582	0.2536	0.9258	0.6661	0.0713	0.8127	-0.1098	0.7958	0.9564	-0.0982	0.8171	0.9570
ME8	MEblue1	80	-0.1779	0.6735	0.9369	-0.1545	0.7148	0.9406	-0.2896	0.4865	0.9303	0.6220	0.0996	0.9099
ME9	MEblue2	151	0.4956	0.2117	0.9258	-0.2537	0.5443	0.9369	0.3203	0.4392	0.9258	-0.5622	0.1469	0.9128
ME10	MEblue3	74	0.3121	0.4517	0.9258	-0.3518	0.3927	0.9258	0.3128	0.4507	0.9258	-0.2730	0.5129	0.9363
ME11	MEblue4	475	0.4658	0.2448	0.9258	0.4660	0.2445	0.9258	-0.1329	0.7537	0.9459	-0.7988	0.0174	0.7480
ME12	MEblueviolet	307	0.3247	0.4326	0.9258	0.1750	0.6786	0.9369	-0.5560	0.1524	0.9128	0.0563	0.8946	0.9705
ME13	MEbrown	276	-0.5149	0.1917	0.9140	-0.1474	0.7275	0.9406	0.3118	0.4521	0.9258	0.3505	0.3947	0.9258
ME14	MEbrown1	157	-0.2855	0.4930	0.9303	-0.3859	0.3451	0.9258	0.2694	0.5188	0.9363	0.4020	0.3235	0.9258
ME15	MEbrown2	1,608	-0.5428	0.1645	0.9128	-0.3409	0.4086	0.9258	0.6738	0.0669	0.8000	0.2099	0.6178	0.9369
ME16	MEbrown3	41,342	0.2451	0.5584	0.9369	0.3842	0.3474	0.9258	-0.7161	0.0457	0.7863	0.0867	0.8382	0.9634
ME17	MEbrown4	169	0.0512	0.9041	0.9705	-0.1158	0.7848	0.9564	-0.5274	0.1792	0.9140	0.5920	0.1221	0.9099
ME18	MEchocolate3	655	-0.6043	0.1125	0.9099	0.1648	0.6965	0.9369	0.4878	0.2201	0.9258	-0.0484	0.9095	0.9705
ME19	MEchocolate4	484	-0.1928	0.6473	0.9369	0.5516	0.1564	0.9128	-0.1908	0.6508	0.9369	-0.1680	0.6909	0.9369
ME20	MEcoral	210	0.1087	0.7977	0.9564	0.1482	0.7261	0.9406	0.2506	0.5494	0.9369	-0.5076	0.1991	0.9140
ME21	MEcoral1	101	-0.5629	0.1463	0.9128	-0.1110	0.7936	0.9564	0.3063	0.4607	0.9272	0.3676	0.3703	0.9258
ME22	MEcoral2	242	-0.7619	0.0280	0.7480	0.2211	0.5987	0.9369	0.4589	0.2528	0.9258	0.0819	0.8470	0.9675
ME23	MEcoral3	466	-0.2632	0.5287	0.9369	-0.2246	0.5928	0.9369	0.6981	0.0542	0.8000	-0.2102	0.6173	0.9369
ME24	MEcoral4	114	-0.0982	0.8170	0.9570	-0.7680	0.0260	0.7480	0.0721	0.8652	0.9696	0.7941	0.0186	0.7480
ME25	MEcornflowerblue	185	-0.1014	0.8111	0.9570	-0.1802	0.6693	0.9369	0.5033	0.2035	0.9220	-0.2217	0.5978	0.9369
ME26	MEcyan	169	0.6701	0.0690	0.8000	-0.4892	0.2186	0.9258	-0.0866	0.8383	0.9634	-0.0942	0.8244	0.9604
ME27	MEdarkgoldenrod4	165	0.0354	0.9338	0.9769	0.2721	0.5144	0.9363	-0.3084	0.4573	0.9272	0.0010	0.9982	0.9983
ME28	MEdarkgreen	415	-0.2010	0.6332	0.9369	0.5155	0.1910	0.9140	-0.7451	0.0339	0.7480	0.4305	0.2870	0.9258
ME29	MEdarkgrey	202	0.0590	0.8897	0.9705	-0.3776	0.3564	0.9258	0.3136	0.4494	0.9258	0.0050	0.9906	0.9983
ME30	MEdarkmagenta	338	-0.5204	0.1861	0.9140	0.1596	0.7058	0.9385	0.6822	0.0623	0.8000	-0.3214	0.4376	0.9258
ME31	MEdarkolivegreen	238	0.2725	0.5139	0.9363	-0.4478	0.2658	0.9258	0.0722	0.8650	0.9696	0.1031	0.8080	0.9570
ME32	MEdarkolivegreen1	1,525	-0.3257	0.4312	0.9258	0.6066	0.1108	0.9099	0.1477	0.7270	0.9406	-0.4287	0.2893	0.9258
ME33	MEdarkolivegreen2	211	-0.2232	0.5952	0.9369	0.5172	0.1893	0.9140	0.2216	0.5979	0.9369	-0.5156	0.1909	0.9140
ME34	MEdarkolivegreen4	625	-0.4336	0.2832	0.9258	0.4973	0.2099	0.9258	0.2038	0.6283	0.9369	-0.2675	0.5218	0.9369
ME35	MEdarkorange	2,549	0.4105	0.3125	0.9258	-0.2736	0.5121	0.9363	0.5775	0.1338	0.9128	-0.7144	0.0465	0.7863

ME36	MEdarkorange2	233	-0.0598	0.8882	0.9705	-0.1714	0.6849	0.9369	-0.2808	0.5005	0.9303	0.5120	0.1946	0.9140
ME37	MEdarkred	446	0.4466	0.2674	0.9258	-0.1476	0.7273	0.9406	0.1106	0.7944	0.9564	-0.4095	0.3137	0.9258
ME38	MEdarkseagreen1	403	-0.3495	0.3961	0.9258	0.3663	0.3722	0.9258	-0.3866	0.3441	0.9258	0.3698	0.3672	0.9258
ME39	MEdarkseagreen2	308	0.0510	0.9045	0.9705	-0.5822	0.1300	0.9099	-0.0498	0.9068	0.9705	0.5809	0.1310	0.9099
ME40	MEdarkseagreen3	155	-0.0009	0.9983	0.9983	0.3249	0.4324	0.9258	-0.3644	0.3748	0.9258	0.0405	0.9242	0.9769
ME41	MEdarkseagreen4	454	-0.6572	0.0766	0.8451	0.5240	0.1825	0.9140	0.2059	0.6248	0.9369	-0.0727	0.8641	0.9696
ME42	MEdarkslateblue	100	-0.2067	0.6233	0.9369	-0.0987	0.8160	0.9570	0.4062	0.3180	0.9258	-0.1007	0.8124	0.9570
ME43	MEdarkturquoise	276	0.1429	0.7357	0.9406	0.2108	0.6163	0.9369	-0.6000	0.1158	0.9099	0.2463	0.5565	0.9369
ME44	MEdarkviolet	210	-0.0757	0.8587	0.9696	-0.1661	0.6942	0.9369	0.2540	0.5439	0.9369	-0.0122	0.9772	0.9946
ME45	MEdeeppink	124	-0.4072	0.3166	0.9258	0.3110	0.4534	0.9258	-0.5859	0.1270	0.9099	0.6821	0.0624	0.8000
ME46	MEdeeppink1	991	-0.0604	0.8870	0.9705	0.6809	0.0630	0.8000	-0.2748	0.5101	0.9363	-0.3457	0.4016	0.9258
ME47	MEdeeppink2	839	0.3778	0.3562	0.9258	-0.1847	0.6615	0.9369	-0.7229	0.0427	0.7863	0.5298	0.1769	0.9140
ME48	MEfirebrick	226	0.2642	0.5272	0.9369	0.3260	0.4307	0.9258	-0.4904	0.2173	0.9258	-0.0998	0.8141	0.9570
ME49	MEfirebrick2	5,573	-0.3003	0.4699	0.9288	-0.0259	0.9515	0.9831	0.5847	0.1279	0.9099	-0.2585	0.5364	0.9369
ME50	MEfirebrick3	156	0.4189	0.3016	0.9258	0.2982	0.4731	0.9300	-0.1677	0.6915	0.9369	-0.5494	0.1584	0.9128
ME51	MEfirebrick4	1,749	0.0927	0.8272	0.9604	-0.5916	0.1224	0.9099	0.3870	0.3436	0.9258	0.1119	0.7920	0.9564
ME52	MEfloralwhite	322	0.4801	0.2285	0.9258	-0.7741	0.0242	0.7480	0.1835	0.6635	0.9369	0.1104	0.7947	0.9564
ME53	MEgreen	554	0.3852	0.3461	0.9258	-0.0637	0.8809	0.9705	-0.1443	0.7331	0.9406	-0.1771	0.6748	0.9369
ME54	MEgreen3	376	-0.0526	0.9015	0.9705	0.5076	0.1991	0.9140	0.2884	0.4886	0.9303	-0.7433	0.0346	0.7480
ME55	MEgreen4	217	-0.0484	0.9093	0.9705	-0.2822	0.4983	0.9303	0.3461	0.4010	0.9258	-0.0155	0.9709	0.9926
ME56	MEgreenyellow	234	0.6340	0.0914	0.8931	0.0148	0.9722	0.9926	-0.6456	0.0838	0.8708	-0.0032	0.9940	0.9983
ME57	MEgrey60	574	-0.6297	0.0943	0.9084	0.4543	0.2582	0.9258	0.5480	0.1596	0.9128	-0.3726	0.3633	0.9258
ME58	MEhoneydew	30	-0.6105	0.1079	0.9099	-0.2454	0.5579	0.9369	0.3290	0.4261	0.9258	0.5269	0.1796	0.9140
ME59	MEhoneydew1	283	-0.4365	0.2796	0.9258	0.3605	0.3803	0.9258	-0.3527	0.3915	0.9258	0.4287	0.2893	0.9258
ME60	MEindianred1	316	-0.0439	0.9179	0.9749	0.0640	0.8804	0.9705	0.2497	0.5509	0.9369	-0.2698	0.5181	0.9363
ME61	MEindianred2	83	-0.2991	0.4718	0.9300	-0.1362	0.7477	0.9459	0.4109	0.3118	0.9258	0.0244	0.9544	0.9831
ME62	MEindianred3	37	-0.2052	0.6259	0.9369	-0.5880	0.1252	0.9099	0.7574	0.0295	0.7480	0.0359	0.9328	0.9769
ME63	MEindianred4	351	-0.1597	0.7057	0.9385	0.8168	0.0133	0.7480	-0.4907	0.2170	0.9258	-0.1665	0.6936	0.9369
ME64	MEivory	1,395	-0.1926	0.6477	0.9369	0.2453	0.5582	0.9369	0.5203	0.1862	0.9140	-0.5730	0.1377	0.9128
ME65	MElavenderblush	1,014	0.2132	0.6123	0.9369	0.4168	0.3042	0.9258	-0.3190	0.4413	0.9258	-0.3111	0.4533	0.9258
ME66	MElavenderblush1	157	0.7771	0.0233	0.7480	-0.3994	0.3269	0.9258	-0.1760	0.6768	0.9369	-0.2017	0.6320	0.9369
ME67	MElavenderblush2	3,271	0.2421	0.5635	0.9369	-0.7129	0.0471	0.7863	0.2229	0.5957	0.9369	0.2480	0.5538	0.9369
ME68	MElavenderblush3	1,696	0.3485	0.3975	0.9258	-0.3370	0.4143	0.9258	-0.3376	0.4134	0.9258	0.3262	0.4304	0.9258
ME69	MElightblue2	510	0.0487	0.9088	0.9705	0.0067	0.9875	0.9983	-0.7429	0.0347	0.7480	0.6875	0.0595	0.8000
ME70	MElightblue3	3,154	0.3129	0.4505	0.9258	0.3471	0.3996	0.9258	-0.3328	0.4205	0.9258	-0.3272	0.4289	0.9258
ME71	MElightblue4	196	0.2606	0.5331	0.9369	-0.6715	0.0682	0.8000	0.5403	0.1668	0.9128	-0.1294	0.7601	0.9459
ME72	MElightcoral	248	0.2868	0.4911	0.9303	-0.5009	0.2061	0.9258	0.3805	0.3524	0.9258	-0.1663	0.6938	0.9369
ME73	MElightcyan	86	0.0517	0.9033	0.9705	-0.1197	0.7777	0.9533	-0.4733	0.2362	0.9258	0.5413	0.1659	0.9128
ME74	MElightcyan1	522	-0.4595	0.2521	0.9258	0.4780	0.2309	0.9258	0.2056	0.6252	0.9369	-0.2242	0.5936	0.9369
ME75	MElightgreen	137	0.1737	0.6808	0.9369	-0.4697	0.2403	0.9258	0.7284	0.0404	0.7814	-0.4325	0.2846	0.9258
ME76	MElightpink1	269	-0.0079	0.9851	0.9983	0.5880	0.1253	0.9099	0.0594	0.8890	0.9705	-0.6394	0.0878	0.8768
ME77	MElightpink2	44	0.1722	0.6835	0.9369	-0.0459	0.9141	0.9724	0.1315	0.7562	0.9459	-0.2578	0.5375	0.9369
ME78	MElightpink3	1,584	0.4890	0.2188	0.9258	0.0462	0.9136	0.9724	-0.7268	0.0411	0.7814	0.1915	0.6495	0.9369
ME79	MElightpink4	2,952	-0.1899	0.6524	0.9369	-0.4433	0.2713	0.9258	0.1080	0.7990	0.9564	0.5252	0.1814	0.9140
ME80	MElightskyblue3	498	-0.7654	0.0269	0.7480	0.3502	0.3951	0.9258	0.3730	0.3628	0.9258	0.0421	0.9211	0.9762
ME81	MElightskyblue4	167	-0.3459	0.4014	0.9258	-0.2249	0.5923	0.9369	0.1269	0.7647	0.9459	0.4439	0.2705	0.9258
ME82	MElightslateblue	1,846	0.1933	0.6465	0.9369	0.3026	0.4662	0.9288	-0.1501	0.7228	0.9406	-0.3458	0.4014	0.9258
ME83	MElightsteelblue	175	0.2465	0.5562	0.9369	0.0351	0.9342	0.9769	0.0331	0.9380	0.9769	-0.3147	0.4478	0.9258
ME84	MElightsteelblue1	450	-0.1962	0.6415	0.9369	0.5967	0.1184	0.9099	0.4249	0.2940	0.9258	-0.8255	0.0116	0.7480
ME85	MElightyellow	205	-0.0522	0.9024	0.9705	0.1378	0.7448	0.9459	0.0017	0.9968	0.9983	-0.0874	0.8369	0.9634
ME86	MEmagenta	325	-0.1667	0.6932	0.9369	-0.2291	0.5853	0.9369	0.5590	0.1498	0.9128	-0.1633	0.6993	0.9379
ME87	MEmagenta1	714	0.1360	0.7481	0.9459	0.1848	0.6613	0.9369	0.3504	0.3949	0.9258	-0.6712	0.0684	0.8000
ME88	MEmagenta2	745	-0.2835	0.4962	0.9303	0.7813	0.0220	0.7480	-0.1644	0.6972	0.9369	-0.3334	0.4197	0.9258
ME89	MEmagenta3	847	-0.3984	0.3283	0.9258	0.4833	0.2251	0.9258	-0.3670	0.3711	0.9258	0.2821	0.4984	0.9303
ME90	MEmagenta4	629	0.0818	0.8473	0.9675	-0.0757	0.8586	0.9696	0.4057	0.3186	0.9258	-0.4118	0.3107	0.9258
ME91	MEmaroon	208	0.3248	0.4324	0.9258	-0.1460	0.7302	0.9406	0.4419	0.2730	0.9258	-0.6208	0.1005	0.9099
ME92	MEmediumorchid	343	0.0329	0.9383	0.9769	-0.2292	0.5850	0.9369	0.2356	0.5743	0.9369	-0.0393	0.9263	0.9769
ME93	MEmediumorchid4	316	-0.1458	0.7305	0.9406	0.5343	0.1725	0.9140	-0.7407	0.0356	0.7480	0.3521	0.3924	0.9258
ME94	MEmediumpurple	227	0.4055	0.3189	0.9258	-0.6928	0.0568	0.8000	0.0028	0.9947	0.9983	0.2844	0.4947	0.9303
ME95	MEmediumpurple1	276	0.4248	0.2942	0.9258	-0.2148	0.6095	0.9369	-0.0557	0.8957	0.9705	-0.1542	0.7153	0.9406
ME96	MEmediumpurple2	102	0.3008	0.4691	0.9288	0.4390	0.2765	0.9258	-0.3796	0.3536	0.9258	-0.3602	0.3808	0.9258
ME97	MEmediumpurple3	280	-0.2331	0.5786	0.9369	-0.5949	0.1198	0.9099	0.6966	0.0549	0.8000	0.1313	0.7566	0.9459
ME98	MEmediumpurple4	404	-0.2844	0.4948	0.9303	0.1068	0.8012	0.9564	0.0338	0.9367	0.9769	0.1437	0.7342	0.9406
ME99	MEmidnightblue	338	0.2032	0.6294	0.9369	-0.1885	0.6548	0.9369	-0.0737	0.8623	0.9696	0.0590	0.8896	0.9705
ME100	MEmistyrose	181	-0.1855	0.6601	0.9369	-0.2457	0.5575	0.9369	0.5648	0.1446	0.9128	-0.1336	0.7524	0.9459
ME101	MEmistyrose4	385	-0.1979	0.6385	0.9369	-0.5558	0.1526	0.9128	0.5243	0.1822	0.9140	0.2294	0.5848	0.9369
ME102	MEmoccasin	182	-0.1890	0.6539	0.9369	0.3214	0.4376	0.9258	-0.3343	0.4184	0.9258	0.2019	0.6315	0.9369
ME103	MEnavajowhite	632	-0.1922	0.6483	0.9369	-0.2507	0.5493	0.9369	0.7274	0.0409	0.7814	-0.2845	0.4946	0.9303
ME104	MEnavajowhite1	355	0.3064	0.4604	0.9272	-0.3846	0.3468	0.9258	-0.2442	0.5600	0.9369	0.3224	0.4361	0.9258
ME105	MEnavajowhite2	145	-0.2026	0.6304	0.9369	-0.2068	0.6232	0.9369	0.6530	0.0792	0.8596	-0.2436	0.5610	0.9369
ME106	MEnavajowhite3	124	-0.1166	0.7834	0.9564	-0.0725	0.8645	0.9696	0.7844	0.0212	0.7480	-0.5953	0.1195	0.9099

ME107	MEorange	128	0.6037	0.1130	0.9099	-0.2084	0.6204	0.9369	-0.3968	0.3304	0.9258	0.0015	0.9971	0.9983
ME108	MEorange4	225	-0.1189	0.7792	0.9535	-0.1265	0.7653	0.9459	-0.2062	0.6242	0.9369	0.4515	0.2614	0.9258
ME109	MEorangered	318	0.0582	0.8911	0.9705	-0.7703	0.0253	0.7480	0.4110	0.3117	0.9258	0.3010	0.4687	0.9288
ME110	MEorangered1	631	-0.4876	0.2203	0.9258	-0.3697	0.3674	0.9258	0.4311	0.2862	0.9258	0.4262	0.2924	0.9258
ME111	MEorangered3	457	0.1053	0.8040	0.9564	0.6931	0.0567	0.8000	-0.0249	0.9534	0.9831	-0.7735	0.0244	0.7480
ME112	MEorangered4	205	-0.1069	0.8011	0.9564	-0.2410	0.5653	0.9369	0.1278	0.7630	0.9459	0.2202	0.6004	0.9369
ME113	MEpaleturquoise	772	0.4757	0.2335	0.9258	-0.4779	0.2310	0.9258	0.2523	0.5466	0.9369	-0.2500	0.5504	0.9369
ME114	MEpaleturquoise4	274	0.4105	0.3124	0.9258	-0.7463	0.0335	0.7480	0.5407	0.1664	0.9128	-0.2050	0.6263	0.9369
ME115	MEpalevioletred	117	0.3566	0.3859	0.9258	-0.2902	0.4856	0.9303	0.3584	0.3834	0.9258	-0.4247	0.2942	0.9258
ME116	MEpalevioletred1	365	0.1538	0.7161	0.9406	0.1985	0.6375	0.9369	-0.3445	0.4033	0.9258	-0.0078	0.9854	0.9983
ME117	MEpalevioletred2	249	-0.3179	0.4429	0.9258	0.5524	0.1556	0.9128	-0.3679	0.3699	0.9258	0.1334	0.7528	0.9459
ME118	MEpalevioletred3	5,975	0.1260	0.7661	0.9459	0.4087	0.3147	0.9258	-0.6452	0.0840	0.8708	0.1105	0.7946	0.9564
ME119	MEpink	96	0.1243	0.7693	0.9464	0.0608	0.8863	0.9705	-0.4368	0.2792	0.9258	0.2517	0.5476	0.9369
ME120	MEpink2	164	-0.2485	0.5529	0.9369	0.0048	0.9909	0.9983	-0.1908	0.6509	0.9369	0.4344	0.2821	0.9258
ME121	MEpink3	516	0.6384	0.0884	0.8768	-0.5685	0.1415	0.9128	0.0912	0.8300	0.9606	-0.1612	0.7030	0.9385
ME122	MEpink4	364	0.3115	0.4526	0.9258	0.0539	0.8990	0.9705	-0.0160	0.9699	0.9926	-0.3494	0.3962	0.9258
ME123	MEplum	288	0.5439	0.1634	0.9128	0.0860	0.8395	0.9634	0.1660	0.6943	0.9369	-0.7960	0.0181	0.7480
ME124	MEplum1	140	0.2717	0.5151	0.9363	-0.3377	0.4133	0.9258	-0.4059	0.3184	0.9258	0.4719	0.2378	0.9258
ME125	MEplum2	171	-0.3387	0.4118	0.9258	-0.3463	0.4008	0.9258	0.5093	0.1974	0.9140	0.1757	0.6773	0.9369
ME126	MEplum3	296	0.9910	0.0000	0.0012	-0.2923	0.4823	0.9303	-0.3525	0.3917	0.9258	-0.3461	0.4010	0.9258
ME127	MEplum4	1,063	0.4183	0.3023	0.9258	0.3292	0.4259	0.9258	-0.2158	0.6078	0.9369	-0.5318	0.1750	0.9140
ME128	MEpowderblue	2,571	0.1516	0.7200	0.9406	-0.4563	0.2558	0.9258	-0.0486	0.9091	0.9705	0.3532	0.3907	0.9258
ME129	MEpurple	170	-0.0160	0.9701	0.9926	0.2100	0.6176	0.9369	0.6045	0.1124	0.9099	-0.7986	0.0175	0.7480
ME130	MEpurple2	1,019	0.3230	0.4352	0.9258	-0.1093	0.7967	0.9564	-0.1991	0.6365	0.9369	-0.0146	0.9726	0.9926
ME131	MEred	3,745	-0.2914	0.4838	0.9303	0.9336	0.0007	0.0793	-0.4489	0.2646	0.9258	-0.1933	0.6465	0.9369
ME132	MEroyalblue	280	-0.5551	0.1532	0.9128	0.1432	0.7351	0.9406	0.2701	0.5176	0.9363	0.1418	0.7377	0.9414
ME133	MEroyalblue2	184	-0.7548	0.0304	0.7480	0.0147	0.9725	0.9926	0.6123	0.1066	0.9099	0.1278	0.7629	0.9459
ME134	MEroyalblue3	495	-0.1677	0.6915	0.9369	0.5936	0.1208	0.9099	-0.1755	0.6776	0.9369	-0.2504	0.5497	0.9369
ME135	MEsaddlebrown	1,139	0.1890	0.6540	0.9369	0.7393	0.0361	0.7480	-0.2461	0.5568	0.9369	-0.6822	0.0624	0.8000
ME136	MEsalmon	7,951	0.3052	0.4623	0.9272	0.5292	0.1774	0.9140	-0.9592	0.0002	0.0376	0.1247	0.7686	0.9464
ME137	MEsalmon1	254	0.4110	0.3117	0.9258	-0.3379	0.4130	0.9258	0.5864	0.1265	0.9099	-0.6596	0.0751	0.8426
ME138	MEsalmon2	175	0.2061	0.6244	0.9369	0.1950	0.6435	0.9369	-0.6738	0.0669	0.8000	0.2727	0.5134	0.9363
ME139	MEsalmon4	591	-0.1138	0.7885	0.9564	-0.3226	0.4357	0.9258	-0.0484	0.9095	0.9705	0.4847	0.2235	0.9258
ME140	MEsienna1	77	0.2959	0.4767	0.9303	-0.0764	0.8574	0.9696	-0.4700	0.2399	0.9258	0.2505	0.5496	0.9369
ME141	MEsienna2	632	-0.0273	0.9488	0.9831	-0.3848	0.3465	0.9258	0.7515	0.0316	0.7480	-0.3394	0.4109	0.9258
ME142	MEsienna3	1,031	-0.3281	0.4276	0.9258	-0.2934	0.4806	0.9303	0.9887	0.0000	0.0012	-0.3672	0.3709	0.9258
ME143	MEsienna4	254	-0.4197	0.3006	0.9258	0.1887	0.6545	0.9369	0.3271	0.4291	0.9258	-0.0960	0.8211	0.9600
ME144	MEskyblue	1,397	0.6053	0.1118	0.9099	-0.6133	0.1059	0.9099	-0.5129	0.1936	0.9140	0.5210	0.1855	0.9140
ME145	MEskyblue1	210	-0.0140	0.9737	0.9926	-0.1451	0.7317	0.9406	0.0920	0.8284	0.9604	0.0671	0.8746	0.9705
ME146	MEskyblue2	621	-0.3897	0.3400	0.9258	0.6423	0.0859	0.8768	-0.0331	0.9379	0.9769	-0.2195	0.6015	0.9369
ME147	MEskyblue3	1,462	0.3194	0.4406	0.9258	-0.9515	0.0003	0.0469	0.2146	0.6098	0.9369	0.4175	0.3034	0.9258
ME148	MEskyblue4	104	-0.2953	0.4776	0.9303	-0.3493	0.3964	0.9258	0.7769	0.0233	0.7480	-0.1323	0.7548	0.9459
ME149	MEslateblue	267	0.0586	0.8904	0.9705	0.4934	0.2141	0.9258	-0.0934	0.8259	0.9604	-0.4585	0.2532	0.9258
ME150	MEslateblue1	927	0.1055	0.8036	0.9564	-0.7740	0.0242	0.7480	0.3574	0.3847	0.9258	0.3111	0.4533	0.9258
ME151	MEsteelblue	749	0.4173	0.3036	0.9258	-0.1520	0.7194	0.9406	-0.0620	0.8841	0.9705	-0.2034	0.6291	0.9369
ME152	MEtan	237	-0.2765	0.5074	0.9363	0.2581	0.5371	0.9369	0.3894	0.3403	0.9258	-0.3710	0.3655	0.9258
ME153	MEtan2	7,213	0.2403	0.5664	0.9369	0.6868	0.0599	0.8000	-0.6232	0.0988	0.9099	-0.3039	0.4643	0.9285
ME154	MEtan3	308	0.4905	0.2172	0.9258	-0.1938	0.6457	0.9369	-0.1233	0.7712	0.9470	-0.1734	0.6814	0.9369
ME155	MEtan4	178	-0.4607	0.2507	0.9258	-0.1014	0.8111	0.9570	0.1570	0.7105	0.9406	0.4051	0.3194	0.9258
ME156	MEthistle	104	0.1306	0.7579	0.9459	-0.2185	0.6032	0.9369	-0.5266	0.1800	0.9140	0.6145	0.1050	0.9099
ME157	MEthistle1	340	0.5280	0.1786	0.9140	-0.7974	0.0178	0.7480	0.3587	0.3829	0.9258	-0.0893	0.8334	0.9629
ME158	MEthistle2	240	0.5472	0.1604	0.9128	-0.4562	0.2559	0.9258	-0.3994	0.3270	0.9258	0.3084	0.4574	0.9272
ME159	MEthistle3	459	0.3307	0.4236	0.9258	0.3584	0.3833	0.9258	-0.2923	0.4824	0.9303	-0.3969	0.3303	0.9258
ME160	MEthistle4	109	0.0061	0.9885	0.9983	-0.3433	0.4051	0.9258	0.1727	0.6825	0.9369	0.1644	0.6972	0.9369
ME161	MEtomato	224	-0.4550	0.2573	0.9258	-0.0677	0.8734	0.9705	-0.0417	0.9219	0.9762	0.5644	0.1450	0.9128
ME162	MEtomato2	117	-0.0716	0.8662	0.9696	0.1387	0.7432	0.9459	-0.2177	0.6046	0.9369	0.1505	0.7220	0.9406
ME163	MEturquoise	137	0.2046	0.6270	0.9369	-0.1806	0.6686	0.9369	0.1316	0.7560	0.9459	-0.1556	0.7130	0.9406
ME164	MEviolet	171	-0.4105	0.3124	0.9258	0.3126	0.4509	0.9258	-0.5505	0.1574	0.9128	0.6483	0.0821	0.8708
ME165	MEwhite	97	-0.1024	0.8094	0.9570	0.1899	0.6524	0.9369	-0.0340	0.9362	0.9769	-0.0535	0.8999	0.9705
ME166	MEwhitesmoke	2,928	0.1599	0.7052	0.9385	0.4461	0.2679	0.9258	0.0772	0.8557	0.9696	-0.6833	0.0617	0.8000
ME167	MEyellow	440	-0.9451	0.0004	0.0543	0.1853	0.6605	0.9369	0.4627	0.2483	0.9258	0.2971	0.4748	0.9303
ME168	MEyellow2	372	0.6739	0.0669	0.8000	-0.0556	0.8959	0.9705	-0.5801	0.1317	0.9099	-0.0381	0.9285	0.9769
ME169	MEyellow3	178	-0.5490	0.1587	0.9128	-0.1829	0.6647	0.9369	0.1708	0.6859	0.9369	0.5611	0.1479	0.9128
ME170	MEyellow4	149	-0.5037	0.2031	0.9220	0.2600	0.5341	0.9369	0.1713	0.6850	0.9369	0.0725	0.8646	0.9696
ME171	MEyellowgreen	578	0.0851	0.8413	0.9639	0.0705	0.8682	0.9704	-0.3254	0.4316	0.9258	0.1698	0.6876	0.9369

Table S17 – Gene ontology enrichment for gene co-expression modules found to be correlated with nest-holder and sneaker males. Conditional enrichment was obtained with unadjusted $P < 0.01$.

(a) plum3 module, positive correlation with nest-holder males

GO ID	GO term	Expected[1]	Observed[2]	pvalue
Biological Process				
GO:0006123	mitochondrial electron transport, cytochrome c to oxygen	0.0545	2	0.0014
GO:0007286	spermatid development	0.0029	1	0.0029
GO:0018108	peptidyl-tyrosine phosphorylation	0.3240	3	0.0041
GO:0046427	positive regulation of JAK-STAT cascade	0.0043	1	0.0043
GO:0042773	ATP synthesis coupled electron transport	0.1004	2	0.0045
GO:0022900	electron transport chain	0.1233	2	0.0067
GO:0007020	microtubule nucleation	0.0086	1	0.0086
GO:0009954	proximal/distal pattern formation	0.0086	1	0.0086
Cellular Component				
GO:0045277	respiratory chain complex IV	0.0666	2	0.0020
Molecular Function				
GO:0004129	cytochrome-c oxidase activity	0.0526	2	0.0013
GO:0016675	oxidoreductase activity	0.0526	2	0.0013
GO:0070064	proline-rich region binding	0.0055	1	0.0055
GO:0004514	nicotinate-nucleotide diphosphorylase (carboxylating) activity	0.0055	1	0.0055
GO:0051082	unfolded protein binding	0.1205	2	0.0065

(b) sienna3 module, positive correlation with sneaker males

GO ID	GO term	Expected[1]	Observed[2]	*pvalue*
Biological Process				
GO:0044335	canonical Wnt signalling pathway involved in neural crest cell differentiation	0.0266	2	0.0002
GO:0000910	cytokinesis	0.8369	6	0.0002
GO:0014823	response to activity	0.1328	3	0.0003
GO:0008643	carbohydrate transport	0.3720	4	0.0005
GO:0044848	biological phase	0.0531	2	0.0010

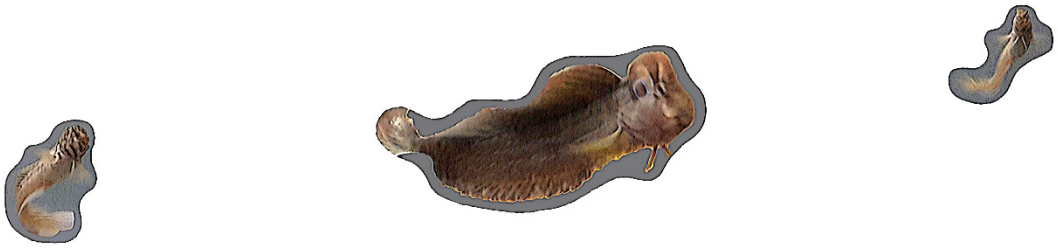
GO:0060037	pharyngeal system development	0.0531	2	0.0010
GO:0051338	regulation of transferase activity	3.1485	10	0.0013
GO:0006955	immune response	1.6606	7	0.0014
GO:0031399	regulation of protein modification process	3.9855	11	0.0024
GO:0006950	response to stress	9.2729	19	0.0026
GO:0045859	regulation of protein kinase activity	2.9758	9	0.0032
GO:0042325	regulation of phosphorylation	4.1715	11	0.0033
GO:0006974	cellular response to DNA damage stimulus	3.5869	10	0.0034
GO:0032094	response to food	0.0930	2	0.0035
GO:0006112	energy reserve metabolic process	0.3321	3	0.0043
GO:0006006	glucose metabolic process	1.6075	6	0.0056
GO:0016051	carbohydrate biosynthetic process	2.1256	7	0.0056
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	0.1196	2	0.0060
GO:0070507	regulation of microtubule cytoskeleton organization	0.3853	3	0.0066
GO:0006302	double-strand break repair	0.3853	3	0.0066
GO:0015780	nucleotide-sugar transport	0.1328	2	0.0074
GO:0051174	regulation of phosphorus metabolic process	4.6763	11	0.0077
GO:0071103	DNA conformation change	1.7403	6	0.0082
GO:0051302	regulation of cell division	0.7971	4	0.0083
GO:0005978	glycogen biosynthetic process	0.1461	2	0.0089
GO:0090090	negative regulation of canonical Wnt signalling pathway	0.1461	2	0.0089
GO:0042312	regulation of vasodilation	0.1461	2	0.0089
Cellular Component				
GO:0098687	chromosomal region	0.3319	4	0.0003
GO:0042613	MHC class II protein complex	0.0553	2	0.0011
GO:0005819	spindle	1.2584	6	0.0016
GO:0000776	kinetochore	0.1245	2	0.0064
Molecular Function				
GO:0004860	protein kinase inhibitor activity	0.1886	3	0.0008
GO:0000049	tRNA binding	0.2290	3	0.0014

GO:0004861	cyclin-dependent protein serine/threonine kinase inhibitor activity	0.1078	2	0.0048
GO:0005338	nucleotide-sugar transmembrane transporter activity	0.1078	2	0.0048
GO:0004731	purine-nucleoside phosphorylase activity	0.1347	2	0.0076
GO:0015238	drug transmembrane transporter activity	0.1482	2	0.0092
GO:0005351	sugar:proton symporter activity	0.1482	2	0.0092

(c) salmon module, negative correlation with sneaker males

GO ID	GO term	Expected[1]	Observed[2]	pvalue
Biological Process				
GO:0015074	DNA integration	0.7528	5	0.0009
GO:0038032	termination of G-protein coupled receptor signalling pathway	1.1994	6	0.0013
GO:0051056	regulation of small GTPase mediated signal transduction	7.5535	17	0.0017
GO:0008277	regulation of G-protein coupled receptor protein signalling pathway	1.4546	6	0.0035
GO:0032387	negative regulation of intracellular transport	0.1021	2	0.0043
GO:0048583	regulation of response to stimulus	15.3621	26	0.0062
GO:0043087	regulation of GTPase activity	6.6220	14	0.0071
Cellular Component				
–				
Molecular Function				
GO:0005227	calcium activated cation channel activity	0.9156	5	0.0023
GO:0030552	cAMP binding	0.0832	2	0.0028
GO:0032266	phosphatidylinositol-3-phosphate binding	0.0951	2	0.0038
GO:0008081	phosphoric diester hydrolase activity	2.1761	7	0.0065
GO:0003964	RNA-directed DNA polymerase activity	0.1427	2	0.0086
GO:0008047	enzyme activator activity	2.3188	7	0.0090

[1]Number of transcripts in each category expected based on the distribution of categories among all transcripts tested. [2]Number of transcripts conferring the enrichment in each category for each module



APPENDIX V |

SUPPLEMENTARY INFORMATION CHAPTER 5

Supplementary information

Cardoso S.D., Gonçalves D., Saraiva J. & Oliveira R.F., 2019. Transcriptomic architecture of courtship sex role reversal in the peacock blenny *Salaria pavo*. *Chapter not submitted for publication*

V.1. Supplementary Figures

Fig. S1 – Boxplots of the shifted logarithm of raw (left column) and normalized counts (right column), for each FPKM threshold (rows) tested in this study to remove low-level expression contigs from the set of forebrain expressed contigs. Further analyses proceeded with a reduced transcriptome whose transcripts had at least 2 mapped FPKM fragments. Each sequenced sample is colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

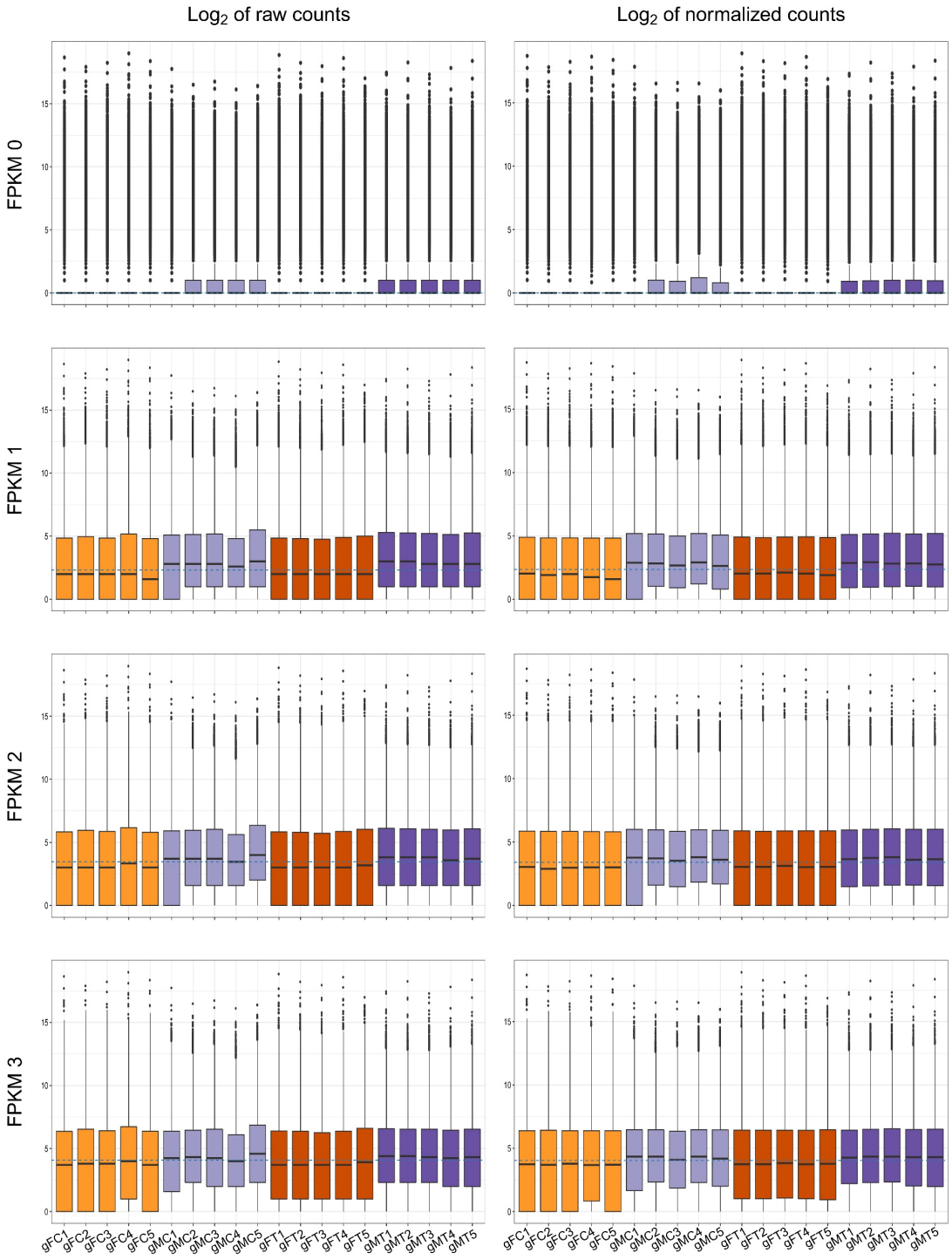


Fig. S2 – Boxplots of the shifted logarithm of raw (left column) and normalized counts (right column), for each FPKM threshold (rows) tested in this study to remove low-level expression contigs from the set of gonad expressed contigs. Further analyses proceeded with a reduced transcriptome whose transcripts had at least 2 mapped FPKM fragments. Each sequenced sample is colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

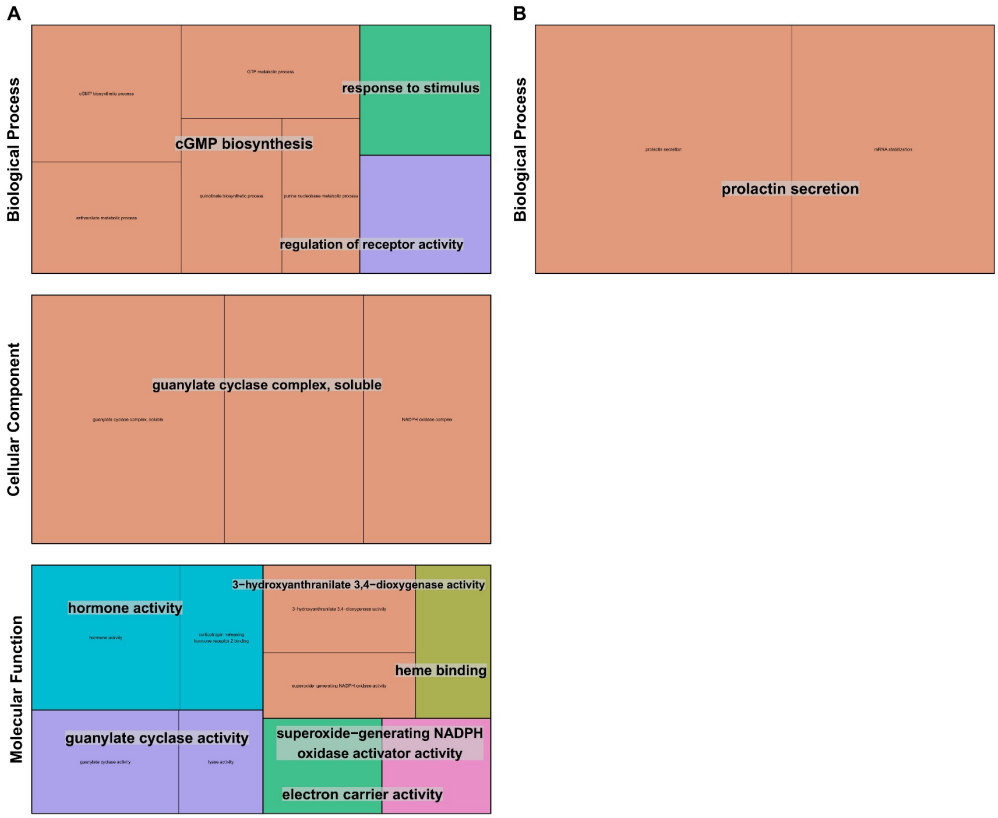


Fig. S3 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the forebrain for the pairwise comparison between A) nest-holder males and B) females from Culatra.

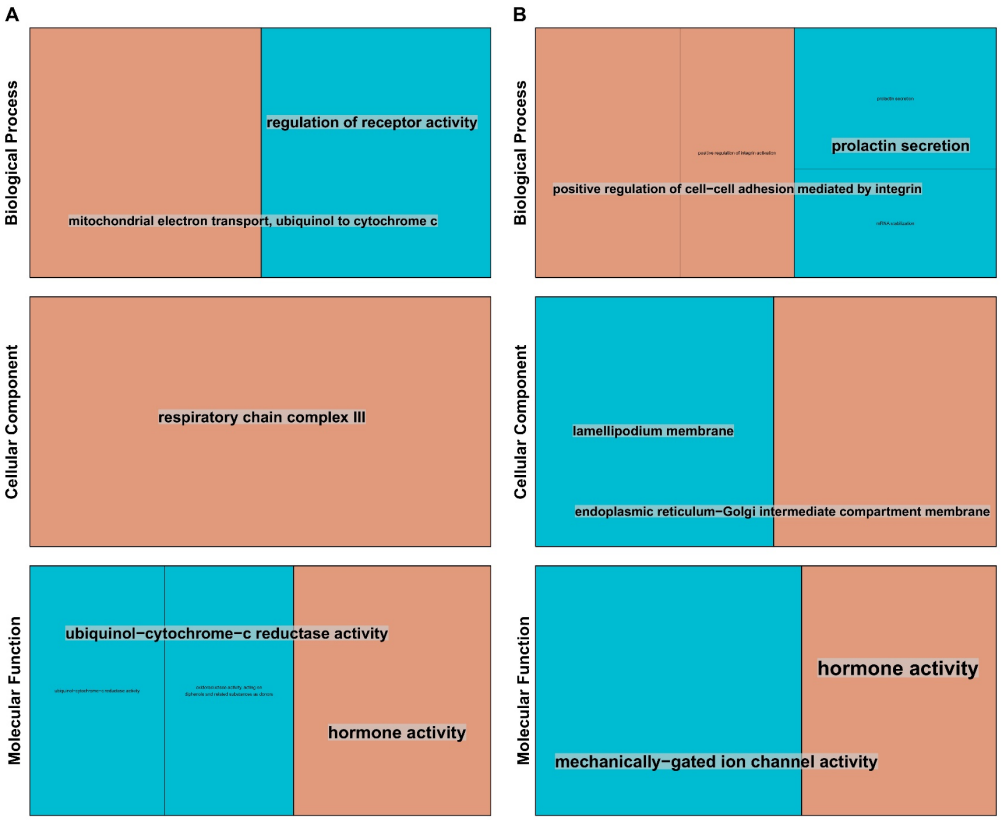


Fig. S4 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the forebrain for the pairwise comparison between A) nest-holder males and B) females from Trieste.

Fig. S5 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the forebrain for the pairwise comparison between A) nest-holders Culatra and B) nest-holders Trieste.

Fig. S6 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the forebrain for the pairwise comparison between A) females Culatra and B) females Trieste.

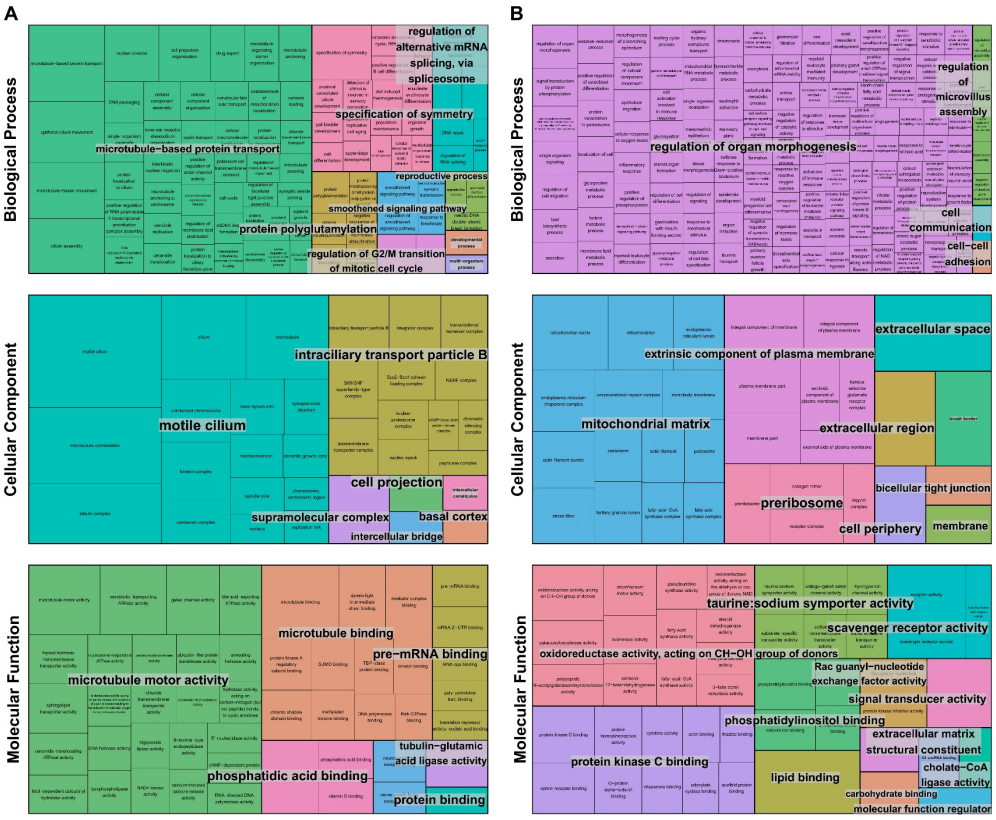


Fig. S7 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the gonad for the pairwise comparison between A) nest-holder males and B) females from Culatra.

Fig. S8 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the gonad for the pairwise comparison between A) nest-holder males and B) females from Trieste.

Fig. S9 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the gonad for the pairwise comparison between A) nest-holders Culatra and B) nest-holders Trieste.

Fig. S10 – Gene Ontology treemap for GO-terms enriched in genes up-regulated in the gonad for the pairwise comparison between A) females Culatra and B) females Trieste.

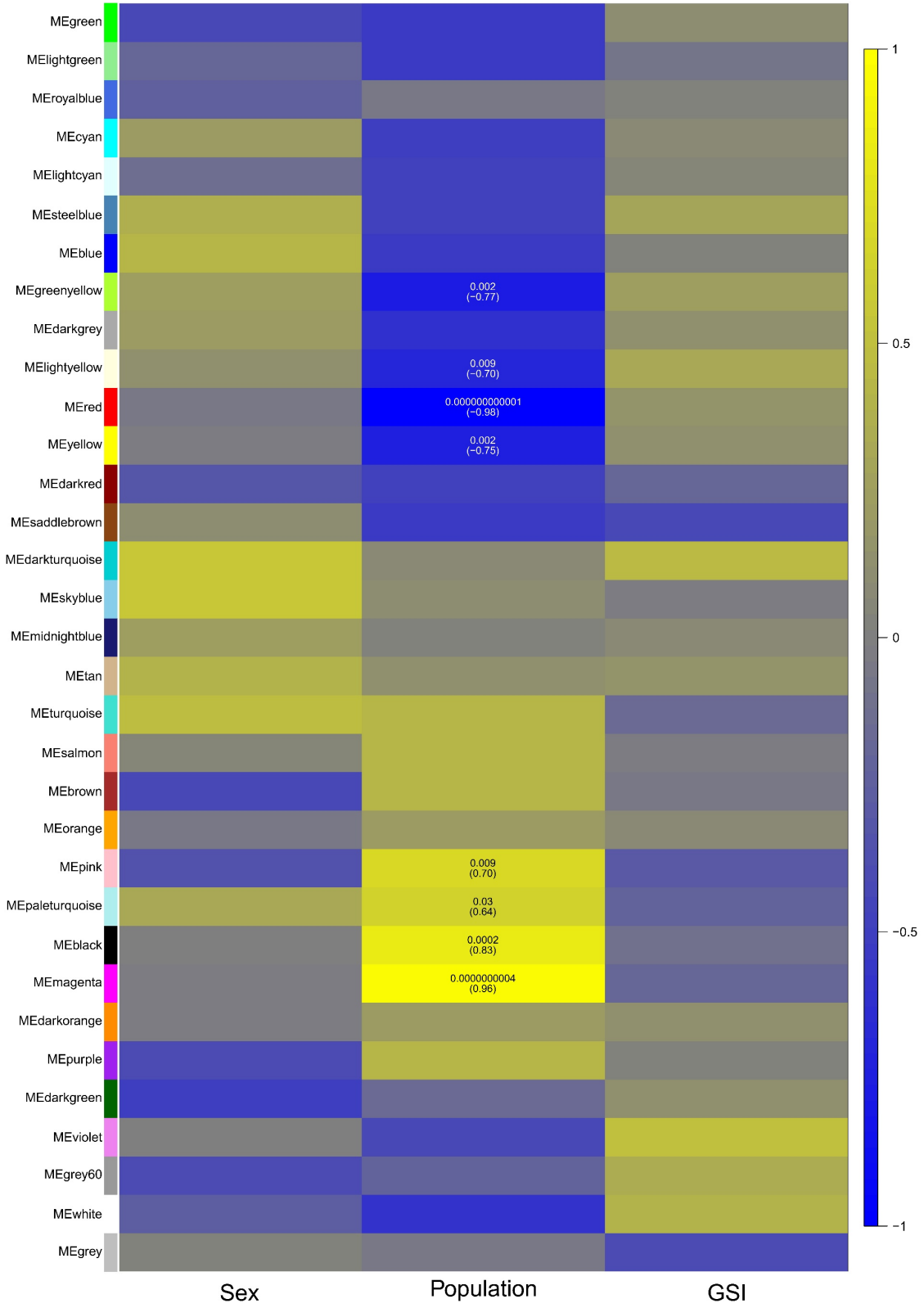


Fig. S11 – Heatmap showing the association between identified forebrain WGCNA gene modules (rows) and each trait under analysis (columns). From the 33 identified modules, eight modules remained significantly associated with 'Population', after correcting the *P*-value for multiple comparisons. Intensity of colour indicates the strength of the correlation coefficient, ranging between negative values (in blue) and positive (in yellow).

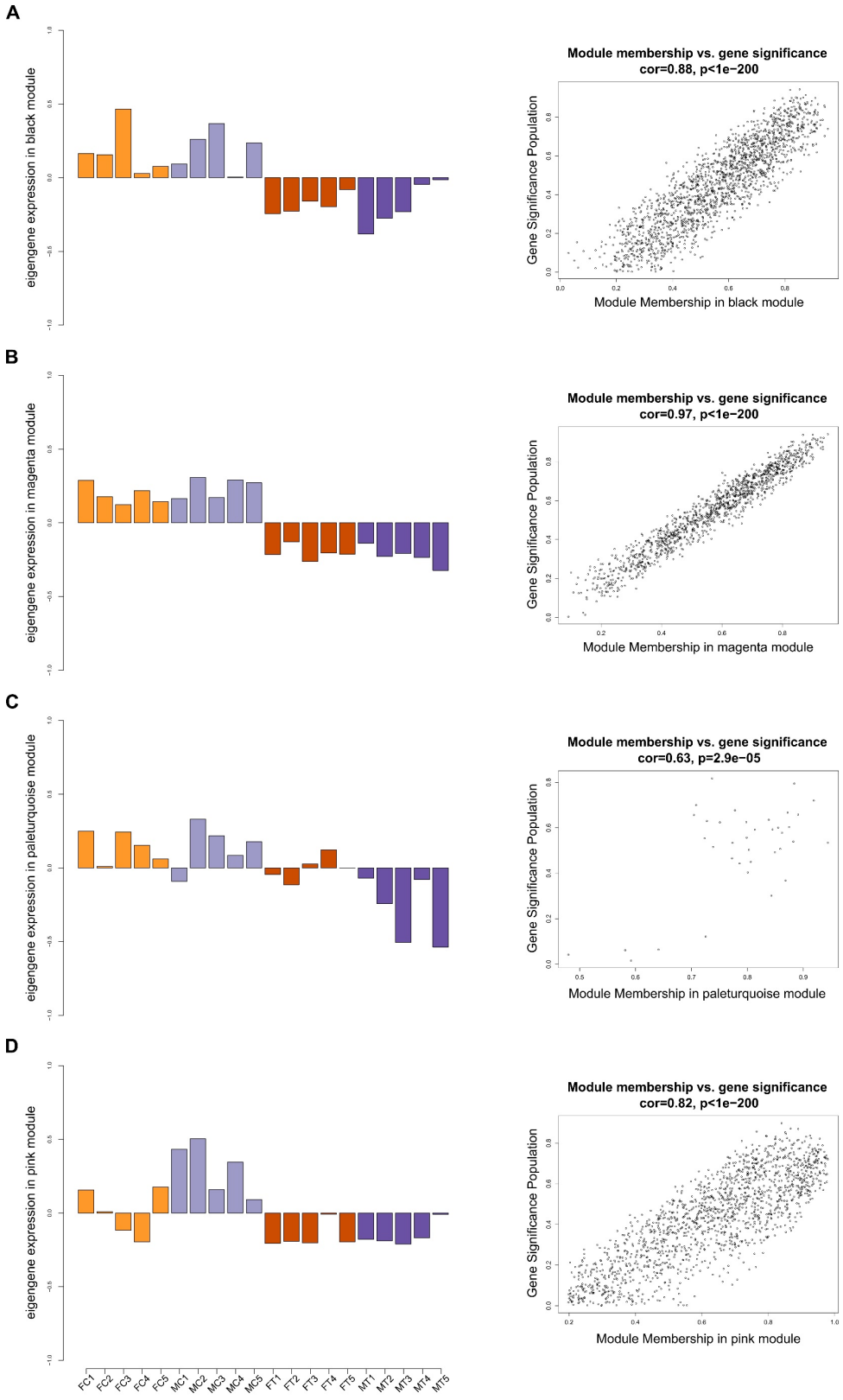


Fig. S12 – Plots for forebrain modules with positive correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

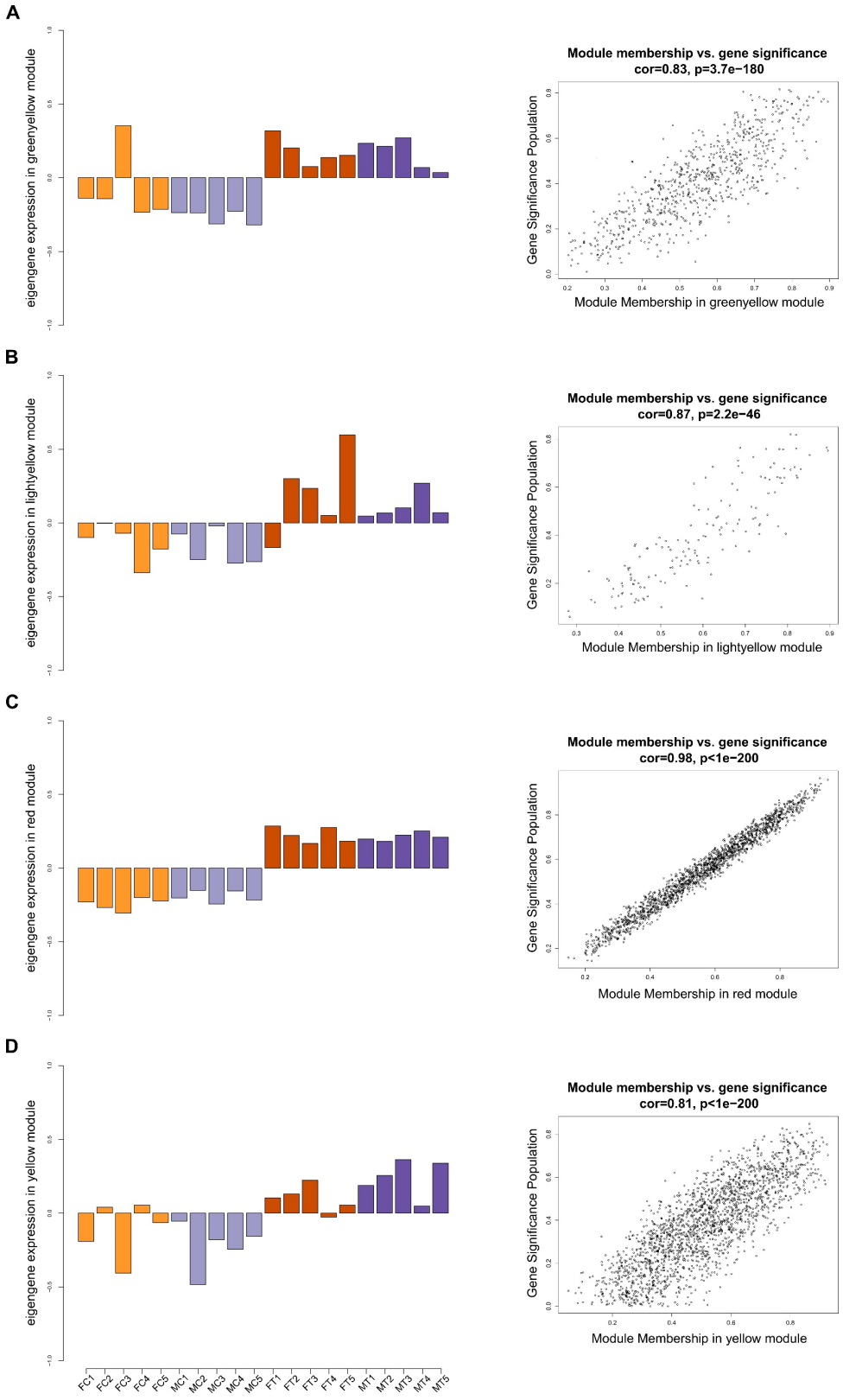


Fig. S13 – Plots for forebrain modules with negative correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light orange for females (FC) and light purple for nest-holder males (MC) from Culatra population, and dark orange for females (FT) and dark purple for nest-holder males (MT) from Trieste population.

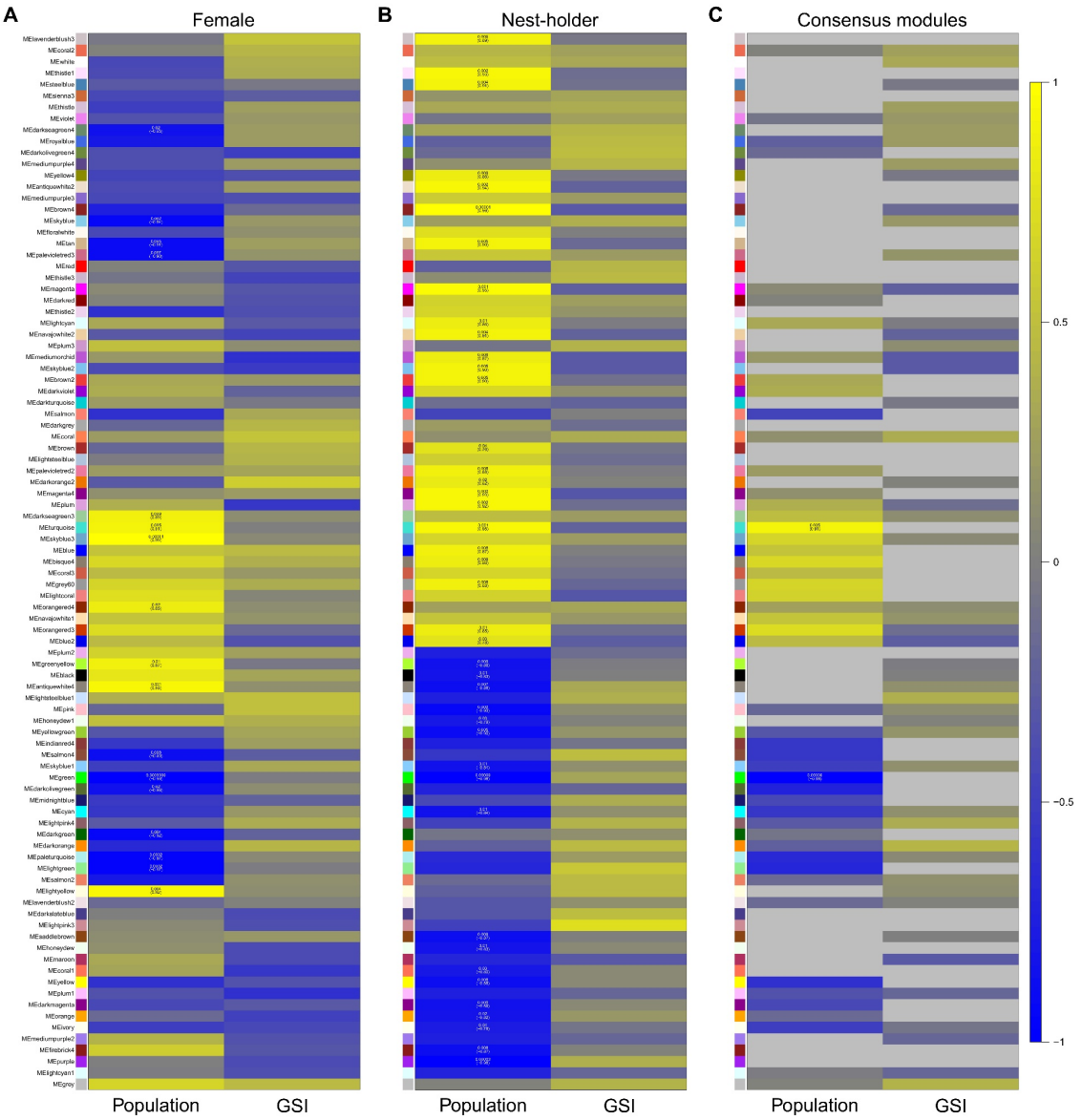


Fig. S14 – Heatmap showing the association between identified gonad WGCNA gene modules (rows) and each trait under analysis (columns) in A) females, B) nest-holder males, and C) consensus network. From the 93 identified modules, 17 modules in females and 42 modules in nest-holder males remained significantly associated with 'Population', after correcting the *P*-value for multiple comparisons, which of these only two modules were consensus between sexes. Intensity of colour indicates the strength of the correlation coefficient, ranging between negative values (in blue) and positive (in yellow).

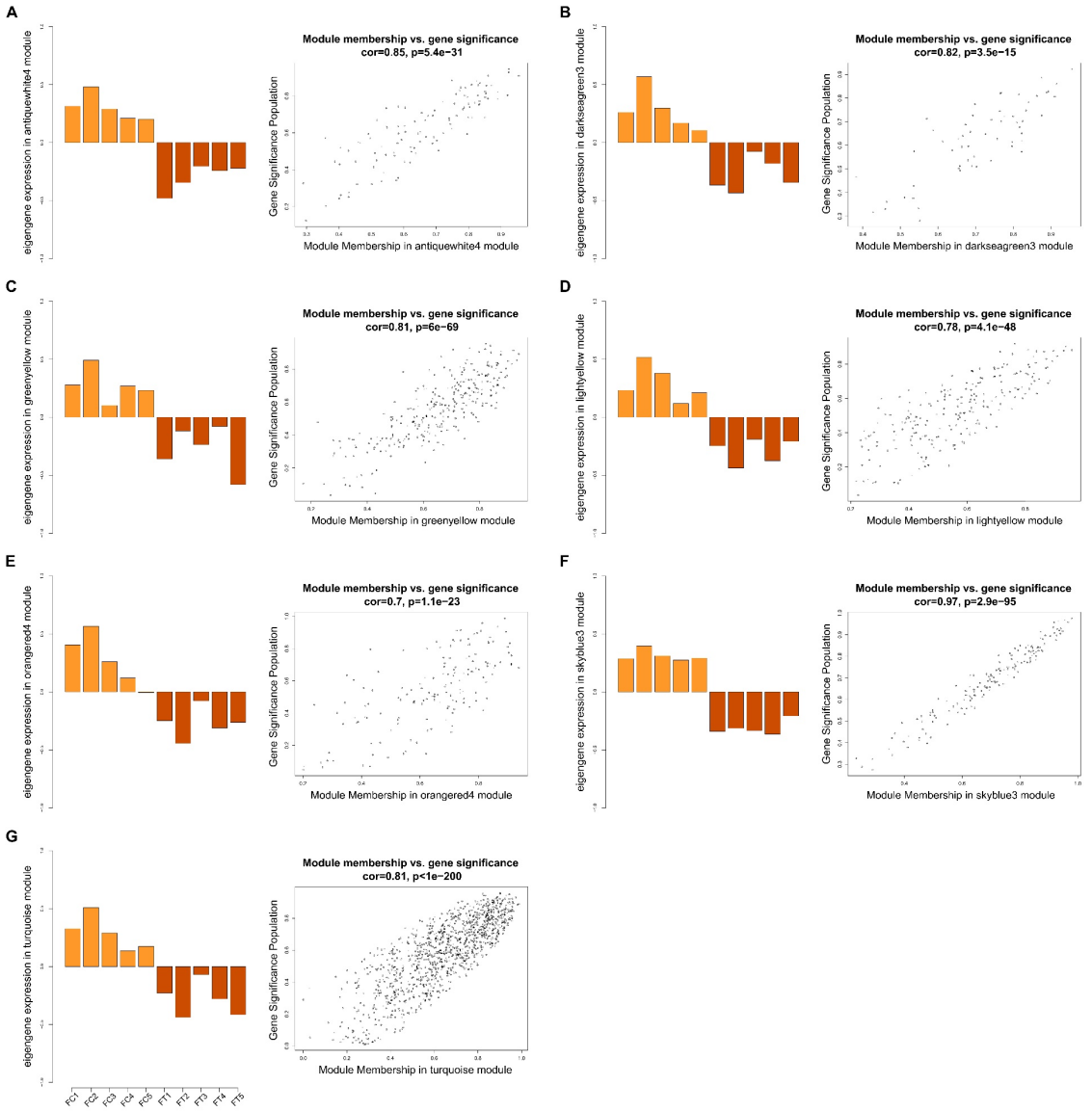


Fig. S15 – Plots for female gonad modules with positive correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light orange for females (FC) from Culatra and dark orange for females (FT) from Trieste.

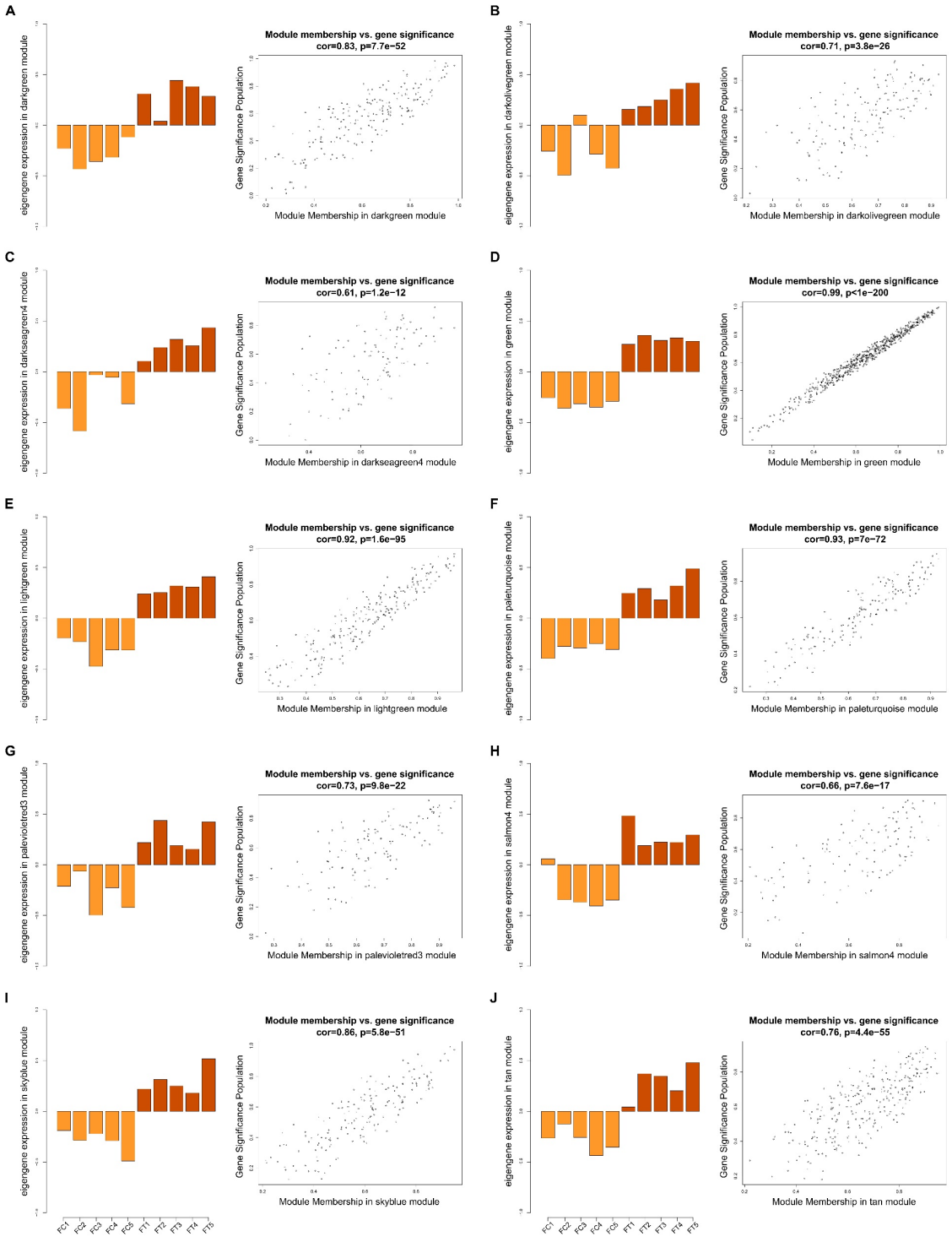


Fig. S16 – Plots for female gonad modules with negative correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light orange for females (FC) from Culatra and dark orange for females (FT) from Trieste.

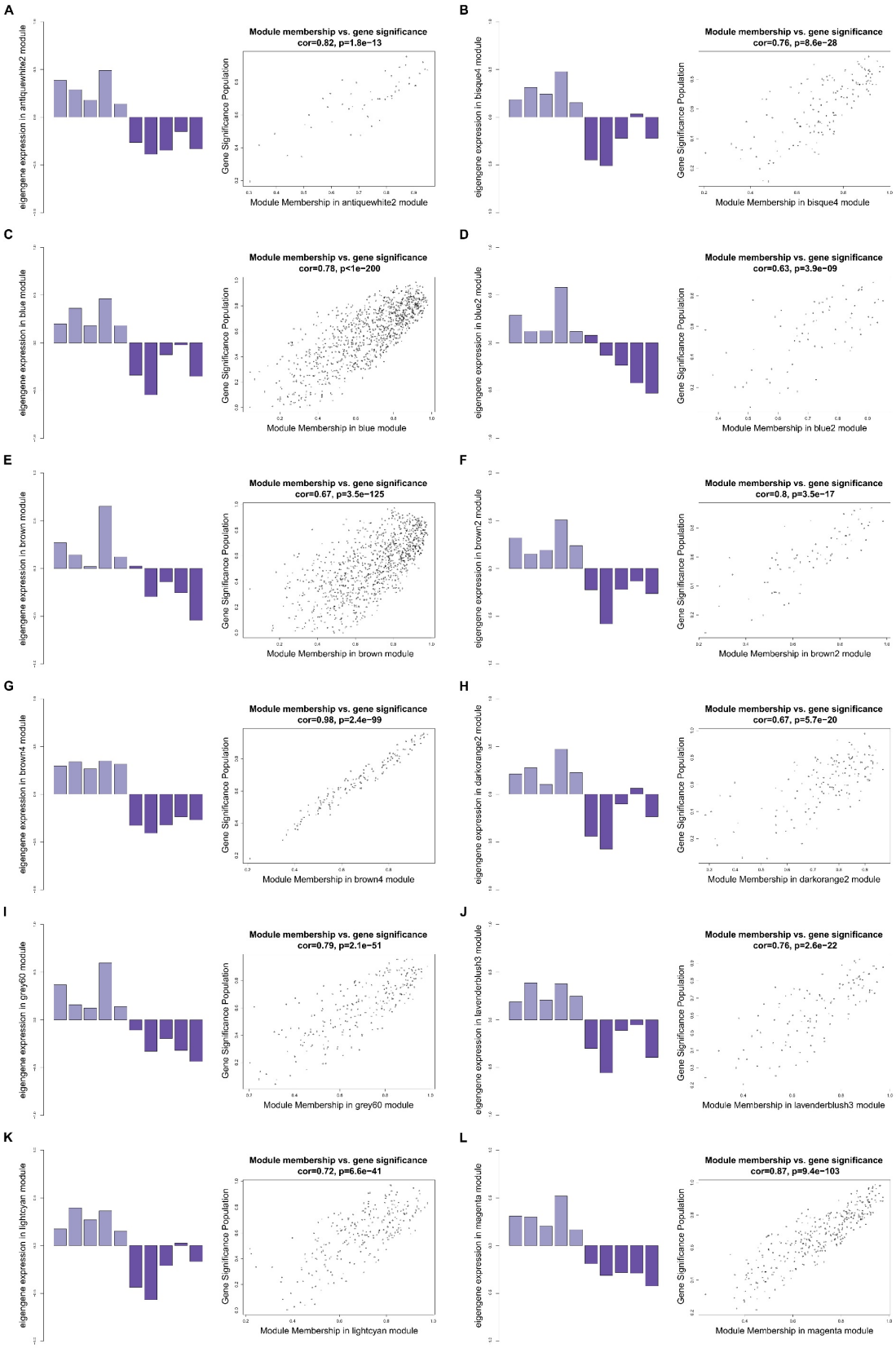


Fig. S17 – Plots for male gonad modules with positive correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light purple for nest-holder males (MC) from Culatra and dark purple for nest-holder males (MT) from Trieste.

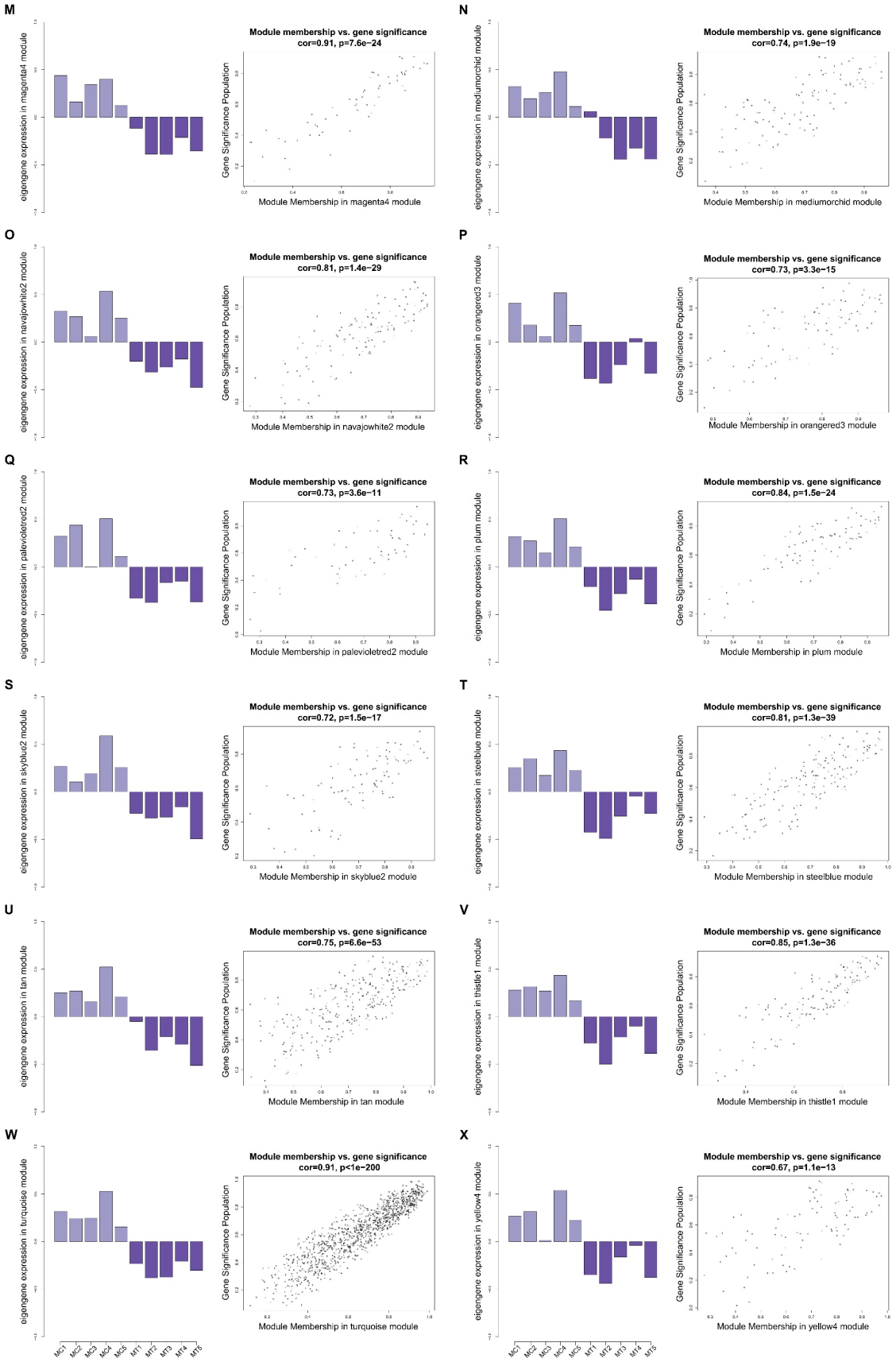


Fig. S17 (*Continuation*)

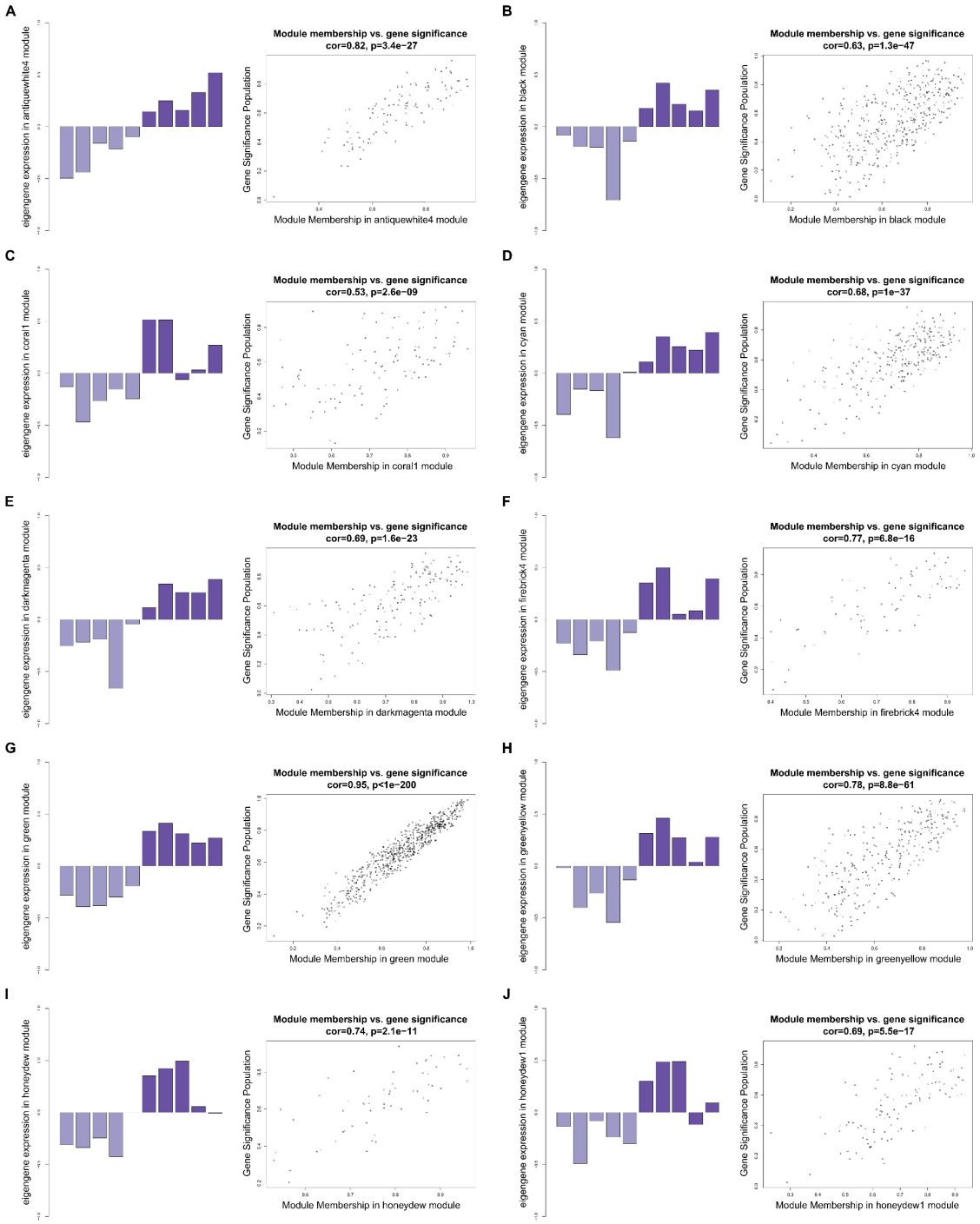


Fig. S18 – Plots for male gonad modules with negative correlation with 'Population' trait, representing expression values of the module eigengene for each sequenced sample (left), and correlation between gene significance for 'Population' and module membership (right), with respective correlations and *P*-values. Bars are colour coded in agreement with light purple for nest-holder males (MC) from Culatra and dark purple for nest-holder males (MT) from Trieste. *(Continued)*

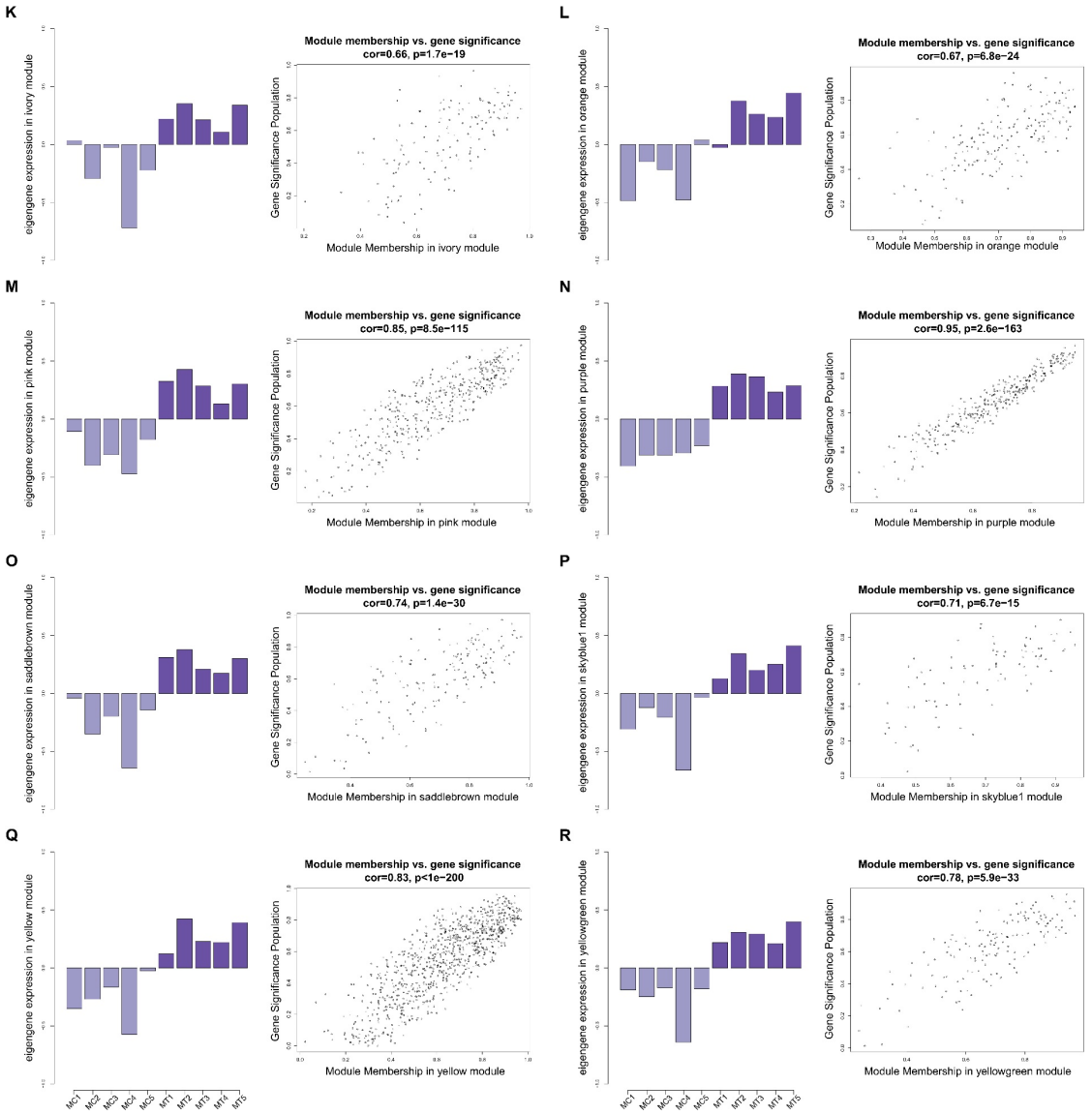


Fig. S18 (*Continuation*)

V.2. Supplementary Tables

Table S1 – RNA quality scores (RIN) and raw and post-trimming read counts per sample, as well the percentage of mapped reads to the transcriptome.

Sample ID	Sample description	Sample RIN	Raw paired-reads	Trimmed paired-reads	% Mappable paired-reads
FC1	Female, Culatra population, replicate 1	B = 8.8	B = 13,368,955	B = 12,953,555	B = 71.85%
		G = 7.8	G = 13,190,346	G = 12,863,799	G = 68.18%
FC2	Female, Culatra population, replicate 2	B = 9.0	B = 12,954,274	B = 12,482,609	B = 69.44%
		G = 7.6	G = 13,542,451	G = 13,104,816	G = 73.43%
FC3	Female, Culatra population, replicate 3	B = 8.9	B = 12,267,503	B = 11,937,482	B = 69.39%
		G = 8.0	G = 12,699,411	G = 12,451,179	G = 71.48%
FC4	Female, Culatra population, replicate 4	B = 8.6	B = 12,786,994	B = 12,585,571	B = 77.00%
		G = 8.5	G = 15,082,702	G = 14,861,399	G = 76.56%
FC5	Female, Culatra population, replicate 5	B = 8.3	B = 13,679,396	B = 13,316,073	B = 73.01%
		G = 8.9	G = 13,405,445	G = 13,048,822	G = 68.69%
FT1	Female, Trieste population, replicate 1	B = 8.4	B = 13,215,145	B = 12,849,865	B = 69.97%
		G = 8.5	G = 13,096,728	G = 12,798,858	G = 70.42%
FT2	Female, Trieste population, replicate 2	B = 8.6	B = 13,156,135	B = 12,748,946	B = 68.98%
		G = 8.7	G = 13,569,840	G = 13,341,647	G = 76.00%
FT3	Female, Trieste population, replicate 3	B = 8.9	B = 11,735,357	B = 11,445,150	B = 72.34%
		G = 7.7	G = 11,544,031	G = 11,184,543	G = 67.11%
FT4	Female, Trieste population, replicate 4	B = 8.4	B = 13,672,193	B = 13,235,110	B = 66.50%
		G = 9.2	G = 13,448,691	G = 13,150,786	G = 68.95%
FT5	Female, Trieste population, replicate 5	B = 8.9	B = 11,234,644	B = 10,923,930	B = 67.32%
		G = 8.3	G = 14,440,784	G = 14,221,466	G = 75.95%

(Continued)

Table S1 (*Continuation*)

Sample ID	Sample description	Sample RIN	Raw paired-reads	Trimmed paired-reads	% Mappable paired-reads
MC1	Nest-holder male, Culatra population, replicate 1	B = 8.6 G = 7.9	B = 14,136,477 G = 12,767,062	B = 13,757,411 G = 12,462,120	B = 71.99% G = 70.41%
MC2	Nest-holder male, Culatra population, replicate 2	B = 8.0 G = 8.5	B = 13,113,292 G = 12,836,792	B = 12,685,505 G = 12,580,939	B = 71.34% G = 67.92%
MC3	Nest-holder male, Culatra population, replicate 3	B = 8.2 G = 8.5	B = 13,422,499 G = 13,173,857	B = 13,051,391 G = 12,879,320	B = 70.21% G = 70.46%
MC4	Nest-holder male, Culatra population, replicate 4	B = 8.3 G = 7.2	B = 11,893,879 G = 10,805,495	B = 11,517,304 G = 10,522,383	B = 69.42% G = 69.69%
MC5	Nest-holder male, Culatra population, replicate 5	B = 8.9 G = 8.8	B = 12,091,000 G = 16,165,396	B = 11,673,413 G = 15,874,031	B = 72.10% G = 73.74%
MT1	Nest-holder male, Trieste population, replicate 1	B = 8.9 G = 8.8	B = 13,315,450 G = 14,708,722	B = 12,974,401 G = 14,367,661	B = 70.95% G = 67.30%
MT2	Nest-holder male, Trieste population, replicate 2	B = 8.8 G = 9.1	B = 09,938,908 G = 15,079,758	B = 09,714,644 G = 14,794,103	B = 68.29% G = 64.72%
MT3	Nest-holder male, Trieste population, replicate 3	B = 8.0 G = 9.0	B = 15,815,608 G = 14,488,911	B = 15,346,604 G = 14,248,504	B = 67.27% G = 68.26%
MT4	Nest-holder male, Trieste population, replicate 4	B = 8.7 G = 8.9	B = 14,958,795 G = 12,636,692	B = 14,436,292 G = 12,344,651	B = 70.57% G = 71.01%
MT5	Nest-holder male, Trieste population, replicate 5	B = 8.7 G = 8.6	B = 14,672,158 G = 13,092,995	B = 14,145,830 G = 12,811,985	B = 72.13% G = 71.44%

B – Information related to forebrain samples; G – Information related to gonad samples

Table S2 – Complete list of differentially expressed genes in the forebrain between nest-holder males Culatra and females Culatra. Positive log2FC indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN314618_c3_g1_i1	*krtap4-3*	keratin-associated protein 4-3-like	1.00E-08	7.0363	0.0141	–
TRINITY_DN332944_c1_g13_i1		MHC class II protein	2.50E-60	6.4401	0.0157	MEred
TRINITY_DN323630_c2_g4_i1	*epb41l2*	Erythrocyte membrane protein band 4.1-like 2	3.60E-10	-6.2341	0.0192	–
TRINITY_DN301266_c0_g1_i6	*ankmy2*	Ankyrin repeat and MYND domain-containing protein 2	1.80E-154	5.5572	0.0083	–
TRINITY_DN309710_c0_g1_i7	*gyc88e*	soluble guanylate cyclase 88E-like2	3.50E-180	5.1930	0.0033	–
TRINITY_DN316049_c0_g1_i5	*gyc88e*	soluble guanylate cyclase 88E-like2	0	5.1201	0.0056	–
TRINITY_DN321378_c0_g1_i1	–	–	–	3.4096	0.0131	–
TRINITY_DN314037_c0_g1_i1	–	–	–	3.3440	0.0199	MEred
TRINITY_DN321715_c3_g1_i1	*ucn3*	Urocortin 3	3.40E-49	2.9853	0.0033	–
TRINITY_DN314371_c3_g1_i1	*znf2*	zinc finger protein 2-like	7.20E-21	2.1736	0.0231	–
TRINITY_DN316326_c0_g4_i7	*mlkl*	Mixed lineage kinase domain like pseudokinase	0	2.0588	0.0192	MEpink
TRINITY_DN330150_c0_g1_i2	*iapp*	islet amyloid polypeptide	1.00E-46	2.0368	0.0033	–
TRINITY_DN296586_c0_g1_i3	*gal*	Galanin/GMAP prepropeptide	3.50E-73	1.8308	3.86E-05	–
TRINITY_DN290277_c0_g1_i1	*vip*	Vasoactive intestinal peptide	3.90E-104	-1.8156	7.50E-05	–
TRINITY_DN304482_c0_g2_i4	*rab20*	RAB20, member RAS oncogene family	3.10E-150	1.5882	0.0033	MEpink
TRINITY_DN310381_c1_g2_i6	*haao*	3-hydroxyanthranilate 3,4-dioxygenase	9.00E-140	1.5176	0.0192	MEpink
TRINITY_DN304658_c0_g1_i1	–	–	–	1.4997	0.0157	–
TRINITY_DN327047_c0_g1_i3	*ncf2*	Neutrophil cytosol factor 2	2.60E-143	1.4450	0.0131	MEpink
TRINITY_DN315194_c1_g2_i1	*ms4a8*	membrane-spanning 4-domains subfamily A member 8	1.70E-18	1.1372	0.0097	MEpink
TRINITY_DN303488_c0_g1_i1	*nxph1*	neurexophilin 1	3.80E-79	-1.1109	0.0272	–
TRINITY_DN328036_c5_g7_i1	*tsn15*	Tetraspanin 15	5.90E-51	1.0677	0.0175	MEpink
TRINITY_DN335277_c5_g3_i6	*vwa5a*	von Willebrand factor A domain-containing protein 5A	0	1.0348	0.0056	MEpink
TRINITY_DN325818_c1_g1_i3	*cmtm6*	CKLF-like MARVEL transmembrane domain containing 6	1.40E-49	1.0324	0.0131	MEpink

Table S3 – Complete list of differentially expressed genes in the forebrain between nest-holder males Trieste and females Trieste. Positive log2FC indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN255860_c0_g2_i1	*mt-cyb*	cytochrome b, mitochondrial	0	5.4046	0.0182	–
TRINITY_DN296586_c0_g1_i3	*gal*	Galanin/GMAP prepropeptide	3.50E-73	2.5856	7.63E-13	–
TRINITY_DN330150_c0_g1_i2	*iapp*	islet amyloid polypeptide	1.00E-46	2.1784	0.0007	–
TRINITY_DN290277_c0_g1_i1	*vip*	Vasoactive intestinal peptide	3.90E-104	-1.7946	3.91E-05	–
TRINITY_DN331546_c1_g2_i2	*piezo1*	Piezo-type mechanosensitive ion channel component 1	0	-1.3723	0.0045	–

Table S4 – List of the top 10% differentially expressed genes in the forebrain between nest-holder males Culatra and nest-holder males Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for nest-holder males Trieste, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males Culatra.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN230527_c0_g1_i1	*eef1as*	Elongation factor 1-alpha, somatic form	0	11.4372	0.0008	–
TRINITY_DN292337_c0_g1_i1	*pabpc1l*	poly(A) binding protein, cytoplasmic 1-like	2.80E-170	10.6952	0.0040	–
TRINITY_DN318227_c0_g1_i6	*ccar1*	cell division cycle and apoptosis regulator 11	8.20E-105	10.1557	0.0143	–
TRINITY_DN303573_c0_g1_i1	*hsp90a.1*	Heat shock protein HSP 90-alpha 1	0	10.0721	0.0173	–
TRINITY_DN300189_c0_g1_i3	*hnrnpa1a*	heterogeneous nuclear ribonucleoprotein A1a	9.40E-147	9.8516	0.0089	–
TRINITY_DN314591_c0_g1_i4	*ncl*	nucleolin-like	3.20E-13	9.7894	0.0065	–
TRINITY_DN10770_c0_g1_i1	*ppifb*	peptidylprolyl isomerase Fb	7.10E-85	9.5522	0.0089	–
TRINITY_DN293439_c0_g1_i2	*ybx1*	Y box-binding protein 1	3.50E-26	9.4950	0.0088	–
TRINITY_DN301962_c0_g1_i2	*hnrnpk*	Heterogeneous nuclear ribonucleoprotein K	5.60E-31	9.3111	0.0212	–
TRINITY_DN304713_c0_g1_i1	*rpl3*	ribosomal protein L3	0	9.2223	0.0116	–
TRINITY_DN249663_c0_g1_i1	*cnbpb*	CCHC-type zinc finger, nucleic acid binding protein b	3.00E-20	9.1203	0.0033	–
TRINITY_DN161064_c0_g1_i1	*rpl10*	60S ribosomal protein L10	1.60E-119	9.1061	0.0278	–
TRINITY_DN271560_c0_g1_i1	*rnasel3*	ribonuclease like 3	4.40E-61	9.0751	9.50E-13	MEpink
TRINITY_DN284873_c0_g1_i2	*hmgb2*	High mobility group protein B2	9.50E-38	9.0042	0.0204	–

TRINITY_DN259495_c0_g1_i1	*gngnb2l1*	Guanine nucleotide-binding protein subunit beta-2-like 1	2.90E-159	8.9395	0.0071	–
TRINITY_DN328764_c0_g1_i1	*hsp90aa2p*	Heat shock protein HSP 90-alpha A2	1.10E-118	8.8911	0.0116	–
TRINITY_DN318420_c4_g3_i2	*cga*	Glycoprotein hormones alpha chain	2.20E-44	8.7317	3.56E-05	MEblack
TRINITY_DN238446_c0_g1_i1	*rpl7*	ribosomal protein L7	1.20E-90	8.6734	0.0336	–
TRINITY_DN257785_c0_g1_i1	*rpl6*	ribosomal protein L6	1.70E-51	8.6221	0.0312	–
TRINITY_DN263618_c0_g1_i1	*rps8b*	ribosomal protein S8b	8.10E-90	8.5900	0.0131	–
TRINITY_DN288777_c0_g1_i1	*rpl15*	ribosomal protein L15	2.40E-103	8.5759	0.0283	–
TRINITY_DN290877_c0_g1_i1	*rplp2*	ribosomal protein, large P2	8.70E-21	8.5427	0.0343	–
TRINITY_DN322216_c0_g1_i3	*ube2e2*	ubiquitin-conjugating enzyme E2E 2	0	8.5016	0.0018	MEmagenta
TRINITY_DN159924_c0_g1_i1	*rps3*	ribosomal protein S3	2.90E-106	8.4095	0.0152	–
TRINITY_DN229021_c0_g1_i1	*rps6*	ribosomal protein S6	4.40E-98	8.3941	0.0380	–
TRINITY_DN226784_c0_g1_i1	*rps18*	Ribosomal protein S18	1.10E-88	8.2081	0.0342	–
TRINITY_DN299540_c0_g1_i1	*csta1*	cystatin-A1-like	1.10E-23	8.1459	6.56E-08	MEpink
TRINITY_DN13493_c0_g1_i1	*rpl9*	Ribosomal protein L9	3.60E-69	8.0954	0.0385	–
TRINITY_DN131712_c0_g1_i1	*rpl12*	ribosomal protein L12	1.20E-91	7.9969	0.0406	–
TRINITY_DN334938_c0_g3_i4	*daam2*	dishevelled associated activator of morphogenesis 2	0	-7.9937	1.49E-10	MEred
TRINITY_DN174726_c0_g1_i1	*ran*	RAN, member RAS oncogene family	3.80E-138	7.9683	0.0379	–
TRINITY_DN508590_c0_g1_i1		hypothetical protein EH28_01512	6.80E-12	7.9323	0.0025	MEpink
TRINITY_DN242038_c0_g1_i1	*rpl27*	ribosomal protein L27	4.60E-57	7.8353	0.0441	–
TRINITY_DN208861_c0_g1_i1	*rps17*	ribosomal protein S17	5.70E-54	7.7812	0.0259	–
TRINITY_DN308766_c0_g2_i3	*prmt1*	protein arginine methyltransferase 1	1.20E-177	7.6085	0.0456	–
TRINITY_DN330398_c3_g8_i2	*ccr3*	C-C chemokine receptor type 3	7.50E-148	7.6030	4.55E-06	MEpink
TRINITY_DN301561_c0_g2_i1	*ctsl.1*	Cathepsin L	3.00E-99	7.5860	9.69E-05	MEpink
TRINITY_DN311005_c0_g1_i1		si:dkey-30j10.5	2.50E-64	7.2111	1.47E-05	MEpink
TRINITY_DN244340_c0_g2_i1	–	–	–	7.1373	1.44E-05	MEpink
TRINITY_DN318729_c9_g4_i6	*wt1-a*	Wilms tumor protein homolog A	1.40E-111	6.9762	7.70E-05	MEpink
TRINITY_DN328562_c0_g4_i1	–	–	–	6.4884	0.0058	MEblack
TRINITY_DN326646_c1_g1_i5	*cc2*	C-C chemokine 2	5.90E-14	6.4787	2.32E-10	MEpink
TRINITY_DN333712_c3_g2_i1	–	–	–	-6.4716	8.37E-08	MEred
TRINITY_DN328023_c0_g1_i1	–	–	–	6.4232	1.75E-07	MEpink
TRINITY_DN311596_c0_g2_i1	*txn*	Thioredoxin	5.00E-36	6.3402	0.0016	MEpink
TRINITY_DN302970_c0_g4_i1	–	–	–	-6.3049	0.0313	MEyellow
TRINITY_DN324431_c3_g5_i1	*rbm41*	RNA binding motif protein 41	6.60E-43	-6.2230	4.39E-06	MEred
TRINITY_DN239614_c0_g4_i1	–	–	–	-6.1856	0.0499	MEred
TRINITY_DN323940_c0_g1_i6	*adarb1*	Double-stranded RNA-specific editase 1	9.80E-153	-6.1482	0.0060	MEred
TRINITY_DN318904_c3_g2_i2	*mmp19*	Matrix metalloproteinase-19	2.00E-174	6.0904	0.0005	MEpink
TRINITY_DN322000_c1_g2_i19	*wt1*	Wilms tumor protein homolog	2.50E-109	6.0100	0.0012	MEpink
TRINITY_DN319886_c1_g1_i2	*gdf5*	growth/differentiation factor 5	4.10E-30	5.9881	0.0003	–
TRINITY_DN335465_c11_g3_i5	*otx1*	orthodenticle homeobox 1	1.20E-34	-5.9162	0.0179	MEred
TRINITY_DN316621_c1_g2_i1	*ppfia3*	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 3	2.60E-68	5.8483	4.55E-06	MEblack
TRINITY_DN301588_c0_g2_i1	–	–	–	-5.7506	2.60E-05	MEred
TRINITY_DN296266_c0_g2_i1	*ctsl.1*	cathepsin L.1	2.80E-121	5.7408	0.0007	MEpink
TRINITY_DN299791_c1_g2_i1	–	–	–	5.6409	0.0035	MEblack
TRINITY_DN238330_c0_g1_i1	–	–	–	-5.6338	0.0047	MEred
TRINITY_DN323630_c2_g4_i1	*epb41l2*	Erythrocyte membrane protein band 4.1-like 2	3.60E-10	-5.6165	0.0041	–
TRINITY_DN334279_c6_g7_i1	*nrxn1a*	neurexin 1a	1.90E-23	-5.3870	0.0009	MEred
TRINITY_DN334281_c4_g16_i1	*ift74*	intraflagellar transport 74	1.20E-42	5.3475	0.0002	MEmagenta
TRINITY_DN323102_c1_g1_i9	*pomca*	proopiomelanocortin a	4.40E-88	5.1837	0.0050	MEblack
TRINITY_DN295262_c0_g10_i1	*ctsa*	cathepsin A	3.20E-17	5.1119	0.0021	MEmagenta
TRINITY_DN309281_c1_g2_i1	–	–	–	-5.0808	0.0003	MEred
TRINITY_DN59465_c0_g2_i1	*ccdc136b*	coiled-coil domain containing 136b	8.30E-40	-5.0165	0.0014	MElightMEyellow
TRINITY_DN198692_c0_g1_i1	–	–	–	-4.9942	0.0004	MEyellow
TRINITY_DN300251_c0_g1_i2	*tmem238a*	transmembrane protein 238a	1.90E-54	4.9875	1.75E-07	MEmagenta
TRINITY_DN301515_c1_g1_i1	–	–	–	4.9040	0.0010	MEpink
TRINITY_DN334971_c3_g10_i3		F-type lectin 3	1.00E-144	4.8677	6.31E-10	MEpink
TRINITY_DN304699_c0_g9_i1	–	–	–	4.7684	0.0004	MEpink
TRINITY_DN332630_c2_g15_i1	–	–	–	4.5725	0.0009	–
TRINITY_DN323589_c2_g3_i1	*pfkpb*	phosphofructokinase, platelet b	1.80E-54	4.4702	0.0059	–
TRINITY_DN322354_c1_g5_i1	–	–	–	-4.4323	0.0022	MEred
TRINITY_DN382982_c0_g1_i1	–	–	–	-4.3737	0.0074	MEred
TRINITY_DN317046_c0_g1_i1	*lifra*	leukemia inhibitory factor receptor alpha a	0	4.2409	4.46E-07	MEpink
TRINITY_DN327638_c0_g3_i1		Fucolectin-4	6.50E-135	4.2295	2.67E-08	MEmagenta
TRINITY_DN293353_c0_g1_i1		hypothetical protein EH28_12542	8.90E-67	-4.1643	0.0046	MEred
TRINITY_DN249374_c0_g1_i2	–	–	–	4.1615	4.99E-05	MEpink
TRINITY_DN330830_c1_g1_i1	*cd209e*	CD209 antigen-like protein E	1.80E-67	4.1535	6.37E-12	MEpink

Table S5 – List of the top 10% differentially expressed genes in the forebrain between females Culatra and females Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for females Trieste, whereas a negative log2FC indicates an up-regulation of the transcript for females Culatra.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN325949_c2_g1_i1	–	–	–	8.9233	4.20E-07	MEblack
TRINITY_DN332944_c1_g13_i1		MHC class II protein	2.50E-60	-8.1249	2.57E-05	MEred
TRINITY_DN314618_c3_g1_i1	*krtap4-3*	keratin-associated protein 4-3-like	1.00E-08	-8.1206	0.0002	–
TRINITY_DN313907_c0_g5_i1	*znf830*	Zinc finger protein 830	1.30E-32	8.1143	0.0006	–
TRINITY_DN318420_c4_g3_i2	*cga*	Glycoprotein hormones alpha chain	2.20E-44	7.9507	0.0005	MEblack
TRINITY_DN333712_c3_g2_i1	–	–	–	-7.5155	8.61E-10	MEred
TRINITY_DN328562_c0_g4_i1	–	–	–	6.8406	0.0075	MEblack
TRINITY_DN323102_c1_g1_i9	*pomca*	proopiomelanocortin a	4.40E-88	6.7352	0.0004	MEblack
TRINITY_DN318173_c0_g2_i2	*gucy1b2*	guanylate cyclase 1, soluble, beta 2	0	-6.7225	0.0006	–
TRINITY_DN323464_c6_g3_i1	*apoda.1*	apolipoprotein Da, duplicate 1	3.50E-112	-6.7064	0.0308	MElightMEyellow
TRINITY_DN301266_c0_g1_i6	*ankmy2*	Ankyrin repeat and MYND domain-containing protein 2	1.80E-154	-6.6793	1.85E-05	–
TRINITY_DN332981_c2_g1_i2	*pde9a*	phosphodiesterase 9A	1.60E-119	-6.6149	0.0005	–
TRINITY_DN294912_c1_g2_i4		parvalbumin, thymic-like	3.90E-66	-6.4731	0.0004	–
TRINITY_DN317937_c1_g1_i1	–	–	–	-6.1633	0.0124	MEred
TRINITY_DN324431_c3_g5_i1	*rbm41*	RNA binding motif protein 41	6.60E-43	-6.0393	2.80E-05	MEred
TRINITY_DN316049_c0_g1_i5	*gyc88e*	soluble guanylate cyclase 88E-like2	0	-5.9813	1.91E-05	–
TRINITY_DN274814_c0_g1_i2	*tspan13b*	tetraspanin 13b	1.50E-21	5.9288	2.38E-05	–
TRINITY_DN309710_c0_g1_i7	*gyc88e*	soluble guanylate cyclase 88E-like2	3.50E-180	-5.8320	1.78E-05	–
TRINITY_DN238330_c0_g1_i1	–	–	–	-5.7271	0.0091	MEred
TRINITY_DN316800_c0_g3_i4	*pde9al*	phosphodiesterase 9A like	3.30E-90	-5.6434	0.0002	–
TRINITY_DN226961_c0_g2_i1	*stk35*	Serine/threonine kinase 35	1.60E-38	-5.6176	0.0003	MEred
TRINITY_DN301588_c0_g2_i1	–	–	–	-5.5899	0.0001	MEred
TRINITY_DN316813_c0_g1_i4	*ptp4a3*	protein tyrosine phosphatase type IVA, member 3	4.60E-105	-5.5177	0.0007	–
TRINITY_DN134029_c0_g2_i1	–	–	–	5.5008	8.39E-05	MEblack
TRINITY_DN322659_c2_g1_i5	–	–	–	5.4329	5.14E-05	MEmagenta
TRINITY_DN299857_c0_g3_i1	*znf502*	Zinc finger protein 502	2.30E-110	-5.4101	0.0016	MEred
TRINITY_DN308820_c0_g1_i2	*itpk1*	Inositol-tetrakisphosphate 1-kinase	5.40E-118	-5.4046	0.0039	–
TRINITY_DN524727_c0_g1_i1	*rpl7a*	ribosomal protein L7a	1.40E-76	-5.3306	0.0001	MEred
TRINITY_DN309281_c1_g2_i1	–	–	–	-5.2992	0.0003	MEred
TRINITY_DN324276_c1_g1_i2	*pde9a*	phosphodiesterase 9A	7.80E-67	-5.2416	0.0002	–
TRINITY_DN334279_c6_g7_i1	*nrxn1a*	neurexin 1a	1.90E-23	-5.2058	0.0035	MEred
TRINITY_DN327097_c1_g2_i1	–	–	–	-5.1934	0.0066	MElightMEyellow
TRINITY_DN290881_c0_g1_i3	*ppp2r2d*	protein phosphatase 2, regulatory subunit B, delta	3.90E-62	-5.1797	0.0001	MEred
TRINITY_DN293353_c0_g1_i1		hypothetical protein EH28_12542	8.90E-67	-5.1469	0.0016	MEred
TRINITY_DN330398_c3_g8_i2	*ccr3*	C-C chemokine receptor type 3	7.50E-148	5.0131	0.0080	MEpink
TRINITY_DN322354_c1_g5_i1	–	–	–	-4.9820	0.0024	MEred
TRINITY_DN320394_c0_g1_i1	*rab19*	RAB19, member RAS oncogene family	5.00E-51	-4.9459	0.0007	MEred

Table S6 – Complete list of differentially expressed genes in the forebrain between nest-holder males and females. Positive log2FoldChange indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN296586_c0_g1_i3	*gal*	Galanin/GMAP prepropeptide	3.50E-73	2.1332	2.42E-12	–
TRINITY_DN315270_c2_g3_i2	*dnajb5*	DnaJ (Hsp40) homolog, subfamily B, member 5	1.10E-96	-1.1608	0.0163	–

Table S7 – List of the top 10% differentially expressed genes in the forebrain between individuals of Culatra and individuals of Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for Trieste individuals, whereas a negative log2FC indicates an up-regulation of the transcript for Culatra individuals.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN313751_c0_g1_i1	*gh1*	growth hormone 1	9.30E-108	8.2332	1.06E-05	MEblack
TRINITY_DN318420_c4_g3_i2	*cga*	Glycoprotein hormones alpha chain	2.20E-44	8.1493	5.66E-09	MEblack
TRINITY_DN183828_c0_g1_i1	–	–	–	7.9087	0.0004	MEblack
TRINITY_DN323947_c1_g9_i1	*elmo2*	Engulfment and cell motility 2	2.10E-70	7.2514	0.0019	MEmagenta
TRINITY_DN283843_c0_g1_i2	*smtla*	somatolactin alpha	4.00E-140	7.1973	0.0006	MEblack
TRINITY_DN320965_c0_g3_i1	*actb1*	actin, beta 1	2.20E-93	7.0706	0.0081	–

TRINITY_DN333712_c3_g2_i1	–	–	–	-7.0024	1.60E-18	MEred
TRINITY_DN285652_c0_g1_i1	*lhb*	luteinizing hormone, beta polypeptide	1.30E-64	6.7227	0.0006	MEblack
TRINITY_DN328562_c0_g4_i1	–	–	–	6.6833	7.77E-07	MEblack
TRINITY_DN330398_c3_g8_i2	*ccr3*	C-C chemokine receptor type 3	7.50E-148	6.6051	1.56E-08	MEpink
TRINITY_DN239614_c0_g4_i1	–	–	–	-6.4868	0.0002	MEred
TRINITY_DN334938_c0_g3_i4	*daam2*	dishevelled associated activator of morphogenesis 2	0	-6.4858	3.30E-16	MEred
TRINITY_DN299540_c0_g1_i1	*csta1*	cystatin-A1-like	1.10E-23	6.4016	4.25E-09	MEpink
TRINITY_DN271560_c0_g1_i1	*rnasel3*	ribonuclease like 3	4.40E-61	6.3371	1.23E-10	MEpink
TRINITY_DN275258_c0_g2_i1	*rpl30*	ribosomal protein L30	5.40E-35	6.2494	0.0044	–
TRINITY_DN335465_c11_g3_i5	*otx1*	orthodenticle homeobox 1	1.20E-34	-6.1624	2.07E-05	MEred
TRINITY_DN324431_c3_g5_i1	*rbm41*	RNA binding motif protein 41	6.60E-43	-6.1404	3.85E-12	MEred
TRINITY_DN323464_c6_g3_i1	*apoda.1*	apolipoprotein Da, duplicate 1	3.50E-112	-6.0719	0.0002	MElightMEyellow
TRINITY_DN314037_c0_g2_i1	–	–	–	-5.8056	0.0338	MEred
TRINITY_DN320965_c0_g1_i1	*actb*	Actin, cytoplasmic 1	2.10E-143	5.7876	0.0077	–
TRINITY_DN301561_c0_g2_i1	*ctsl.1*	Cathepsin L	3.00E-99	5.7760	1.67E-05	MEpink
TRINITY_DN323102_c1_g1_i9	*pomca*	proopiomelanocortin a	4.40E-88	5.6980	9.51E-07	MEblack
TRINITY_DN301588_c0_g2_i1	–	–	–	-5.6793	1.18E-10	MEred
TRINITY_DN238330_c0_g1_i1	–	–	–	-5.6782	3.88E-09	MEred
TRINITY_DN313907_c0_g5_i1	*znf830*	Zinc finger protein 830	1.30E-32	5.5945	0.0003	–
TRINITY_DN311005_c0_g1_i1	*si:dkey-30j10.5*	si:dkey-30j10.5	2.50E-64	5.5438	1.18E-06	MEpink
TRINITY_DN115187_c0_g1_i1	*rps27.2*	ribosomal protein S27, isoform 2	3.60E-30	5.5040	0.0294	–
TRINITY_DN318904_c3_g2_i2	*mmp19*	Matrix metalloproteinase-19	2.00E-174	5.4986	1.55E-06	MEpink
TRINITY_DN296266_c0_g2_i1	*ctsl.1*	cathepsin L.1	2.80E-121	5.4135	1.78E-05	MEpink
TRINITY_DN322000_c1_g2_i19	*wt1*	Wilms tumor protein homolog	2.50E-109	5.3596	1.43E-05	MEpink
TRINITY_DN323940_c0_g1_i6	*adarb1*	Double-stranded RNA-specific editase 1	9.80E-153	-5.3292	5.19E-05	MEred
TRINITY_DN334279_c6_g7_i1	*nrxn1a*	neurexin 1a	1.90E-23	-5.3056	2.63E-07	MEred
TRINITY_DN244340_c0_g2_i1	–	–	–	5.2287	5.31E-06	MEpink
TRINITY_DN324431_c3_g8_i1	*rbm41*	RNA binding motif protein 41	6.60E-43	-5.2065	0.0008	MEred
TRINITY_DN299791_c1_g2_i1	–	–	–	5.2001	8.81E-06	MEblack
TRINITY_DN309281_c1_g2_i1	–	–	–	-5.1912	3.63E-09	MEred
TRINITY_DN242478_c0_g1_i1		hypothetical protein	7.70E-13	5.0856	0.0196	–
TRINITY_DN328023_c0_g1_i1	–	–	–	5.0241	9.96E-11	MEpink
TRINITY_DN368561_c0_g2_i1	*mef2d*	myocyte enhancer factor 2d	2.40E-36	4.9807	0.0451	–
TRINITY_DN334485_c3_g7_i1	–	–	–	-4.9261	4.17E-05	MEred
TRINITY_DN323277_c0_g1_i1	*rps25*	ribosomal protein S25	1.50E-32	4.8930	0.0226	–
TRINITY_DN299857_c0_g3_i1	*znf502*	Zinc finger protein 502	2.30E-110	-4.8768	3.37E-06	MEred
TRINITY_DN174726_c0_g1_i1	*ran*	RAN, member RAS oncogene family	3.80E-138	4.8676	0.0463	–
TRINITY_DN322273_c0_g2_i1	*stbd1*	starch binding domain 1	6.20E-19	-4.8531	0.0113	MEred
TRINITY_DN295262_c0_g10_i1	*ctsa*	cathepsin A	3.20E-17	4.8487	1.01E-05	MEmagenta
TRINITY_DN327097_c1_g2_i1	–	–	–	-4.8456	1.31E-05	MElightMEyellow
TRINITY_DN300251_c0_g1_i2	*tmem238a*	transmembrane protein 238a	1.90E-54	4.8052	2.04E-12	MEmagenta
TRINITY_DN328764_c0_g1_i1	*hsp90aa2p*	Heat shock protein HSP 90-alpha A2	1.10E-118	4.7643	0.0482	–
TRINITY_DN322354_c1_g5_i1	–	–	–	-4.6960	1.04E-06	MEred
TRINITY_DN293353_c0_g1_i1		hypothetical protein EH28_12542	8.90E-67	-4.6716	1.91E-06	MEred
TRINITY_DN382982_c0_g1_i1	–	–	–	-4.6288	5.19E-06	MEred
TRINITY_DN508590_c0_g1_i1		hypothetical protein EH28_01512	6.80E-12	4.5327	0.0149	MEpink
TRINITY_DN318729_c9_g4_i6	*wt1-a*	Wilms tumor protein homolog A	1.40E-111	4.5102	0.0001	MEpink
TRINITY_DN335461_c0_g3_i1	–	–	–	-4.4205	0.0001	MEgreenMEyellow
TRINITY_DN274814_c0_g1_i2	*tspan13b*	tetraspanin 13b	1.50E-21	4.4044	6.30E-07	–
TRINITY_DN134029_c0_g2_i1	–	–	–	4.3839	5.36E-07	MEblack
TRINITY_DN290881_c0_g1_i3	*ppp2r2d*	protein phosphatase 2, regulatory subunit B, delta	3.90E-62	-4.3741	1.58E-08	MEred
TRINITY_DN325949_c2_g2_i1	–	–	–	-4.3562	0.0022	MEred
TRINITY_DN325949_c2_g1_i1	–	–	–	4.2584	0.0014	MEblack
TRINITY_DN326646_c1_g1_i5	*cc2*	C-C chemokine 2	5.90E-14	4.2500	1.04E-07	MEpink
TRINITY_DN306313_c1_g2_i1	–	–	–	-4.2153	1.50E-06	MEred
TRINITY_DN301515_c1_g1_i1	–	–	–	4.1931	1.60E-05	MEpink
TRINITY_DN311596_c0_g2_i1	*txn*	Thioredoxin	5.00E-36	4.1337	0.0013	MEpink
TRINITY_DN226961_c0_g2_i1	*stk35*	Serine/threonine kinase 35	1.60E-38	-4.1046	2.87E-05	MEred
TRINITY_DN317546_c0_g1_i1	–	–	–	-4.1021	9.05E-06	MEred
TRINITY_DN249374_c0_g1_i2	–	–	–	4.0861	4.70E-07	MEpink
TRINITY_DN322659_c2_g1_i5	–	–	–	4.0833	2.19E-11	MEmagenta
TRINITY_DN332630_c2_g15_i1	–	–	–	4.0724	3.62E-05	–
TRINITY_DN314037_c0_g5_i1	–	–	–	4.0363	0.0009	MEmagenta
TRINITY_DN299846_c3_g2_i1	*tmem200b*	transmembrane protein 200B	2.00E-33	-3.8309	0.0008	MEred
TRINITY_DN335374_c0_g1_i1		Uncharacterized protein	7.30E-30	-3.7858	0.0035	–
TRINITY_DN334971_c3_g10_i3		F-type lectin 3	1.00E-144	3.7550	7.16E-11	MEpink
TRINITY_DN329526_c1_g1_i7		PHD-finger domain-containing protein	1.90E-75	-3.6969	2.05E-09	MEred
TRINITY_DN304699_c0_g9_i1	–	–	–	3.6489	1.37E-05	MEpink
TRINITY_DN317046_c0_g1_i1	*lifra*	leukemia inhibitory factor receptor alpha a	0	3.6108	1.47E-08	MEpink

Table S8 – List of the top 1% differentially expressed genes in the gonads between nest-holder males Culatra and females Culatra ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN322657_c1_g1_i9	*morn3*	MORN repeat-containing protein 3	8.80E-114	14.3136	7.37E-50	MEyellow4
TRINITY_DN311595_c0_g1_i1	*lrrc51*	leucine-rich repeat-containing 51-like	1.60E-70	13.9560	7.44E-48	MEyellow4
TRINITY_DN324669_c5_g2_i4	*rbm46*	RNA binding motif protein 46	7.10E-119	13.9135	8.81E-50	MEtan
TRINITY_DN305096_c0_g1_i2	*pifo*	Primary Cilia Formation	1.30E-72	13.6610	1.32E-46	MEmagenta
TRINITY_DN326350_c0_g3_i1	–	–	–	-13.6267	4.87E-46	–
TRINITY_DN319386_c1_g4_i9	*muc5ac*	Mucin-5AC	4.30E-74	-13.5688	6.39E-48	MEskyblue
TRINITY_DN318255_c2_g6_i1	*lrrc18a*	Leucine rich repeat containing 18a	3.80E-88	13.3329	7.88E-43	MEyellow4
TRINITY_DN322719_c0_g1_i2	*tex36*	testis expressed 36	3.90E-43	13.2130	2.89E-40	MEnavajowhite2
TRINITY_DN324025_c9_g1_i11	*cfap77*	cilia- and flagella-associated protein 77	2.90E-76	13.1946	8.75E-67	–
TRINITY_DN335134_c1_g1_i2	*plcz1*	Phospholipase C Zeta 1	4.60E-117	13.1449	4.79E-39	–
TRINITY_DN330388_c2_g13_i4	*tcf7l1b*	transcription factor 7-like 1-B	4.50E-25	13.1033	2.39E-41	–
TRINITY_DN330476_c0_g3_i1	*ttll9*	tubulin tyrosine ligase-like family, member 9	0	13.0624	7.26E-43	MEskyblue2
TRINITY_DN320180_c0_g2_i9	*lef1*	lymphoid enhancer-binding factor 1-like	4.30E-07	13.0298	9.60E-49	MEpalevioletred3
TRINITY_DN329742_c2_g1_i2	–	–	–	12.9778	3.07E-38	–
TRINITY_DN327204_c2_g1_i8	*styxl1*	Serine/threonine/tyrosine interacting-like 1	3.40E-49	12.9328	2.14E-36	MEyellow4
TRINITY_DN312005_c7_g3_i1	–	–	–	12.8550	2.56E-39	–
TRINITY_DN330232_c0_g4_i2	*ccdc37*	coiled-coil domain-containing 37-like	2.70E-07	12.8402	3.73E-36	–
TRINITY_DN322741_c1_g1_i2	*plcz1*	Phospholipase C Zeta 1	1.80E-28	12.8143	1.21E-91	MEbisque4
TRINITY_DN315851_c0_g1_i16	*cep55*	Centrosomal protein 55	1.60E-98	12.8044	1.57E-48	MEbrown4
TRINITY_DN334452_c6_g1_i27	*cfap65*	Cilia- and flagella-associated protein 65	0	12.7618	3.71E-40	MEantiquewhite2
TRINITY_DN322408_c1_g2_i1	*rsph1*	radial spoke head 1 homolog	1.90E-116	12.7256	6.97E-48	MEyellow4
TRINITY_DN332695_c1_g4_i5	–	–	–	12.6961	1.77E-37	–
TRINITY_DN325165_c0_g2_i1	*faah*	Fatty-acid amide hydrolase 1	4.90E-164	12.6565	2.75E-36	–
TRINITY_DN290990_c0_g2_i2	–	–	–	-12.6466	1.89E-47	MEbrown
TRINITY_DN321492_c6_g1_i7	*spaca4l*	sperm acrosome associated 4 like	3.00E-57	-12.5683	4.68E-30	–
TRINITY_DN323199_c3_g1_i3	*drc7*	Coiled-coil domain containing 135	2.20E-87	12.5038	2.44E-39	–
TRINITY_DN333753_c6_g2_i2	*insl3*	insulin-like 3 (Leydig cell)	1.40E-41	12.4955	5.33E-121	MEblue
TRINITY_DN312099_c4_g6_i1	*znf146*	zinc finger protein OZF	4.80E-06	12.4734	1.05E-36	–
TRINITY_DN334848_c6_g8_i4	*lrrc18a*	leucine rich repeat containing 18a	3.70E-68	12.4492	1.32E-43	–
TRINITY_DN313860_c3_g3_i2	–	–	–	12.4254	8.86E-34	–
TRINITY_DN320436_c0_g1_i2		hypothetical protein EH28_03918	3.00E-12	-12.3728	8.35E-38	–
TRINITY_DN314116_c0_g2_i6	*izumo1*	izumo sperm-egg fusion protein 1	2.80E-154	12.3718	7.51E-38	MEyellow4
TRINITY_DN312610_c1_g1_i3	*dnal1*	Dynein light chain 1, axonemal	1.40E-48	12.3633	2.68E-35	–
TRINITY_DN330053_c0_g2_i13	*lrrc71*	leucine-rich repeat-containing 713	1.90E-99	12.3372	9.03E-38	MEyellow4
TRINITY_DN327701_c1_g1_i18	*ccdc40*	Coiled-coil domain containing 40	2.40E-165	12.3362	3.70E-46	–
TRINITY_DN313235_c0_g1_i1	–	–	–	12.3334	4.00E-36	MEpalevioletred3
TRINITY_DN331876_c1_g12_i3	*brdt*	Bromodomain testis-specific protein	2.00E-58	12.3166	2.27E-37	–
TRINITY_DN315665_c3_g6_i2	*plcd1b*	phospholipase C, delta 1b	2.10E-93	12.3149	1.55E-140	MEturquoise
TRINITY_DN271377_c0_g2_i2	–	–	–	12.3102	1.69E-35	–
TRINITY_DN326174_c0_g2_i6	*pih1d2*	PIH1 domain containing 2	2.80E-92	12.3097	3.90E-35	MEblue
TRINITY_DN323669_c0_g1_i2	*btg4*	B-cell translocation gene 4	8.50E-105	-12.3052	5.09E-78	–
TRINITY_DN330887_c1_g2_i2	*tsnaxip1*	Translin-associated factor X-interacting protein 1	1.60E-75	12.2903	1.12E-36	MEtan
TRINITY_DN295757_c0_g2_i1	–	–	–	12.2825	7.92E-35	–
TRINITY_DN264543_c0_g1_i1	*cldnd*	Claudin	4.00E-55	-12.2734	1.25E-38	–
TRINITY_DN318954_c2_g1_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	1.20E-142	12.2672	5.34E-141	–
TRINITY_DN318324_c1_g3_i2	*ribc1*	RIB43A domain with coiled-coils 1	4.50E-129	12.2630	1.33E-44	MEturquoise
TRINITY_DN310543_c1_g1_i3	*wfdc2*	WAP four-disulfide core domain 2	1.10E-22	-12.2402	6.52E-43	MEmediumorchid
TRINITY_DN321616_c6_g2_i2	*clu*	Clusterin	1.30E-66	12.2390	1.59E-216	–
TRINITY_DN321979_c0_g1_i1	*h2b*	histone H2B-like	5.60E-21	-12.2285	5.06E-69	MEbrown
TRINITY_DN330535_c1_g1_i2		si:dkey-27p23.3	1.90E-27	12.2231	1.92E-36	MEnavajowhite2
TRINITY_DN319154_c0_g1_i1	*h2af1al*	H2A histone family member 1a like	2.20E-40	-12.2046	1.20E-68	–
TRINITY_DN317478_c1_g1_i4	*cfap161*	cilia- and flagella-associated protein 161	4.20E-144	12.1982	8.98E-88	MEpalevioletred3
TRINITY_DN312279_c0_g1_i3	*ankef1*	Ankyrin repeat and EF-hand domain-containing protein 1	1.90E-172	12.1432	5.20E-34	–
TRINITY_DN331020_c0_g1_i3	*zp2l1*	zona pellucida glycoprotein 2, like 1	7.70E-171	-12.1000	2.45E-61	–
TRINITY_DN320566_c0_g1_i3	–	–	–	12.0980	2.99E-34	–
TRINITY_DN300289_c0_g1_i2	*–4*	–	–	-12.0866	2.99E-43	MEbrown
TRINITY_DN332654_c0_g2_i8	*zp2*	Zona pellucida sperm-binding protein 2	0	-12.0722	1.95E-169	–
TRINITY_DN335169_c3_g1_i7	*sycp1*	Synaptonemal complex protein 1	9.80E-163	12.0386	0	MEtan
TRINITY_DN330536_c0_g7_i1	–	–	–	12.0342	7.13E-32	–
TRINITY_DN328841_c0_g1_i2	–	–	–	-12.0224	2.86E-147	MEbrown
TRINITY_DN334894_c0_g5_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	4.80E-110	12.0070	5.08E-36	–
TRINITY_DN326743_c4_g7_i2	–	–	–	11.9990	1.38E-32	–
TRINITY_DN284263_c0_g1_i2	*ctsz*	cathepsin Z	2.00E-61	-11.9903	4.13E-36	–
TRINITY_DN331894_c1_g9_i1	–	–	–	-11.9801	6.94E-45	MEgrey60

TRINITY_DN323411_c0_g3_i3		calcium-binding protein P-like	4.50E-17	-11.9773	1.07E-32	–
TRINITY_DN331836_c2_g4_i3		waprin-Phi1-like	4.20E-20	-11.9581	5.60E-47	MEmediumorchid
TRINITY_DN331672_c0_g1_i6	*prss27*	protease, serine, 27	4.20E-91	-11.9308	4.13E-34	–
TRINITY_DN318617_c4_g3_i3		calcium-binding protein P-like	4.10E-18	-11.9291	4.91E-72	MEbrown
TRINITY_DN334343_c1_g6_i2	–	–	–	11.9165	3.94E-32	–
TRINITY_DN324581_c0_g3_i3	–	–	–	-11.9000	1.84E-24	–
TRINITY_DN332496_c0_g3_i5		si:dkey-57n24.6	0	11.8917	7.77E-35	MEthistle1
TRINITY_DN334409_c1_g4_i6	*cfap52*	Cilia and flagella associated protein 52	5.90E-36	11.8570	2.38E-34	–
TRINITY_DN313583_c1_g11_i1	*ccdc70*	coiled-coil domain-containing 70-like	2.10E-07	11.8550	6.18E-30	–
TRINITY_DN305057_c0_g1_i1	*krt18*	Keratin, type I cytoskeletal 18	0	-11.8423	1.36E-36	MEblue
TRINITY_DN305336_c0_g1_i1	*spag16*	Sperm-associated antigen 16 protein	3.20E-103	11.8312	4.01E-35	–
TRINITY_DN330473_c0_g1_i1	*org*	oogenesis-related gene	1.10E-28	-11.8256	4.14E-185	–
TRINITY_DN333153_c4_g2_i1		sperm-associated antigen 8	2.50E-26	11.8076	4.56E-34	MEyellow4
TRINITY_DN331293_c0_g5_i11	*hipk1a*	Homeodomain-interacting protein kinase 1	1.90E-71	11.8035	2.16E-32	–
TRINITY_DN330703_c7_g3_i7	*larp4b*	La ribonucleoprotein domain family, member 4B	0	11.8026	2.14E-34	MEyellow4
TRINITY_DN317561_c0_g2_i3	*zp3a.1*	zona pellucida glycoprotein 3a, tandem duplicate 1	1.70E-127	-11.7998	1.94E-98	–
TRINITY_DN242493_c0_g1_i2		calcium-binding protein P-like	1.90E-14	-11.7934	2.10E-84	MEmediumorchid
TRINITY_DN273068_c0_g1_i1	–	–	–	-11.7731	2.62E-86	–
TRINITY_DN320318_c1_g1_i1	*ctsz*	Cathepsin Z	7.40E-115	-11.7537	3.19E-66	–
TRINITY_DN310439_c0_g1_i1	*mme*	Metalloendopeptidase	2.30E-93	-11.7419	1.98E-38	MEorangered4
TRINITY_DN313189_c1_g1_i5	*gle1*	GLE1 RNA export mediator homolog (yeast)	5.10E-20	11.7228	6.72E-32	–
TRINITY_DN329214_c4_g1_i1	–	–	–	11.7182	1.17E-29	–
TRINITY_DN312005_c7_g1_i2	–	–	–	11.7128	1.07E-39	MEturquoise
TRINITY_DN331185_c0_g1_i5	–	–	–	-11.6462	2.38E-34	MEmediumorchid
TRINITY_DN332295_c2_g1_i1		Si:dkey-90a13.10	1.30E-50	11.6401	2.07E-39	–
TRINITY_DN331894_c1_g1_i7	*ttn*	titin	1.10E-14	-11.6274	2.47E-124	MEgrey60
TRINITY_DN328768_c0_g1_i1		P17/29C-like protein DDB_G0287399	1.80E-09	-11.6259	7.21E-132	–
TRINITY_DN331659_c0_g2_i1	–	–	–	-11.6185	7.86E-58	–
TRINITY_DN333462_c6_g3_i1	*h1m*	Linker histone H1M	4.50E-52	-11.5966	9.44E-141	MEtan
TRINITY_DN323206_c0_g7_i1	–	–	–	11.5966	6.33E-29	MEmagenta
TRINITY_DN323179_c3_g3_i8	*wap*	whey acidic protein-like	3.40E-12	-11.5943	5.32E-34	MEmediumorchid
TRINITY_DN303744_c0_g1_i1	*syce2*	synaptonemal complex central element protein 2-like	9.40E-51	11.5724	8.07E-51	MEbrown
TRINITY_DN326248_c1_g9_i1	–	–	–	11.5672	1.15E-29	–
TRINITY_DN323657_c0_g1_i3	*sparcl1*	SPARC-like 1	2.80E-113	-11.5593	9.18E-124	–
TRINITY_DN303996_c0_g1_i4	–	–	–	-11.5503	1.47E-36	MEbrown
TRINITY_DN318617_c4_g2_i4		calcium-binding protein P-like	8.60E-15	-11.5476	2.47E-44	MEmediumorchid
TRINITY_DN257796_c0_g1_i4	–	–	–	-11.5472	1.04E-41	MEblue
TRINITY_DN327029_c0_g3_i4		extensin-like	1.10E-13	-11.5434	1.71E-33	–
TRINITY_DN328307_c2_g2_i2	*krt9*	keratin, type I cytoskeletal 9-like	1.10E-07	-11.5345	6.40E-39	MEmediumorchid
TRINITY_DN318255_c2_g3_i1	*lrrc18a*	Leucine rich repeat containing 18a	4.70E-42	11.5277	5.01E-32	MEyellow4
TRINITY_DN325682_c0_g1_i2	*lrp2*	Low-density lipoprotein receptor-related protein 2	9.10E-135	-11.5158	1.33E-195	MEbrown
TRINITY_DN333199_c5_g6_i1	–	–	–	11.5106	1.93E-31	–
TRINITY_DN325908_c0_g1_i2		Hypothetical protein EH28_18302	2.00E-07	11.4939	6.43E-31	–
TRINITY_DN316170_c0_g1_i1	*nexn*	nexilin-like	7.70E-17	11.4876	3.77E-22	–
TRINITY_DN310726_c0_g1_i5		Estrogen-regulated protein	1.10E-79	-11.4866	1.92E-12	–
TRINITY_DN373777_c0_g1_i1	–	–	–	-11.4797	2.84E-33	MEblue
TRINITY_DN314371_c4_g3_i1	–	–	–	11.4791	1.64E-32	MEantiquewhite2
TRINITY_DN335638_c2_g1_i1	*dnhd1*	Dynein heavy chain domain-containing protein 1	2.90E-109	11.4772	3.06E-50	–
TRINITY_DN307318_c0_g3_i4	*odf3b*	outer dense fiber of sperm tails 3b	4.90E-109	11.4673	1.94E-37	–
TRINITY_DN305805_c0_g1_i1	*dnaaf4*	Dynein axonemal assembly factor 4	4.30E-70	11.4622	2.76E-31	MEbrown2
TRINITY_DN305854_c0_g1_i1	*hes2.1*	hes family bHLH transcription factor 2, tandem duplicate 1	3.40E-31	-11.4504	6.88E-16	–
TRINITY_DN324538_c2_g3_i2	*tbx16*	T-box 16	5.70E-116	-11.4481	1.84E-32	MEskyblue
TRINITY_DN318617_c3_g1_i5	*anxa1*	annexin A1	1.10E-155	-11.4361	1.67E-157	–
TRINITY_DN331792_c2_g1_i6	*fam187a*	Ig-like V-type domain-containing protein FAM187A	0	11.4248	1.79E-32	MEyellow4
TRINITY_DN317063_c0_g2_i7	–	–	–	11.4197	1.35E-29	–
TRINITY_DN326095_c4_g1_i1		gastrula zinc finger protein xFG20-1-like	8.60E-21	11.4034	3.87E-31	–
TRINITY_DN322187_c1_g5_i1	*cfap45*	cilia- and flagella-associated protein 45	8.40E-06	11.3971	7.22E-31	–
TRINITY_DN326185_c0_g2_i3	*ctss*	Cathepsin S	3.90E-124	-11.3915	2.45E-51	–
TRINITY_DN320683_c0_g2_i2	*gch1*	GTP cyclohydrolase 1	2.70E-102	11.3810	5.65E-12	MEorange
TRINITY_DN325805_c0_g2_i9	*drc3*	Dynein regulatory complex subunit 3	2.20E-164	11.3805	2.20E-77	MEpalevioletred2
TRINITY_DN326350_c0_g2_i9	–	–	–	-11.3689	1.91E-45	–
TRINITY_DN335379_c3_g3_i1	*cluap1*	Clusterin-associated protein 1 homolog	1.80E-110	11.3659	3.43E-30	–
TRINITY_DN329009_c2_g3_i11	*ktn1*	kinectin-like	6.90E-49	11.3545	8.23E-33	MEantiquewhite2

Table S9 – List of the top 1% differentially expressed genes in the gonads between nest-holder males Trieste and females Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN290955_c0_g1_i1	*actb*	Actin, cytoplasmic 1	0	-22.1081	1.92E-13	MEturquoise
TRINITY_DN331894_c1_g1_i7	*ttn*	titin	1.10E-14	-14.9200	1.60E-55	MEgrey60
TRINITY_DN328768_c0_g1_i1		P17/29C-like protein DDB_G0287399	1.80E-09	-14.5602	5.69E-52	–
TRINITY_DN318617_c4_g3_i3		calcium-binding protein P-like	4.10E-18	-14.3386	4.83E-49	MEbrown
TRINITY_DN314727_c0_g1_i1	*zp3b*	zona pellucida glycoprotein 3b	0	-14.2302	7.92E-104	–
TRINITY_DN273068_c0_g1_i1	–	–	–	-14.1331	1.26E-47	–
TRINITY_DN319386_c1_g5_i1		dentin sialophosphoprotein-like	1.40E-58	-13.9669	2.50E-47	MEbrown
TRINITY_DN320924_c0_g12_i2	–	–	–	-13.9654	5.63E-46	MElavenderblush3
TRINITY_DN319386_c1_g4_i9	*muc5ac*	Mucin-5AC	4.30E-74	-13.8596	1.37E-48	MEskyblue
TRINITY_DN320924_c0_g7_i3	–	–	–	-13.8434	1.01E-47	MEskyblue
TRINITY_DN300289_c0_g1_i2	–	–	–	-13.8290	2.13E-46	MEbrown
TRINITY_DN317561_c0_g2_i3	*zp3a.1*	zona pellucida glycoprotein 3a, tandem duplicate 1	1.70E-127	-13.7596	2.75E-47	–
TRINITY_DN242493_c0_g1_i2		calcium-binding protein P-like	1.90E-14	-13.7442	1.93E-52	MEmediumorchid
TRINITY_DN311544_c0_g1_i2	*cldnd*	claudin d	1.90E-52	-13.6895	1.17E-46	–
TRINITY_DN331185_c0_g1_i5	–	–	–	-13.4058	3.29E-43	MEmediumorchid
TRINITY_DN324249_c0_g1_i13	*atg16l1*	Autophagy-related protein 16-1	5.00E-37	-13.3312	5.62E-42	MEskyblue
TRINITY_DN321620_c0_g1_i2	*zp3a.2*	zona pellucida glycoprotein 3a, tandem duplicate 2	0	-13.3231	1.67E-44	–
TRINITY_DN257796_c0_g1_i4	–	–	–	-13.3135	6.31E-45	MEblue
TRINITY_DN256186_c0_g1_i1	–	–	–	-13.2998	2.32E-41	MEmediumorchid
TRINITY_DN328841_c0_g1_i2	–	–	–	-13.2743	1.44E-102	MEbrown
TRINITY_DN323179_c3_g2_i3	*wfdc18*	WAP four-disulfide core domain protein 18-like	3.70E-11	-13.2098	7.28E-61	MEmediumorchid
TRINITY_DN330630_c1_g1_i1	–	–	–	-13.1885	1.44E-42	MEbrown
TRINITY_DN264543_c0_g1_i1	*cldnd*	Claudin	4.00E-55	-13.1744	3.85E-43	–
TRINITY_DN334506_c0_g2_i4	*zp3*	Zona pellucida sperm-binding protein 3	2.20E-13	-13.1719	1.64E-49	–
TRINITY_DN323179_c3_g3_i8	*wap*	whey acidic protein-like	3.40E-12	-13.1612	5.36E-43	MEmediumorchid
TRINITY_DN321616_c6_g2_i2	*clu*	Clusterin	1.30E-66	13.1477	4.44E-253	–
TRINITY_DN326185_c0_g2_i3	*ctss*	Cathepsin S	3.90E-124	-13.1245	3.82E-61	–
TRINITY_DN317005_c0_g2_i8		uncharacterized protein LOC108875206	5.00E-82	-13.0274	3.21E-96	–
TRINITY_DN329742_c2_g1_i2	–	–	–	13.0166	2.22E-38	–
TRINITY_DN320318_c1_g1_i1	*ctsz*	Cathepsin Z	7.40E-115	-12.9943	1.10E-42	–
TRINITY_DN328307_c2_g2_i2	*krt9*	keratin, type I cytoskeletal 9-like	1.10E-07	-12.9838	5.76E-47	MEmediumorchid
TRINITY_DN332677_c0_g1_i7	*g2e3*	G2/M phase-specific E3 ubiquitin-protein ligase	0	-12.9760	7.16E-43	–
TRINITY_DN332654_c0_g2_i8	*zp2*	Zona pellucida sperm-binding protein 2	0	-12.9577	2.04E-139	–
TRINITY_DN320438_c0_g3_i10	*iqub*	IQ and ubiquitin-like domain-containing protein	2.60E-152	12.9008	1.35E-42	MEskyblue3
TRINITY_DN320885_c0_g1_i2	–	–	–	-12.8865	3.31E-75	–
TRINITY_DN318324_c1_g3_i2	*ribc1*	RIB43A domain with coiled-coils 1	4.50E-129	12.8564	2.24E-41	MEturquoise
TRINITY_DN295757_c0_g2_i1	–	–	–	12.8537	6.02E-38	–
TRINITY_DN318617_c4_g2_i4		calcium-binding protein P-like	8.60E-15	-12.8500	1.35E-53	MEmediumorchid
TRINITY_DN327701_c1_g1_i18	*ccdc40*	Coiled-coil domain containing 40	2.40E-165	12.8417	5.82E-42	–
TRINITY_DN326743_c4_g7_i2	–	–	–	12.8323	9.65E-37	–
TRINITY_DN333462_c6_g3_i1	*h1m*	Linker histone H1M	4.50E-52	-12.7703	5.12E-123	MEtan
TRINITY_DN310543_c1_g1_i3	*wfdc2*	WAP four-disulfide core domain 2	1.10E-22	-12.7554	1.74E-40	MEmediumorchid
TRINITY_DN326505_c1_g1_i1	*zpc1*	Zona pellucida protein C1	2.70E-136	-12.7523	1.22E-69	MEbrown
TRINITY_DN318546_c1_g1_i3	*slc47a1*	solute carrier family 47 (multidrug and toxin extrusion), member 1	0	12.7186	5.49E-17	MEpurple
TRINITY_DN331836_c2_g4_i3		waprin-Phi1-like	4.20E-20	-12.6803	2.51E-51	MEmediumorchid
TRINITY_DN335100_c2_g2_i7	*zp3.2*	zona pellucida glycoprotein 3, tandem duplicate 2	0	-12.6715	2.28E-70	–
TRINITY_DN334894_c0_g5_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	4.80E-110	12.6581	8.15E-40	–
TRINITY_DN330679_c0_g2_i3	*yeats2*	YEATS domain-containing protein 2	5.00E-07	-12.6538	3.96E-57	–
TRINITY_DN330232_c0_g4_i2	*ccdc37*	coiled-coil domain-containing 37-like	2.70E-07	12.6502	4.69E-35	–
TRINITY_DN305336_c0_g1_i1	*spag16*	Sperm-associated antigen 16 protein	3.20E-103	12.5809	3.79E-39	–
TRINITY_DN323330_c2_g2_i3	–	–	–	-12.5673	1.48E-71	MEmediumorchid
TRINITY_DN322187_c1_g5_i1	*cfap45*	cilia- and flagella-associated protein 45	8.40E-06	12.5644	3.52E-37	–
TRINITY_DN331020_c0_g1_i3	*zp2l1*	zona pellucida glycoprotein 2, like 1	7.70E-171	-12.5509	5.65E-61	–
TRINITY_DN303996_c0_g1_i4	–	–	–	-12.5362	4.40E-36	MEbrown
TRINITY_DN334452_c6_g1_i27	*cfap65*	Cilia- and flagella-associated protein 65	0	12.5207	1.30E-38	MEantiquewhite2
TRINITY_DN327661_c2_g1_i1	*stx11a*	syntaxin 11a	6.90E-152	-12.5178	1.28E-39	MEbrown
TRINITY_DN321979_c0_g1_i1	*h2b*	histone H2B-like	5.60E-21	-12.5156	1.24E-71	MEbrown
TRINITY_DN325682_c0_g1_i2	*lrp2*	Low-density lipoprotein receptor-related protein 2	9.10E-135	-12.5077	3.67E-156	MEbrown
TRINITY_DN326350_c0_g3_i1	–	–	–	-12.5048	4.23E-59	–
TRINITY_DN316170_c0_g1_i1	*nexn*	nexilin-like	7.70E-17	12.4997	6.54E-26	–
TRINITY_DN314775_c0_g1_i3		uncharacterized protein LOC104919597	4.40E-50	-12.4807	2.92E-37	–
TRINITY_DN331480_c1_g1_i1	–	–	–	-12.4604	3.09E-37	MEsteelblue
TRINITY_DN322001_c5_g1_i3	*g0s2*	G0/G1 switch regulatory protein 2	1.40E-31	12.4402	2.03E-254	MEblack

TRINITY_DN329159_c3_g2_i2	*wag22*	WAG22 antigen-like	1.30E-06	-12.4353	1.77E-80	MEmediumorchid
TRINITY_DN331480_c1_g9_i1	–	–	–	-12.4218	3.78E-77	MEblue
TRINITY_DN312279_c0_g1_i3	*ankef1*	Ankyrin repeat and EF-hand domain-containing protein 1	1.90E-172	12.3935	2.65E-35	–
TRINITY_DN318617_c3_g1_i5	*anxa1*	annexin A1	1.10E-155	-12.3906	1.83E-138	–
TRINITY_DN319154_c0_g1_i1	*h2af1al*	H2A histone family member 1a like	2.20E-40	-12.3535	1.90E-81	–
TRINITY_DN305057_c0_g1_i1	*krt18*	Keratin, type I cytoskeletal 18	0	-12.3170	2.00E-38	MEblue
TRINITY_DN322132_c1_g1_i1	*f11*	coagulation factor XI-like	5.40E-159	-12.3074	4.93E-09	MEturquoise
TRINITY_DN318954_c2_g1_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	1.20E-142	12.2315	3.02E-162	–
TRINITY_DN321492_c6_g1_i7	*spaca4l*	sperm acrosome associated 4 like	3.00E-57	-12.2108	2.30E-28	–
TRINITY_DN290990_c0_g2_i2	–	–	–	-12.2013	4.83E-55	MEbrown
TRINITY_DN322741_c1_g1_i2	*plcz1*	Phospholipase C Zeta 1	1.80E-28	12.1904	2.76E-83	MEbisque4
TRINITY_DN326952_c3_g2_i1	–	–	–	-12.1289	2.24E-69	MEtan
TRINITY_DN312610_c1_g1_i3	*dnal1*	Dynein light chain 1, axonemal	1.40E-48	12.1009	8.54E-34	–
TRINITY_DN315665_c3_g6_i2	*plcd1b*	phospholipase C, delta 1b	2.10E-93	12.0904	1.64E-116	MEturquoise
TRINITY_DN329214_c4_g1_i1	–	–	–	12.0893	2.05E-31	–
TRINITY_DN301575_c0_g2_i2	–	–	–	-12.0756	3.95E-84	–
TRINITY_DN333018_c2_g1_i5		uncharacterized protein LOC1065282722	5.70E-95	-12.0654	6.28E-53	–
TRINITY_DN271377_c0_g2_i2	–	–	–	12.0644	4.55E-34	–
TRINITY_DN330536_c0_g7_i1	–	–	–	12.0618	6.13E-32	–
TRINITY_DN373777_c0_g1_i1	–	–	–	-12.0487	1.90E-36	MEblue
TRINITY_DN331480_c1_g4_i1		dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1-like	6.00E-07	-12.0377	2.49E-34	–
TRINITY_DN333167_c2_g4_i1	*cfap45*	cilia- and flagella-associated protein 45	9.10E-27	12.0278	2.24E-31	–
TRINITY_DN329750_c0_g4_i2	–	–	–	12.0113	9.64E-34	–
TRINITY_DN330473_c0_g1_i1	*org*	oogenesis-related gene	1.10E-28	-11.9835	1.92E-202	–
TRINITY_DN276489_c0_g2_i2	*fabp1b.1*	fatty acid binding protein 1b, tandem duplicate 1	1.20E-74	-11.9575	7.67E-42	–
TRINITY_DN305805_c0_g1_i1	*dnaaf4*	Dynein axonemal assembly factor 4	4.30E-70	11.9562	7.38E-34	MEbrown2
TRINITY_DN310726_c0_g1_i5		Estrogen-regulated protein	1.10E-79	-11.9265	7.07E-14	–
TRINITY_DN326787_c0_g3_i9	*papola*	Poly(A) polymerase alpha	2.80E-146	11.8923	4.13E-34	–
TRINITY_DN313860_c3_g3_i2	–	–	–	11.8726	7.17E-31	–
TRINITY_DN326555_c2_g1_i6	*slco1e1*	solute carrier organic anion transporter family, member 1E1	3.30E-151	11.8613	1.13E-33	MEpurple
TRINITY_DN330367_c1_g1_i10	–	–	–	-11.8507	1.35E-35	MEbisque4
TRINITY_DN324676_c1_g2_i1	*cfap157*	cilia and flagella associated protein 157	8.20E-13	11.8431	1.40E-65	–
TRINITY_DN305096_c0_g1_i2	*pifo*	Primary Cilia Formation	1.30E-72	11.7605	2.62E-53	MEmagenta
TRINITY_DN314795_c0_g1_i2	*cldne*	claudin e	1.40E-103	-11.7425	2.68E-52	MEbrown
TRINITY_DN334694_c0_g6_i7	*dnah6*	Dynein heavy chain 6, axonemal	0	11.7328	3.37E-33	MElightyellow
TRINITY_DN296038_c0_g1_i4	*dnal1*	Dynein light chain 1, axonemal	3.20E-30	11.7129	2.86E-32	–
TRINITY_DN331876_c1_g12_i3	*brdt*	Bromodomain testis-specific protein	2.00E-58	11.6875	1.06E-33	–
TRINITY_DN320283_c2_g3_i5	–	–	–	11.6826	2.51E-33	–
TRINITY_DN325165_c0_g2_i1	*faah*	Fatty-acid amide hydrolase 1	4.90E-164	11.6693	5.86E-31	–
TRINITY_DN332695_c1_g4_i5	–	–	–	11.6479	1.15E-31	–
TRINITY_DN326248_c1_g9_i1	–	–	–	11.6060	8.72E-30	–
TRINITY_DN324538_c2_g3_i2	*tbx16*	T-box 16	5.70E-116	-11.5995	3.16E-33	MEskyblue
TRINITY_DN331894_c1_g9_i1	–	–	–	-11.5969	1.11E-40	MEgrey60
TRINITY_DN284263_c0_g1_i2	*ctsz*	cathepsin Z	2.00E-61	-11.5835	1.63E-38	–
TRINITY_DN291875_c0_g1_i1	–	–	–	11.5746	8.25E-34	MEpurple
TRINITY_DN320180_c0_g2_i9	*lef1*	lymphoid enhancer-binding factor 1-like	4.30E-07	11.5557	4.76E-69	MEpalevioletred3
TRINITY_DN313189_c1_g1_i5	*gle1*	GLE1 RNA export mediator homolog (yeast)	5.10E-20	11.5338	7.65E-31	–
TRINITY_DN330476_c0_g3_i1	*ttll9*	tubulin tyrosine ligase-like family, member 9	0	11.5321	1.11E-39	MEskyblue2
TRINITY_DN312005_c7_g1_i2	–	–	–	11.5257	1.25E-32	MEturquoise
TRINITY_DN310364_c1_g6_i5		hypothetical protein XENTR_v90030530mg	1.40E-06	11.5254	2.50E-30	–
TRINITY_DN326192_c1_g6_i1	–	–	–	11.5055	3.90E-31	–
TRINITY_DN329401_c0_g3_i3	*pcloa*	piccolo presynaptic cytomatrix protein a	4.80E-13	-11.4909	1.53E-90	MEblue
TRINITY_DN334343_c1_g7_i2	–	–	–	11.4906	5.93E-30	–
TRINITY_DN283261_c0_g3_i2	*ccnb1ip1*	E3 ubiquitin-protein ligase CCNB1IP1	1.30E-135	11.4901	1.13E-32	–
TRINITY_DN318247_c0_g1_i1		extensin-like	1.90E-39	-11.4852	2.97E-181	–
TRINITY_DN302684_c0_g1_i1	–	–	–	11.4830	1.48E-30	–
TRINITY_DN312005_c7_g3_i1	–	–	–	11.4673	1.84E-31	–

Table S10 – List of the top 1% differentially expressed genes in the gonads between nest-holder males Culatra and nest-holder males Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for nest-holder males Trieste, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males Culatra.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN323206_c0_g7_i1	–	–	–	11.0376	1.10E-23	MEmagenta
TRINITY_DN323371_c0_g1_i3	*chit1*	Chitotriosidase-1	4.80E-129	-10.6083	2.16E-06	MEsalmon4
TRINITY_DN304165_c0_g1_i2		guanylin-like	1.50E-38	-9.9299	1.95E-05	MEsalmon4
TRINITY_DN327562_c1_g1_i2		Collagen alpha-1(XXI) chain	0	-9.4590	1.42E-05	MEyellowgreen

TRINITY_DN319503_c0_g1_i2	*lect2l*	leukocyte cell-derived chemotaxin 2 like	1.90E-73	-9.2962	7.84E-05	MEsalmon4
TRINITY_DN246676_c0_g1_i1	*fabp6*	Fatty acid binding protein 6, ileal (gastrotropin)	1.60E-83	-8.9907	7.33E-06	MEsalmon4
TRINITY_DN183828_c0_g1_i1	–	–	–	8.7978	0.0216	MEpalevioletred2
TRINITY_DN317294_c0_g1_i1	–	–	–	-8.7533	0.0002	MEsalmon4
TRINITY_DN334281_c4_g16_i1	*ift74*	intraflagellar transport 74	1.20E-42	8.6670	5.32E-09	MEturquoise
TRINITY_DN331257_c4_g1_i2	*sult1st6*	sulfotransferase family 1, cytosolic sulfotransferase 6	6.50E-105	-8.6021	9.21E-07	MEyellow
TRINITY_DN306732_c0_g3_i1	*ctsb*	Cathepsin B	1.20E-161	-8.5966	0.0002	MEsalmon4
TRINITY_DN323255_c0_g1_i2	*pdzk1*	PDZ domain containing 1	3.20E-137	-8.4803	2.14E-06	MEsalmon4
TRINITY_DN330305_c1_g1_i9	*myo15aa*	Myosin XVAa	1.20E-179	-8.4693	1.53E-10	MElightgreen
TRINITY_DN316205_c0_g1_i5	*iyd*	Iodotyrosine deiodinase 1	1.50E-123	-8.4345	1.01E-06	MEivory
TRINITY_DN262496_c0_g1_i1	*apoa4b.1*	apolipoprotein A-IV b, tandem duplicate 1	8.60E-53	-8.0419	0.0001	MEsalmon4
TRINITY_DN298209_c0_g1_i1	*apoa4*	Apolipoprotein A IV	6.00E-136	-8.0003	3.06E-05	MEorange
TRINITY_DN293475_c0_g1_i1	*dupd1*	Dual specificity phosphatase DUPD1	2.80E-126	-7.8963	1.14E-09	MEyellow
TRINITY_DN314051_c2_g1_i1	*krt18*	Keratin, type I cytoskeletal 18	1.00E-134	-7.7786	2.12E-06	MEsalmon4
TRINITY_DN314818_c0_g1_i1	*tmigd1*	transmembrane and immunoglobulin domain containing 1	1.70E-82	-7.7101	0.0015	MEsalmon4
TRINITY_DN264397_c0_g2_i1	–	–	–	7.6626	1.45E-05	MEturquoise
TRINITY_DN296749_c0_g1_i1	*tm4sf5*	transmembrane 4 L six family member 5	1.00E-83	-7.5703	3.14E-05	MEsalmon4
TRINITY_DN315017_c1_g1_i2		uncharacterized protein LOC109962628	4.10E-72	7.5275	3.02E-08	MEmagenta

Table S11 – List of the top 1% differentially expressed genes in the gonads between females Culatra and females Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for females Trieste, whereas a negative log2FC indicates an up-regulation of the transcript for females Culatra.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN266467_c0_g2_i2	*pgm1*	phosphoglucomutase 1	0	9.7147	8.09E-06	MEorangered4
TRINITY_DN319667_c1_g1_i1	*angpt5*	angiopoietin-related protein 5-like	2.10E-66	-8.6419	4.76E-07	MEgreen
TRINITY_DN304134_c0_g1_i1	*npvf*	Neuropeptide VF	1.20E-82	-8.5297	0.0059	MEgreen
TRINITY_DN335461_c0_g3_i1	–	–	–	-8.2642	1.68E-13	MEgreen
TRINITY_DN291202_c0_g2_i2	*camk1d*	calcium/calmodulin-dependent protein kinase type 1D-like	7.10E-15	-8.2539	0.0008	MEskyblue
TRINITY_DN322848_c2_g1_i5	*nptx1*	Neuronal pentraxin-1	0	-8.2372	0.0012	–
TRINITY_DN268068_c0_g1_i1	–	–	–	8.1754	0.0033	MEturquoise
TRINITY_DN332944_c1_g13_i1		MHC class II protein	2.50E-60	-7.9456	0.0002	MEgreen
TRINITY_DN309105_c0_g1_i1	*b4galnt2*	Beta-1,4 N-acetylgalactosaminyltransferase 2	2.00E-101	-7.8363	0.0235	MEgreen

Table S12 – List of the top 1% differentially expressed genes in the gonads between nest-holder males and females ordered by fold-change (FC). Positive log2FoldChange indicates an up-regulation of the transcript for females, whereas a negative log2FC indicates an up-regulation of the transcript for nest-holder males.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN323206_c0_g7_i1	–	–	–	25.4700	4.07E-28	MEmagenta
TRINITY_DN280316_c0_g2_i1	–	–	–	-23.9793	1.60E-15	MEskyblue
TRINITY_DN319386_c1_g4_i9	*muc5ac*	Mucin-5AC	4.30E-74	-13.7160	3.17E-95	MEskyblue
TRINITY_DN329742_c2_g1_i2	–	–	–	12.9973	1.83E-75	–
TRINITY_DN326350_c0_g3_i1	–	–	–	-12.9847	6.89E-87	–
TRINITY_DN300289_c0_g1_i2	–	–	–	-12.9545	8.23E-80	MEbrown
TRINITY_DN318617_c4_g3_i3		calcium-binding protein P-like	4.10E-18	-12.8518	1.01E-105	MEbrown
TRINITY_DN321616_c6_g2_i2	*clu*	Clusterin	1.30E-66	12.7872	0	–
TRINITY_DN330232_c0_g4_i2	*ccdc37*	coiled-coil domain-containing 37-like	2.70E-07	12.7486	8.96E-70	–
TRINITY_DN264543_c0_g1_i1	*cldnd*	Claudin	4.00E-55	-12.7267	1.07E-79	–
TRINITY_DN273068_c0_g1_i1	–	–	–	-12.6566	3.40E-135	–
TRINITY_DN311595_c0_g1_i1	*lrrc51*	leucine-rich repeat-containing 51-like	1.60E-70	12.6528	3.00E-71	MEyellow4
TRINITY_DN334452_c6_g1_i27	*cfap65*	Cilia- and flagella-associated protein 65	0	12.6467	1.43E-77	MEantiquewhite2
TRINITY_DN305096_c0_g1_i2	*pifo*	Primary Cilia Formation	1.30E-72	12.6328	1.51E-81	MEmagenta
TRINITY_DN242493_c0_g1_i2		calcium-binding protein P-like	1.90E-14	-12.6205	2.44E-139	MEmediumorchid
TRINITY_DN327701_c1_g1_i18	*ccdc40*	Coiled-coil domain containing 40	2.40E-165	12.6049	1.44E-77	–
TRINITY_DN295757_c0_g2_i1	–	–	–	12.5681	1.19E-71	–
TRINITY_DN318324_c1_g3_i2	*ribc1*	RIB43A domain with coiled-coils 1	4.50E-129	12.5644	1.01E-76	MEturquoise
TRINITY_DN328768_c0_g1_i1		P17/29C-like protein DDB_G0287399	1.80E-09	-12.5551	1.79E-197	–
TRINITY_DN328841_c0_g1_i2	–	–	–	-12.5451	1.83E-255	MEbrown
TRINITY_DN322741_c1_g1_i2	*plcz1*	Phospholipase C Zeta 1	1.80E-28	12.5356	9.51E-162	MEbisque4
TRINITY_DN331894_c1_g1_i7	*ttn*	titin	1.10E-14	-12.4956	1.04E-153	MEgrey60
TRINITY_DN310543_c1_g1_i3	*wfdc2*	WAP four-disulfide core domain 2	1.10E-22	-12.4948	2.81E-98	MEmediumorchid
TRINITY_DN317561_c0_g2_i3	*zp3a.1*	zona pellucida glycoprotein 3a, tandem duplicate 1	1.70E-127	-12.4771	1.49E-145	–
TRINITY_DN324669_c5_g2_i4	*rbm46*	RNA binding motif protein 46	7.10E-119	12.4726	2.06E-76	MEtan

TRINITY_DN332654_c0_g2_i8	*zp2*	Zona pellucida sperm-binding protein 2	0	-12.4709	0	–
TRINITY_DN326743_c4_g7_i2	–	–	–	12.4255	2.19E-67	–
TRINITY_DN290990_c0_g2_i2	–	–	–	-12.4128	4.59E-101	MEbrown
TRINITY_DN321492_c6_g1_i7	*spaca4l*	sperm acrosome associated 4 like	3.00E-57	-12.4097	2.94E-63	–
TRINITY_DN257796_c0_g1_i4	–	–	–	-12.4035	1.08E-77	MEblue
TRINITY_DN321979_c0_g1_i1	*h2b*	histone H2B-like	5.60E-21	-12.3756	2.27E-138	MEbrown
TRINITY_DN320924_c0_g12_i2	–	–	–	-12.3443	7.21E-75	MElavenderblush3
TRINITY_DN334894_c0_g5_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	4.80E-110	12.3329	5.31E-75	–
TRINITY_DN331836_c2_g4_i3		waprin-Phi1-like	4.20E-20	-12.3132	8.86E-101	MEmediumorchid
TRINITY_DN331020_c0_g1_i3	*zp2l1*	zona pellucida glycoprotein 2, like 1	7.70E-171	-12.3061	6.59E-220	–
TRINITY_DN330476_c0_g3_i1	*ttll9*	tubulin tyrosine ligase-like family, member 9	0	12.2963	2.06E-74	MEskyblue2
TRINITY_DN319154_c0_g1_i1	*h2af1al*	H2A histone family member 1a like	2.20E-40	-12.2819	1.47E-146	–
TRINITY_DN312279_c0_g1_i3	*ankef1*	Ankyrin repeat and EF-hand domain-containing protein 1	1.90E-172	12.2735	5.68E-68	–
TRINITY_DN320924_c0_g7_i3	–	–	–	-12.2691	6.53E-71	MEskyblue
TRINITY_DN318954_c2_g1_i2	*slco4a1*	Solute carrier organic anion transporter family member 4A1	1.20E-142	12.2490	9.17E-298	–
TRINITY_DN320318_c1_g1_i1	*ctsz*	Cathepsin Z	7.40E-115	-12.2475	1.73E-103	–
TRINITY_DN331185_c0_g1_i5	–	–	–	-12.2455	7.75E-89	MEmediumorchid
TRINITY_DN312610_c1_g1_i3	*dnal1*	Dynein light chain 1, axonemal	1.40E-48	12.2384	1.07E-67	–
TRINITY_DN320180_c0_g2_i9	*lef1*	lymphoid enhancer-binding factor 1-like	4.30E-07	12.2289	6.20E-117	MEpalevioletred3
TRINITY_DN305336_c0_g1_i1	*spag16*	Sperm-associated antigen 16 protein	3.20E-103	12.2209	4.11E-72	–
TRINITY_DN312005_c7_g3_i1	–	–	–	12.2186	7.55E-67	–
TRINITY_DN315665_c3_g6_i2	*plcd1b*	phospholipase C, delta 1b	2.10E-93	12.2157	1.23E-234	MEturquoise
TRINITY_DN271377_c0_g2_i2	–	–	–	12.1929	3.00E-68	–
TRINITY_DN332695_c1_g4_i5	–	–	–	12.1918	5.18E-67	–
TRINITY_DN325165_c0_g2_i1	*faah*	Fatty-acid amide hydrolase 1	4.90E-164	12.1780	2.45E-65	–
TRINITY_DN313860_c3_g3_i2	–	–	–	12.1761	2.57E-62	–
TRINITY_DN318255_c2_g6_i1	*lrrc18a*	Leucine rich repeat containing 18a	3.80E-88	12.1653	5.30E-68	MEyellow4
TRINITY_DN314727_c0_g1_i1	*zp3b*	zona pellucida glycoprotein 3b	0	-12.1405	4.81E-80	–
TRINITY_DN333462_c6_g3_i1	*h1m*	Linker histone H1M	4.50E-52	-12.1358	3.31E-253	MEtan
TRINITY_DN326185_c0_g2_i3	*ctss*	Cathepsin S	3.90E-124	-12.1327	4.91E-116	–
TRINITY_DN323179_c3_g3_i8	*wap*	whey acidic protein-like	3.40E-12	-12.1319	2.03E-109	MEmediumorchid
TRINITY_DN318546_c1_g1_i3	*slc47a1*	solute carrier family 47 (multidrug and toxin extrusion), member 1	0	12.1060	1.69E-22	MEpurple
TRINITY_DN321620_c0_g1_i2	*zp3a.2*	zona pellucida glycoprotein 3a, tandem duplicate 2	0	-12.0952	3.13E-88	–
TRINITY_DN305057_c0_g1_i1	*krt18*	Keratin, type I cytoskeletal 18	0	-12.0824	5.36E-73	MEblue
TRINITY_DN330536_c0_g7_i1	–	–	–	12.0481	1.45E-62	–
TRINITY_DN322408_c1_g2_i1	*rsph1*	radial spoke head 1 homolog	1.90E-116	12.0370	2.75E-104	MEyellow4
TRINITY_DN303996_c0_g1_i4	–	–	–	-12.0320	2.45E-68	MEbrown
TRINITY_DN332677_c0_g1_i7	*g2e3*	G2/M phase-specific E3 ubiquitin-protein ligase	0	-12.0158	1.40E-77	–
TRINITY_DN330388_c2_g13_i4	*tcf7l1b*	transcription factor 7-like 1-B	4.50E-25	12.0135	2.77E-70	–
TRINITY_DN322187_c1_g5_i1	*cfap45*	cilia- and flagella-associated protein 45	8.40E-06	12.0101	1.78E-65	–
TRINITY_DN316170_c0_g1_i1	*nexn*	nexilin-like	7.70E-17	12.0093	1.00E-54	–
TRINITY_DN327204_c2_g1_i8	*styxl1*	Serine/threonine/tyrosine interacting-like 1	3.40E-49	12.0026	2.10E-46	MEyellow4
TRINITY_DN331876_c1_g12_i3	*brdt*	Bromodomain testis-specific protein	2.00E-58	12.0024	3.54E-70	–
TRINITY_DN328307_c2_g2_i2	*krt9*	keratin, type I cytoskeletal 9-like	1.10E-07	-11.9945	1.03E-86	MEmediumorchid
TRINITY_DN310726_c0_g1_i5		Estrogen-regulated protein	1.10E-79	-11.9923	5.45E-25	–
TRINITY_DN318617_c4_g2_i4		calcium-binding protein P-like	8.60E-15	-11.9883	4.45E-145	MEmediumorchid
TRINITY_DN322657_c1_g1_i9	*morn3*	MORN repeat-containing protein 3	8.80E-114	11.9845	1.10E-140	MEyellow4
TRINITY_DN333753_c6_g2_i2	*insl3*	insulin-like 3 (Leydig cell)	1.40E-41	11.9802	1.81E-186	MEblue
TRINITY_DN325682_c0_g1_i2	*lrp2*	Low-density lipoprotein receptor-related protein 2	9.10E-135	-11.9666	0	MEbrown
TRINITY_DN335100_c2_g2_i7	*zp3.2*	zona pellucida glycoprotein 3, tandem duplicate 2	0	-11.9538	5.45E-191	–
TRINITY_DN317005_c0_g2_i8		uncharacterized protein LOC108875206	5.00E-82	-11.9252	2.24E-232	–
TRINITY_DN329214_c4_g1_i1	–	–	–	11.9153	2.34E-59	–
TRINITY_DN330473_c0_g1_i1	*org*	oogenesis-related gene	1.10E-28	-11.9086	0	–
TRINITY_DN318617_c3_g1_i5	*anxa1*	annexin A1	1.10E-155	-11.8954	2.70E-287	–
TRINITY_DN312099_c4_g6_i1	*znf146*	zinc finger protein OZF	4.80E-06	11.8615	9.45E-66	–
TRINITY_DN324025_c9_g1_i11	*cfap77*	cilia- and flagella-associated protein 77	2.90E-76	11.8446	2.88E-185	–
TRINITY_DN330053_c0_g2_i13	*lrrc71*	leucine-rich repeat-containing 713	1.90E-99	11.8284	3.27E-67	MEyellow4
TRINITY_DN331894_c1_g9_i1	–	–	–	-11.8082	8.97E-83	MEgrey60
TRINITY_DN323179_c3_g2_i3	*wfdc18*	WAP four-disulfide core domain protein 18-like	3.70E-11	-11.8081	4.10E-149	MEmediumorchid
TRINITY_DN323199_c3_g1_i3	*drc7*	Coiled-coil domain containing 135	2.20E-87	11.8054	4.72E-69	–
TRINITY_DN334848_c6_g8_i4	*lrrc18a*	leucine rich repeat containing 18a	3.70E-68	11.7979	2.06E-118	–
TRINITY_DN284263_c0_g1_i2	*ctsz*	cathepsin Z	2.00E-61	-11.7791	6.99E-71	–
TRINITY_DN373777_c0_g1_i1	–	–	–	-11.7658	8.36E-69	MEblue
TRINITY_DN324249_c0_g1_i13	*atg16l1*	Autophagy-related protein 16-1	5.00E-37	-11.7558	4.76E-62	MEskyblue
TRINITY_DN305805_c0_g1_i1	*dnaaf4*	Dynein axonemal assembly factor 4	4.30E-70	11.7099	4.13E-64	MEbrown2
TRINITY_DN314116_c0_g2_i6	*izumo1*	izumo sperm-egg fusion protein 1	2.80E-154	11.6957	5.83E-67	MEyellow4
TRINITY_DN326505_c1_g1_i1	*zpc1*	Zona pellucida protein C1	2.70E-136	-11.6880	2.92E-112	MEbrown
TRINITY_DN323330_c2_g2_i3	–	–	–	-11.6841	7.97E-163	MEmediumorchid
TRINITY_DN312005_c7_g1_i2	–	–	–	11.6509	1.44E-65	MEturquoise

TRINITY_DN313189_c1_g1_i5	*gle1*	GLE1 RNA export mediator homolog (yeast)	5.10E-20	11.6316	1.51E-61	–
TRINITY_DN329159_c3_g2_i2	*wag22*	WAG22 antigen-like	1.30E-06	-11.6022	4.54E-204	MEmediumorchid
TRINITY_DN326248_c1_g9_i1	–	–	–	11.5867	2.91E-58	–
TRINITY_DN330535_c1_g1_i2		si:dkey-27p23.3	1.90E-27	11.5617	2.45E-64	MEnavajowhite2
TRINITY_DN320436_c0_g1_i2		hypothetical protein EH28_03918	3.00E-12	-11.5602	2.12E-65	–
TRINITY_DN311544_c0_g1_i2	*cldnd*	claudin d	1.90E-52	-11.5593	2.57E-193	–
TRINITY_DN320566_c0_g1_i3	–	–	–	11.5555	1.35E-61	–
TRINITY_DN324538_c2_g3_i2	*tbx16*	T-box 16	5.70E-116	-11.5550	3.61E-63	MEskyblue
TRINITY_DN323669_c0_g1_i2	*btg4*	B-cell translocation gene 4	8.50E-105	-11.5485	6.27E-203	–
TRINITY_DN320283_c2_g3_i5	–	–	–	11.5197	1.03E-63	–
TRINITY_DN323657_c0_g1_i3	*sparcl1*	SPARC-like 1	2.80E-113	-11.5160	3.18E-240	–
TRINITY_DN331659_c0_g2_i1	–	–	–	-11.5155	2.15E-121	–
TRINITY_DN331480_c1_g4_i1		dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1-like	6.00E-07	-11.4769	9.54E-61	–
TRINITY_DN331480_c1_g9_i1	–	–	–	-11.4503	1.06E-201	MEblue
TRINITY_DN296038_c0_g1_i4	*dnal1*	Dynein light chain 1, axonemal	3.20E-30	11.4352	1.05E-60	–
TRINITY_DN326174_c0_g2_i6	*pih1d2*	PIH1 domain containing 2	2.80E-92	11.4317	5.37E-58	MEblue
TRINITY_DN322719_c0_g1_i2	*tex36*	testis expressed 36	3.90E-43	11.4249	3.54E-59	MEnavajowhite2
TRINITY_DN309732_c0_g1_i2	*myb*	v-myb avian myeloblastosis viral oncogene homolog	1.60E-58	11.4241	9.23E-63	MEyellowgreen
TRINITY_DN315851_c0_g1_i16	*cep55*	Centrosomal protein 55	1.60E-98	11.4228	1.11E-141	MEbrown4
TRINITY_DN326095_c4_g1_i1		gastrula zinc finger protein xFG20-1-like	8.60E-21	11.4103	2.36E-61	–
TRINITY_DN335379_c3_g3_i1	*cluap1*	Clusterin-associated protein 1 homolog	1.80E-110	11.3986	1.27E-59	–
TRINITY_DN334506_c0_g2_i4	*zp3*	Zona pellucida sperm-binding protein 3	2.20E-13	-11.3950	8.89E-164	–
TRINITY_DN322001_c5_g1_i3	*g0s2*	G0/G1 switch regulatory protein 2	1.40E-31	11.3860	1.61E-265	MEblack
TRINITY_DN333153_c4_g2_i1		sperm-associated antigen 8	2.50E-26	11.3760	6.15E-62	MEyellow4
TRINITY_DN335169_c3_g1_i7	*sycp1*	Synaptonemal complex protein 1	9.80E-163	11.3733	0	MEtan
TRINITY_DN256186_c0_g1_i1	–	–	–	-11.3712	3.82E-105	MEmediumorchid
TRINITY_DN301575_c0_g2_i2	–	–	–	-11.3516	1.65E-210	–
TRINITY_DN331672_c0_g1_i6	*prss27*	protease, serine, 27	4.20E-91	-11.3321	2.58E-78	–
TRINITY_DN320885_c0_g1_i2	–	–	–	-11.3311	2.56E-174	–
TRINITY_DN302684_c0_g1_i1	–	–	–	11.3287	4.03E-58	–
TRINITY_DN325908_c0_g1_i2		Hypothetical protein EH28_18302	2.00E-07	11.3278	1.92E-58	–

Table S13 – List of the top 1% differentially expressed genes in the gonads between individuals of Culatra and individuals of Trieste ordered by fold-change (FC). Positive log2FC indicates an up-regulation of the transcript for Trieste individuals, whereas a negative log2FC indicates an up-regulation of the transcript for Culatra individuals.

Gene identifier	Gene Symbol	Gene Description	E-value	log2FC	padj	WGCNA
TRINITY_DN327562_c1_g1_i2		Collagen alpha-1(XXI) chain	0	-9.0863	9.48E-05	MEyellowgreen
TRINITY_DN298900_c1_g3_i4		Uromodulin	5.60E-105	-8.7750	0.0052	–
TRINITY_DN183828_c0_g1_i1	–	–	–	8.1004	0.0003	MEpalevioletred2
TRINITY_DN334281_c4_g16_i1	*ift74*	intraflagellar transport 74	1.20E-42	8.0862	3.11E-10	MEturquoise
TRINITY_DN322848_c2_g1_i5	*nptx1*	Neuronal pentraxin-1	0	-7.7794	0.0005	–

Table S14 – Summary of correlations between modules' eigengene (ME) and trait 'Population' for each network, with respective adjusted *P*-values, module size and number of differentially expressed genes (DEGs), in any of the comparisons, present in the module. Additionally, top GO term for each category, obtained from conditional enrichment analysis with unadjusted *P*-value < 0.01, are also shown.

a) Forebrain

Module name	Size	Cor	Padj	DEG	Top GO term	Figure ref.
MEblack	1703	0.83	0.0002	97	**BP:** G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger; **CC:** voltage-gated potassium channel complex; **MF:** voltage-gated potassium channel activity;	Fig. S12A
MEmagenta	1205	0.96	4.00E-10	119	**BP:** RNA processing; **CC:** mitochondrial inner membrane; **MF:** isocitrate dehydrogenase (NADP+) activity	Fig. S12B
MEpaleturquoise	37	0.64	0.0265	0	**BP:** midbrain-hindbrain boundary structural organization; **CC:** Ctf18 RFC-like complex; **MF:** ethanolamine kinase activity	Fig. S12C
MEpink	1522	0.70	0.0087	341	**BP:** cell activation involved in immune response; **CC:** vacuolar part; **MF:** serine-type endopeptidase activity	Fig. S12D
MEgreenyellow	704	−0.77	0.0021	47	**BP:** nucleic acid-templated transcription; **CC:** transcription factor complex; **MF:** transcription factor activity, sequence-specific DNA binding	Fig. S13A
MElightyellow	147	−0.70	0.0087	11	**BP:** ribosomal protein import into nucleus; **CC:** microtubule minus-end; **MF:** GABA-gated chloride ion channel activity	Fig. S13B
MEred	1706	−0.98	1.24E-12	265	**BP:** lipid oxidation; **CC:** shelterin complex; **MF:** succinate-semialdehyde dehydrogenase [NAD(P)+] activity	Fig. S13C
MEyellow	1975	−0.75	0.0024	90	**BP:** homophilic cell adhesion via plasma membrane adhesion molecules; **CC:** intrinsic component of membrane; **MF:** axon guidance receptor activity	Fig. S13D

b) Female Gonad

Module name	Size	Cor	Padj	DEGs	Top GO term	Figure ref.
MEantiquewhite4	107	0.95	0.0013	91	**BP:** negative regulation of alkaline phosphatase activity; **CC:** cell periphery; **MF:** receptor activity	Fig. S15A
MEdarkseagreen3	58	0.89	0.0089	32	**BP:** embryonic body morphogenesis; **CC:** site of DNA damage; **MF:** serine-pyruvate transaminase activity	Fig. S15B

MEgreenyellow	291	0.87	0.0144	235	**BP:** positive regulation of superoxide anion generation; **CC:** integral component of membrane; **MF:** carbohydrate transmembrane transporter activity	Fig. S15C
MElightyellow	229	0.92	0.0036	170	**BP:** positive regulation of phosphorylation of RNA polymerase II C-terminal domain serine 2 residues; **CC:** cyclin K-CDK12 complex; **MF:** fibroblast growth factor-activated receptor activity	Fig. S15D
MEorangered4	152	0.85	0.0195	88	**BP:** galactose catabolic process; **CC:** nucleosome; **MF:** acylglycerol lipase activity	Fig. S15E
MEskyblue3	154	0.99	9.51E-06	123	**BP:** cilium assembly; **CC:** BBSome; **MF:** DNA topoisomerase type I activity	Fig. S15F
MEturquoise	1127	0.91	0.0054	782	**BP:** ciliary basal body organization; **CC:** CA3 pyramidal cell dendrite; **MF:** intermembrane lipid transfer activity	Fig. S15G
MEdarkgreen	199	−0.92	0.0037	119	**BP:** peptide metabolic process; **CC:** ribosome; **MF:** structural constituent of ribosome	Fig. S16A
MEdarkolivegreen	162	−0.86	0.0170	114	**BP:** regulation of T cell receptor signaling pathway; **CC:** endoplasmic reticulum quality control compartment; MF: HMG box domain binding	Fig. S16B
MEdarkseagreen4	111	−0.85	0.0191	93	**BP:** regulation of vesicle size; CC: intracellular organelle part; **MF:** MH2 domain binding	Fig. S16C
MEgreen	659	−0.99	9.24E-07	515	**BP:** physiological muscle hypertrophy; **CC:** integral component of membrane; **MF:** estradiol 17-beta-dehydrogenase activity	Fig. S16D
MElightgreen	232	−0.97	0.0002	171	**BP:** positive regulation of glycogen biosynthetic process; **CC:** microvillus; **MF:** methyltransferase activity	Fig. S16E
MEpaleturquoise	163	−0.97	0.0002	119	**BP:** plasma membrane repair; **CC:** membrane raft; **MF:** vascular endothelial growth factor receptor binding	Fig. S16F
MEpalevioletred3	123	−0.90	0.0071	111	**BP:** microtubule bundle formation; **CC:** core mediator complex; **MF:** kinetochore binding	Fig. S16G
MEsalmon4	124	−0.89	0.0080	91	**BP:** protein-chromophore linkage; **CC:** extracellular exosome; **MF:** peptide binding	Fig. S16H
MEskyblue	170	−0.94	0.0022	121	**BP:** positive regulation of adenosine receptor signaling pathway; **CC:** AMPA glutamate receptor complex; **MF:** calcium ion transmembrane transporter activity	Fig. S16I
MEtan	286	−0.91	0.0051	226	**BP:** recombinational repair; **CC:** nuclear proteasome complex; **MF:** TBP-class protein binding	Fig. S16J

c) Male Gonad

Module name	Size	Cor	Padj	DEGs	Top GO term	Figure ref.
MEantiquewhite2	51	0.94	0.0018	44	**BP:** negative regulation of mRNA splicing, via spliceosome; **CC:** cyclin D2-CDK6 complex; **MF:** mRNA CDS binding	Fig. S17A
MEbisque4	141	0.86	0.0088	93	**BP:** tryptophanyl-tRNA aminoacylation; **CC:** cyclin B1-CDK1 complex; **MF:** tryptophan-tRNA ligase activity	Fig. S17B
MEblue	986	0.87	0.0077	588	**BP:** nucleic acid metabolic process; **CC:** nucleus; **MF:** chromatin binding	Fig. S17C
MEblue2	71	0.78	0.0350	43	**BP:** formation of cytoplasmic translation initiation complex; **CC:** eukaryotic 48S preinitiation complex; **MF:** dUTP diphosphatase activity	Fig. S17D
MEbrown	953	0.76	0.0449	616	**BP:** RNA splicing, via transesterification reactions; **CC:** intracellular organelle lumen; **MF:** RNA binding	Fig. S17E
MEbrown2	72	0.90	0.0054	52	**BP:** clustering of voltage-gated sodium channels; **CC:** cytosol; **MF:** inositol-1,4,5-trisphosphate 3-kinase activity	Fig. S17F
MEbrown4	141	0.99	1.13E-05	112	**BP:** regulation of molecular function, epigenetic; **CC:** Scc2-Scc4 cohesin loading complex; **MF:** chromatin insulator sequence binding	Fig. S17G
MEdarkorange2	143	0.82	0.0199	113	**BP:** tubulin complex assembly; **CC:** centrosome; **MF:** histone kinase activity (H3-T3 specific)	Fig. S17H
MEgrey60	235	0.88	0.0077	142	**BP:** RNA splicing, via transesterification reactions; **CC:** ribonucleoprotein complex; **MF:** nucleic acid binding	Fig. S17I
MElavenderblush3	112	0.89	0.0057	81	**BP:** DNA modification; **CC:** condensed nuclear chromosome, centromeric region; **MF:** chromo shadow domain binding	Fig. S17J
MElightcyan	248	0.86	0.0095	163	**BP:** regulation of cellular component organization; **CC:** microtubule plus-end; **MF:** phospholipase D activator activity	Fig. S17K
MEmagenta	330	0.95	0.0013	228	**BP:** DNA integration; **CC:** proteasome core complex, beta-subunit complex; **MF:** threonine-type endopeptidase activity	Fig. S17L
MEmagenta4	60	0.93	0.0024	40	**BP:** nucleoside triphosphate biosynthetic process; **CC:** Swi5-Sfr1 complex; **MF:** formate-tetrahydrofolate ligase activity	Fig. S17M
MEmediumorchid	105	0.87	0.0088	70	**BP:** collagen fibril organization; **CC:** cis-Golgi network; **MF:** peptidase inhibitor activity	Fig. S17N
MEnavajowhite2	122	0.91	0.0039	103	**BP:** NADP biosynthetic process; **CC:** cell tip; **MF:** NAD+ kinase activity	Fig. S17O

MEorangered3	84	0.85	0.0112	65	**BP:** positive regulation of release of cytochrome c from mitochondria; **CC:** translation initiation ternary complex; **MF:** membrane insertase activity	Fig. S17P
MEpalevioletred2	60	0.88	0.0077	43	**BP:** protein localization to cell surface; **CC:** Shu complex; **MF:** FMN adenylyltransferase activity	Fig. S17Q
MEplum	88	0.92	0.0024	66	**BP:** regulation of multivesicular body size involved in endosome transport; **CC:** ESCRT II complex; **MF:** fumarylacetoacetase activity	Fig. S17R
MEskyblue2	102	0.90	0.0054	77	**BP:** positive regulation of testosterone secretion; **CC:** lumenal side of Golgi membrane; **MF:** aspartic endopeptidase activity, intramembrane cleaving	Fig. S17S
MEsteelblue	165	0.91	0.0037	127	**BP:** bicellular tight junction assembly; **CC:** plasma membrane bounded cell projection; **MF:** 1-phosphatidylinositol 4-kinase activity	Fig. S17T
MEtan	286	0.90	0.0054	226	**BP:** recombinational repair; **CC:** nuclear proteasome complex; **MF:** TBP-class protein binding	Fig. S17U
MEthistle1	127	0.93	0.0023	86	**BP:** positive regulation of telomere maintenance via telomerase; **CC:** SCF ubiquitin ligase complex; **MF:** NADPH-hemoprotein reductase activity	Fig. S17V
MEturquoise	1127	0.95	0.0013	782	**BP:** ciliary basal body organization; **CC:** CA3 pyramidal cell dendrite; **MF:** intermembrane lipid transfer activity	Fig. S17W
MEyellow4	95	0.86	0.0088	84	**BP:** synaptonemal complex assembly; **CC:** basal cortex; **MF:** polyamine binding	Fig. S17X
MEantiquewhite4	107	−0.88	0.0071	91	**BP:** negative regulation of alkaline phosphatase activity; **CC:** cell periphery; **MF:** receptor activity	Fig. S18A
MEblack	418	−0.83	0.0148	346	**BP:** 'de novo' IMP biosynthetic process; **CC:** peroxisome; **MF:** phosphoribosylformylglycinamidine cyclo-ligase activity	Fig. S18B
MEcoral1	110	−0.80	0.0257	77	**BP:** antigen processing and presentation; **CC:** MHC class II protein complex; **MF:** threonine-type endopeptidase activity	Fig. S18C
MEcyan	268	−0.84	0.0129	171	**BP:** cellular response to cAMP; **CC:** cell cortex; **MF:** aconitate hydratase activity	Fig. S18D
MEdarkmagenta	157	−0.86	0.0092	111	**BP:** negative regulation of protein homodimerization activity; **CC:** catenin complex; **MF:** receptor tyrosine kinase binding	Fig. S18E
MEfirebrick4	75	−0.87	0.0077	60	**BP:** oxygen transport; CC: lytic vacuole; **MF:** oxygen binding	Fig. S18F
MEgreen	659	−0.98	8.53E-05	515	**BP:** physiological muscle hypertrophy; **CC:** integral component of membrane; **MF:** estradiol 17-beta-dehydrogenase activity	Fig. S18G

MEgreenyellow	291	−0.86	0.0095	235	**BP:** positive regulation of superoxide anion generation; **CC:** integral component of membrane; **MF:** carbohydrate transmembrane transporter activity	Fig. S18H
MEhoneydew	59	−0.83	0.0148	44	**BP:** transport; **CC:** heterotrimeric G-protein complex; **MF:** neurotransmitter:sodium symporter activity	Fig. S18I
MEhoneydew1	111	−0.79	0.0323	82	**BP:** oxidoreduction coenzyme metabolic process; **CC:** phosphopyruvate hydratase complex; **MF:** formaldehyde dehydrogenase activity	Fig. S18J
MEivory	145	−0.78	0.0366	100	**BP:** cellular monovalent inorganic cation homeostasis; **CC:** plasma membrane part; **MF:** sugar transmembrane transporter activity	Fig. S18K
MEorange	173	−0.82	0.0199	125	**BP:** positive regulation of fibroblast growth factor receptor signaling pathway; **CC:** basolateral plasma membrane; **MF:** inorganic cation transmembrane transporter activity	Fig. S18L
MEpink	407	−0.93	0.0024	311	**BP:** organic acid catabolic process; **CC:** mitochondrion; **MF:** formate-tetrahydrofolate ligase activity	Fig. S18M
MEpurple	321	−0.98	2.81E-05	230	**BP:** G1 to G0 transition; **CC:** elastic fiber; **MF:** calcidiol 1-monooxygenase activity	Fig. S18N
MEsaddlebrown	169	−0.87	0.0087	127	**BP:** response to folic acid; **CC:** integral component of membrane; **MF:** heme transporter activity	Fig. S18O
MEskyblue1	89	−0.84	0.0122	66	**BP:** response to progesterone; **CC:** NF-kappaB p50/p65 complex; **MF:** transcription factor activity, sequence-specific DNA binding	Fig. S18P
MEyellow	880	−0.88	0.0077	618	**BP:** regulation of cell motility; **CC:** receptor complex; **MF:** macromolecular complex binding	Fig. S18Q
MEyellowgreen	155	−0.90	0.0048	112	**BP:** cellular potassium ion homeostasis; **CC:** sodium:potassium-exchanging ATPase complex; **MF:** sodium:potassium-exchanging ATPase activity	Fig. S18R

Gene Ontology categories: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF).