



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PHD PROGRAMME IN
Innovative technologies and sustainable use of Mediterranean Sea fishery and
biological resources (FishMed-PhD)

Cycle XXXVII

Settore Consorsuale: 05/B1 – ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/05 – ZOOLOGIA

**Integrative perspectives on bivalve sex
determination: A comparative and evolutionary
analysis across phylogeny**

Candidate: Filippo Nicolini

PhD Coordinator:
prof. Stefano Goffredo

Supervisor:
prof. Andrea Luchetti

Final exam 2025

Table of contents

Table of contents	i
List of abbreviations	1
Abstract	4
Chapter 1: Introduction	6
1.1 The diversity of sexual processes in animals	6
1.2 Genetic sex determination and the evolution of sex-determination related genes .	9
1.3 Unravelling sex determination in bivalves	10
Chapter 2: Project outlook and objectives	14
Chapter 3: Bivalves as emerging model systems to study the mechanisms and evolution of sex determination: a genomic point of view	16
3.1 Open yet inspiring topics in bivalve sex determination	18
3.1.1 Transitions between environmental and genetic sex determination	18
3.1.2 Evolution of sex chromosomes	19
3.1.3 Mito-nuclear interactions	20
3.1.4 Evolution of sex-determination related genes	22
3.2 The case of the Dmrt gene family in bivalves	23
3.3 Conclusions: bivalves as new models in the study of sex determination	26
3.4 Acknowledgments	31
3.5 Data Availability	31
Chapter 4: Identification of putative sex-determination related genes in bivalves through comparative molecular evolutionary analyses	32
4.1 Introduction	33
4.2 Materials and methods	36
4.2.1 Dataset of bivalve annotated genomes and transcriptomes	36

4.2.2	Identification and classification of Dmrt, Sox and Fox genes in bivalves	36
4.2.3	Sequence diversity of bivalve single-copy orthogroups	38
4.2.4	Mammals and <i>Drosophila</i> spp. as test datasets	39
4.2.5	GO-term enrichment	39
4.3	Results	41
4.3.1	Genomic and transcriptomic datasets	41
4.3.2	The Dmrt, Sox, and Fox complements in bivalves	41
4.3.3	Amino acid sequence divergence of Dmrt, Sox, and Fox genes in bivalves	43
4.3.4	Dmrt, Sox, and Fox genes, and amino acid sequence divergence in the test datasets	50
4.4	Discussion	52
4.4.1	A new manually-curated and phylogenetic-based reference dataset of Dmrt, Sox, and Fox genes in bivalves	52
4.4.2	High amino acid sequence divergence identifies putative sex-determining genes	56
4.5	Conclusions	61
Chapter 5: Localisation of three sex-related genes and the germline marker <i>Vasa/Vasa</i> in the early developmental stages of <i>Mytilus galloprovincialis</i>		63
5.1	Introduction	64
5.2	Materials and methods	66
5.2.1	Time-series gene expression	66
5.2.2	Sample collection, MitoTracker staining and fixation	67
5.2.3	mRNA <i>in-situ</i> hybridization chain reaction (HCR)	68
5.2.4	Immunolocalization of Vasa	70
5.3	Results	71
5.3.1	Differential gene expression analysis of <i>Vasa</i> and SRGs in embryo time- series	71
5.3.2	mRNA <i>in-situ</i> HCR of <i>Vasa</i> and SRGs	74
5.3.3	Immunolocalization of Vasa	75
5.4	Discussion	77
5.4.1	Sperm mitochondria are not detected after 12 hpf because of MitoTracker misincorporation or fading	77

5.4.2 Exploring the processes of SD in the <i>M. galloprovincialis</i> early development	80
5.4.3 Primordial germ cells are specified by both preformation and epigenesis in <i>M. galloprovincialis</i>	83
5.5 Conclusion	86
Chapter 6: Conclusions	88
References	113
Appendix	114
Supplementary figures	120
Supplementary tables	137
Data availability	217

List of abbreviations

AASD	amino acid sequence divergence
CMS	cytoplasmatic male sterility
CUE	coupling of ubiquitin conjugation to endoplasmic reticulum degradation [domain]
DEAD/DEAH-box	Asp-Glu-Ala-Asp/Asp-Glu-Ala-His box
DGE	differential gene expression
dpf	days post fertilization
DM	<i>dsx</i> and <i>mab-3</i> [domain]
DMA	DM-associated [domain]
Dmrt	<i>dsx</i> and <i>mab-3</i> related transcription factor
Dmrt-1L	<i>Dmrt 1-like</i>
Dm-W	<i>dsx</i> and <i>mab-3</i> related gene <i>W</i>
Dmy	<i>dsx</i> and <i>mab-3</i> related gene <i>Y</i>
DSFG	Dmrt, Sox, and Fox gene
dsx	<i>doublesex</i>
DUI	doubly uniparental inheritance
ESD	environmental sex determination
FASW	filtered artificial sea water
FHA	forkhead-associated [domain]
Fox	forkhead box

GC	germ cell
GO	gene ontology
GRN	gene regulatory network
GSD	genetic sex determination
HCR	hybridization chain reaction
HeSC	heteromorphic sex chromosome
HMG	high mobility group [box domain]
HMM	hidden Markov model
HoSC	homomorphic sex chromosome
hpf	hours post fertilization
mab-3	<i>male abnormal-3</i>
MCL	Markov clustering algorithm
ML	maximum likelihood
mRNA-ISH	mRNA <i>in-situ</i> hybridization
Mya	million years ago
ORF	open reading frame
PBS	1× phosphate-buffered saline
PBS-Tw	1× PBS with 0.1% Tween 20
PCA	principal component analyses
PFA	paraformaldehyde
PGC	primordial germ cell
pPGC	presumptive primordial germ cell
PSC	pluripotent stem cell
qRT-PCR	quantitative real-time polymerase chain reaction

RNAi RNA interference

RT room temperature

SC sex chromosome

SCO single-copy orthogroup

SD sex determination

SDf sex differentiation

SDG sex-determining gene

Sox *Sry*-related HMG-box

SRG sex-determination related gene

Sry *Sex-determining region of chromosome Y*

SSC-Tw 5× saline-sodium citrate with 0.1% Tween 20

Sxl *Sex-lethal*

TBS 1× Tris-buffered saline

TBS-Tx 1× TBS with Triton X-100

tra *transformer*

TSD temperature-dependent [environmental] sex determination

Abstract

Sex determination (SD) in bivalves remains a largely unexplored field, despite the socio-economic importance of many species. This can be traced back to the diversity of mechanisms observed across species, involving both genetic and environmental factors and apparently lacking heteromorphic sex chromosomes, which hamper a straightforward scientific research. This study presents an integrative approach that combines comparative genomics, phylogenetic analyses, and *in-situ* visualisation techniques to investigate the molecular basis of SD in bivalve molluscs. Using a broad phylogenetic and genomic framework, key components of gene families known for their roles in SD across Metazoa were identified and analysed through the lens of comparative genomics. Particularly, considering that sex-determining genes tend to evolve faster than genes not involved in SD, we leveraged the tools of molecular evolution to identify highly-divergent genes among the *Dmrt*, *Sox*, and *Fox* gene families. Both *Dmrt-1L* and *Sox-H* were found to be included in the group of bivalve fast-evolving genes, giving support to previous works which appointed them as tightly involved with male SD in bivalves. To further investigate the roles of these genes, mRNA *in-situ* hybridization chain reaction (HCR) was employed to look at their transcription patterns during the embryonic and early larval stages of *Mytilus galloprovincialis*, along with expression patterns of *Fox-L2*—a gene previously associated with bivalve female SD, and of the germline marker *Vasa/Vasa*. Both *Dmrt-1L* and *Sox-H* were found to be not transcribed during the sampled stages (up until 72 hours post fertilization), while *Fox-L2* showed an increasing sex-unbiased expression with the onset of gastrulation. Therefore, SD is likely not happening during these early developmental stages. This observation aligns with the expression of *Vasa/Vasa*, whose specification of primordial germ cells (PGCs) seemed to be relying on a mixed process of preformation and epigenesis. Before gastrulation, both *Vasa/Vasa* is homogeneously present in all blastomeres, thus not labelling presumptive PGCs univocally. The process of PGCs formation seems instead to start after the formation of the

larvae, when *Vasa*-positive cells begin to accumulate in two lateral areas at both sides of the larvae. Therefore, SD can be expected to occur only after this stage. The present work shows the importance of employing an integrative analysis when investigating overlooked processes in non-model organisms. Particularly, this contributes a foundational reference for SD in bivalves, broadening our understanding of the genetic factors shaping reproductive biology in this ecologically and economically significant group.

Chapter 1

Introduction

1.1 The diversity of sexual processes in animals

The process of sex determination (SD) has been traditionally associated with the very first steps of gonad differentiation, where an initial trigger (or master switch) activates the molecular pathway that establishes the sexual identity of an organism. According to this view, two alternative types of SD have been traditionally recognised: the genetic sex determination (GSD) and the environmental sex determination (ESD), depending on whether the very first cues are of genetic or environmental origin. All the downstream events of gonad and morphological sex-specific development (i.e., after SD) have been instead appointed as sex differentiation (SDf), which consists of the entire set of morphogenetic, molecular, and physiological events leading to the full maturation of testes or ovaries and secondary sexual characters (**Uller and Helanterä, 2011; Bear and Monteiro, 2013; Beukeboom and Perrin, 2014**). GSD is found in many classical model systems—such as the mouse *Mus musculus*, the fruit fly *Drosophila melanogaster*, and the nematode *Caenorhabditis elegans*, but is by far the most prevalent SD system in animals, as it occurs in the majority of vertebrates and arthropods (**Bachtrög et al., 2014; Beukeboom and Perrin, 2014**). It encompasses a variety of sex-determining cues, ranging from single master genes (e.g., *Sex-determining region of chromosome Y (Sry)* in eutherians), to polygenic networks (as in the zebrafish *Danio rerio*) and chromosome countings (as in *D. melanogaster* and *C. elegans*). Conversely, ESD is more uncommon and is found mainly in reptiles, fishes, insects, crustaceans, annelids, nematodes, and rotifers (reviewed in **Korpelainen, 1990; Bachtrög et al., 2014**). It relies on several initial stimuli of environmental origin, such as light, food availability, and population density, though the most common process is the temperature-dependent SD

(TSD; **Bachtrög et al., 2014**; **Beukeboom and Perrin, 2014**). Currently, the molecular basis by which an environmental signal is transduced into the canalization of the male or female developmental pathway is unknown (**Bachtrög et al., 2014**; **Capel, 2017**).

Lately, a growing number of studies have challenged the traditionally binary views of both GSD/ESD and SD/SDf (**Bear and Monteiro, 2013**; **Bachtrög et al., 2014**; **Beukeboom and Perrin, 2014**; **Todd et al., 2016**; **Capel, 2017**). On the one hand, the characterisation of SD in new species has shown that GSD and ESD represent the ends of a continuum of mixed conditions, rather than two mutually exclusive phenomena. For example, in the red-eared slider turtle *Trachemys scripta*, a species with TSD, it has been shown that pairs of gonads cultured separately at the same pivotal temperature (i.e., the temperature producing 50% of males and 50% of females in offspring) tend to predominantly differentiate into one sex. Therefore, an underlying genetic/epigenetic mechanism controlling SD should exist when temperature effect is absent (**Mork et al., 2014**). In the Australian bearded-dragon *Pogona vitticeps*, some ZZ males were caught to sex-revert to fertile females in the wild after incubation at high temperatures, constituting a natural case of transition from GSD to ESD (in the form of TSD; **Holleley et al., 2015**). On the other hand, instead, the high evolutionary dynamics and the variable expression patterns of the genes involved in the processes of gonad commitment and development make the distinction between SD and SDf increasingly blurred (**Bear and Monteiro, 2013**; **Bachtrög et al., 2014**; **Beukeboom and Perrin, 2014**; **Todd et al., 2016**; **Capel, 2017**). In fact, considering that the SD of an organism may follow different routes, such as being triggered only in presumptive gonads—which then establish the sexual identity of the rest of the organism through hormone signalling (gonadal SD; a process traditionally associated to mammals), or occurring independently in every cell of the developing embryo (cell-autonomous SD; a process traditionally associated to fruit flies and nematodes), the tempo and modes of SDf may vary significantly among species and may not even necessarily depend on or originate from SD itself (**Bear and Monteiro, 2013**; **Capel, 2017**). Classic examples of the independence between SD and SDf are provided by gynandromorph animals (where both the male and the female phenotypes are found in the same organism in a bilaterally-distributed fashion), but also by the marsupial mammal tammar wallaby (*Macropus* [now *Notamacropus*] *eugenii*). In gynandromorph chickens (*Gallus gallus*), for example, it has been shown that the male half of the animal is made up mainly by ZZ cells, while the female half by ZW cells, and that each is selectively susceptible to either male or female gonad-secreted hormones, respectively.

This shows how in birds SD, which is initiated in the presumptive gonads, is independent from and occurs later than SDf, which is instead triggered by a cell-autonomous mechanism (**Zhao et al., 2010**). Similarly, in *N. eugenii* the somatic differentiation of the scrotum and the pouch precedes that of the gonads, indicating that their development (as part of SDf) is triggered by genes linked to the X chromosome, rather than by sex-differentiation inducing hormones as in other mammals (**Renfree et al., 1996**).

Considering this complex scenario, **Uller and Helanterä (2011)** proposed a unified and broad-scope definition for SD, that is, “the processes within an embryo leading to the formation of differentiated gonads as either testes or ovaries”, thus excluding any actual distinction between environmental/genetic initial triggers or the downstream effectors. However, it can be argued that this definition should be even more expanded to encompass not only the embryonic stage of the animal life cycle but also adulthood, since cases of sex reversals (sequential hermaphroditism) legitimately express proper SD processes also during post-embryonic life stages. For example, fishes represent a noteworthy example of how the establishment of sexual fate is not an irreversible process in an organism’s life, but instead a tradeoff trait, involving antagonistic gene regulatory networks (GRNs), that can be switched from one side to the other (**Todd et al., 2016; Capel, 2017**). Sex reversal in fishes is typically prompted by environmental signals, such as population density, sex ratio at spawning, and social factors, but also by the attainment of a threshold size and/or age (reviewed in **Todd et al., 2016**). Regardless of its biology and direction (i.e., from males to females [protandry], from females to males [protogyny], or both ways), sex reversal often results in the complete restructuring of gonads, as well as in remarkable changes in both somatic morphology and behaviour. Therefore, sequential hermaphroditism can be considered the expression of a genuine SD program inducing a second round of SDf in adult organisms (**Todd et al., 2016**).

Overall, decades of studies have revealed that SD is strikingly diverse among animals, even in closely related species. Therefore, the research effort is currently devoted to further characterising sex-determining processes in new species, as well as to understand how this fundamental aspect of animal development may sustain such a high plasticity among species (**Beukeboom and Perrin, 2014; Todd et al., 2016; Capel, 2017**).

1.2 Genetic sex determination and the evolution of sex-determination related genes

In its most intimate core, animal SD is the manifestation of complex GRNs where, according to **Wilkins (1995)**, the downstream actors appear to be nearly conserved across species, while the master top regulators (the commonly recognized sex determinants, such as the *Sry* in therians or the ratio between sex and autosomal chromosomes in *Drosophila*) are often the most variable part (**Matson and Zarkower, 2012; Mullon et al., 2012; Bachtrog et al., 2014; Beukeboom and Perrin, 2014**). Such a differential pattern of molecular evolution is considered to be the direct result of the mechanism by which a sex-determining cascade is assembled. Particularly, it has been proposed that a GRN may take on a role in SD through a retrograde growth, i.e., by progressively adding upper regulators in a bottom-up process (**Stothard and Pilgrim, 2003; Mullon et al., 2012; Capel, 2017**). This mechanism regards the SD cascade in *Drosophila* species (**Mullon et al., 2012**), *C. elegans* (**Stothard and Pilgrim, 2003**), and vertebrates, although the latter case has been questioned several times (reviewed in **Capel, 2017**). Remarkably, it appears that some gene families are more prone than others to be recruited in SD, as either primary sex-determining genes (SDGs) or in some key part of the cascade. For example, components of the *dsx* and *mab-3* related transcription factor (Dmrt) gene family have a main role as bottom effectors in the SD cascade of many animal species, as seen in *D. melanogaster* with *doublesex* (*dsx*) and in *C. elegans* with *male abnormal-3* (*mab-3*), but also in other invertebrates and the majority of vertebrates. In these cases, Dmrt genes dictate the sex-specific development in response to the primary SD decision (**Matson and Zarkower, 2012**). Nevertheless, paralogs of *Dmrt-1* have also repeatedly and independently taken on the role as SDGs in several vertebrate species, as in the medaka fish *Oryzias latipes* with *dsx* and *mab-3* related gene *Y* (*Dmy*), in the African clawed frog *Xenopus laevis* with *dsx* and *mab-3* related gene *W* (*Dm-W*), and in *G. gallus* with the Z-linked *Dmrt-1* (reviewed in **Matson and Zarkower, 2012; Mawaribuchi et al., 2019**). A similar but even more conserved sex-determining genetic axis is found in insects, where *transformer* (*tra*) directs the sex-specific splicing of *dsx* in almost every species investigated so far (**Verhulst et al., 2010; Bopp et al., 2014**). The GRN in which the *tra-dsx* module is placed, is instead more diversified and species-specific, as the top- and bottom-most parts are highly divergent, resulting in a SD cascade that can be represented by an hour-glass model (**Bopp et al., 2014**). Similarly, other highly-conserved genes involved in SD has been identified,

particularly as downstream effectors in vertebrates: these includes for example *Fox-L2* from the forkhead box (Fox) gene family and *Sox-9* from the *Sry*-related HMG-box (Sox) gene families, acting in the female- and male-specific cascades, respectively (**Capel, 2017**).

The significance of molecular evolution in shaping SDGs is also evident in the wider category of sex-determination related genes (SRGs), which includes all the genes that are responsible for the specification, development and maintenance of the sexual identity. For example, transcriptionally sex-biased genes often tend to evolve faster than unbiased genes at the level of protein sequences. In particular, male-biased genes generally show higher rate of sequence evolution in comparison to both female-biased and unbiased counterparts, as it has been repeatedly observed in well-studied organisms—such as fruit flies, nematodes, mice and primates (reviewed in **Parsch and Ellegren, 2013; Grath and Parsch, 2016**), but also in other emerging systems, such as the water flea *Daphnia pulex* (**Eads et al., 2007**), aphids (**Purandare et al., 2014**), and two wasp species of the genus *Nasonia* (**Wang et al., 2015**). That said, growing evidence is also showing cases in which female-biased genes have higher rates of sequence evolution than male-biased genes, such as in mosquitoes of the genus *Anopheles* (**Papa et al., 2017**), and European and Manila clams of the genus *Ruditapes* (**Ghiselli et al., 2018**). High rates of molecular evolution in SRGs is particularly evident in organisms with sex chromosomes (SCs)—both in XY/ZW and X0 systems, such as fruit flies, birds and mammals, where the so-called fast-X (or fast-Z) effect has been extensively reported (**Vicoso and Charlesworth, 2006; Mank et al., 2007; Meisel and Connallon, 2013**). In these species, accelerated sequence evolution is seen in general for genes residing on the X (or Z) chromosomes (i.e., the chromosomes determining the homogametic sex) with respect to genes of the autosomal chromosomes, and it could be explained by both adaptive and non-adaptive processes. In fact, the higher ratio of non-synonymous to synonymous mutations (dN/dS , or ω) may result from positive selection, driven either by natural or sexual selection (as in *Drosophila*), as well as form genetic drift (as in birds; **Vicoso and Charlesworth, 2006; Meisel and Connallon, 2013; Parsch and Ellegren, 2013; Grath and Parsch, 2016**).

1.3 Unravelling sex determination in bivalves

Bivalves are the second largest clade in molluscs, counting more than 23,000 species ([Catalogue of Life](#); accessed on 15/10/2024) distributed at all depths and in all marine environments, as well

as in some freshwater habitats. Thanks to their high diversity and biological peculiarities, they have been proposed as promising model organisms for investigating a wide array of biological, ecological and evolutionary issues (**Milani and Ghiselli, 2020; Ghiselli et al., 2021; Nicolini, Ghiselli, et al., 2023**). However, despite the socio-economic and scientific importance, the knowledge concerning the molecular basis of bivalve reproduction and SD is still quite limited (**Breton et al., 2018; Nicolini, Ghiselli, et al., 2023**). Clues from various works seem to suggest that both genetic and environmental factors are involved in SD, though the exact process by which sex is determined and gonad commitment is established is, currently, still unknown.

In the attempt to identify SRGs (including SDGs), and clarify whether a single genetic determinant or a parliamentary decision exist, several differential gene expression (DGE) analyses have been recently performed on a variety of species (e.g., **Milani et al., 2013; Teaniniuraite-moana et al., 2014; Zhang et al., 2014; Chen et al., 2017; Capt et al., 2018; Ghiselli et al., 2018; Shi et al., 2018**). Particularly, some of the genes that were found to be differentially expressed between gonads of different sex were systematically retrieved, such as those belonging to the Dmrt, Sox, and Fox gene (DSFG) families. To this regard, **Zhang et al. (2014)** proposed a working model for the sex-determining pathway of the Pacific oyster *Crassostrea gigas* in which: *Sox-H* promotes male gonad development by activating *Dmrt 1-like* (*Dmrt-1L*), and inhibiting *Fox-L2*; *Fox-L2*, when not inhibited by the pair *Sox-H/Dmrt-1L*, promotes the female gonad development. Additionally, *Fox-L2* has been appointed as the female SDG—following a ZW inheritance system, in *Patinopacten yessoensis* and *Chlamys farreri*, based on the analysis of read coverage and of the distribution of sexually dimorphic single-nucleotide polymorphisms (SNPs; **Han et al., 2022**). However, both the SD model in *C. gigas* and the role of *Fox-L2* as the female SDG in *P. yessoensis* and *C. farreri*, have never been fully tested from a functional point of view (e.g., through gene editing or knock-down), and thus remain only hypothesis. Overall, much of the recent research effort on bivalve SRGs (including DS-FGs) has indeed been limited to their molecular cloning, differential transcription, and tissue localization (**Liang et al., 2019; Sun et al., 2022**), and few works have directly investigated the biological functions so far, mostly through post-transcriptional silencing of target mRNAs (RNA interference [RNAi]). For example, **Liang et al. (2019)** studied the role of *Sox2* in the spermatogenesis of the Zhikong scallop *C. farreri* and found that it likely regulates proliferation of spermatogonia and apoptosis of spermatocytes, since its knockdown resulted in the loss of male germ cells. **Wang et al. (2020)** proposed that in the female gonads of the freshwater

mussel *Hyriopsis cumingii*, *Fox-L2* might be related to the *Wnt/β-catenin* signalling pathway, which takes part in ovarian differentiation also in vertebrates. **Sun et al. (2022)**, instead, found instead that in *C. gigas*, *Fox-L2* and *Dmrt-1L* mRNA knockdown results in the size reduction of female and male mature gonads, respectively. The challenge in identifying SDGs, if they exist, is partly due also to the apparent lack of heteromorphic sex chromosomes (HeSCs) in all the bivalve species investigated to date (**Breton et al., 2018; Han et al., 2022**). In fact, any evidence of SCs has only been found in four scallop species (*Amusium japonicum*, *C. farreri*, *Placopecten magellanicus*, *P. yessoensis*), where they have been described as homomorphic sex chromosomes (HoSCs; **Han et al., 2022**). Though, considering that DSFGs generally work in a coordinated manner to regulate many developmental processes also in other animal species, including the SD cascade itself (see **Section 1.2**), it is reasonable to assume that they play similar roles also in bivalves.

Our understanding of the environmental influences on SD is possibly even more limited. Given that bivalves exhibit a wide array of reproductive strategies—ranging from strict gonochorism to sequential (either protandrous or protogynous) and simultaneous hermaphroditism, as well as the so-called ‘alternative’ and ‘rhythymical sexuality’ (reviewed in **Breton et al., 2018**), they represent an excellent model to investigate the mechanisms of ESD. Temperature, food availability, social factors, and xenobiotics all seem to influence SD, or at least to trigger sex reversal in several hermaphroditic species (mainly belonging to the Ostreida and Pectinida orders). As a matter of fact, ESD has been investigated only in adult individuals through sex-ratio studies, thus few or no experiments are available for the very first round of SD (i.e., that encompassing the first gonad specification cycle). The Pacific oyster *C. gigas*, along with other oyster species, is one of the most studied bivalves not only for GSD (as mentioned above), but also for ESD. It has been shown that the sex ratio of adults is influenced by the incubation temperature of immature spats: at 18 °C, the sex ratio is skewed towards females, while at 28 °C it favours males, with some simultaneous hermaphrodites also observed (**Santerre et al., 2013**). Considering that SD in *C. gigas* may be also under genetic control (**Santerre et al., 2013; Zhang et al., 2014**; reviewed in **Breton et al., 2018**), these observations contribute to the growing evidence that a mixture of different factors govern SD in the species. A similar hypothesis of a mixed SD system has been suggested also for other species, such as *Crassostrea corteziensis* (**Chávez-Villalba et al., 2008**), *Pinctada margaritifera* (**Teaniniuraitemoana et al., 2016**), and *Mytilus edulis* (**Dalpé et al., 2022**), although more insightful and thorough investigations are

needed (**Dalpé et al., 2022**).

Clearly, bivalves represent a dazzling example of how the traditional representation of sex as genetically- or environmentally-determined, as well as the distinction between SD and SDf, can no longer be assumed as strictly dichotomous. A multifactorial model, in which many genes and environmental cues act in concert to establish the sexual identity of the individual, seems to better explain the extreme diversity of bivalve SD systems (**Breton et al., 2018**). Nonetheless, much work still needs to be done, especially in the functional characterisation of the molecular ground plan. Functional assays employing RNAi and clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (CRISPR-Cas9) techniques (e.g., **Wang et al., 2020; Sun et al., 2022; Wang et al., 2022**) are finally making their way into the study of bivalve biology and have been proved essential instruments also for the investigation of sex-related traits. However, very few works have made extensive use of the comparative and integrative approach in bivalve studies so far, which hampers the possibility to infer general patterns for such a vast and diverse class of organisms (**Milani and Ghiselli, 2020**).

Chapter 2

Project outlook and objectives

This PhD project focuses on understanding the evolutionary dynamics and molecular mechanisms underlying sex determination (SD) in bivalve molluscs. The research has leveraged a wide array of analytical tools—from comparative genomics, to transcriptomics, *in-situ* hybridization, and immunolocalization, in order to investigate sex-determination related genes (SRGs) across various species with an integrative and comparative approach. Particularly, special attention was given to the Dmrt, Sox, and Fox gene (DSFG) families, which are widely-recognised key actors in the SD process of the majority of animal species, including bivalves. Each major area of analysis in my research, together with its objectives, is presented in a dedicated chapter, resulting in three distinct sections.

Chapter 3, which consists of a perspective piece published in *Genome Biology and Evolution*, will examine bivalves as emerging model organisms in SD research, by reviewing their genomic and biological characteristics. Bivalve offers valuable insights into several topics, including (i) the transitions between environmental and genetic SD, (ii) the evolution of sex chromosomes (SCs), (iii) the tentatively interaction between mitochondrial inheritance and SD, and (iv) the evolutionary history of SRGs. Particularly, this chapter wants to emphasise the importance of establishing a comprehensive evolutionary genomics framework for studying SD across bivalve species.

Chapter 4 will explore the molecular evolution of some key SRGs. Using a broad genomic context that includes more than 40 annotated bivalve genomes and transcriptomes, this chapter aims to uncover how these genes have evolved and their potential roles in SD, by also adopting a cross-species validation assay. The analysis will focus on the evolution of the DSFG families, by using the tools of molecular evolution to assess whether some of them are tightly involved

in SD. Mammals and *Drosophila* spp. will be used as positive and negative control datasets, respectively, to validate the reliability of the approach.

Chapter 5 will focus on the expression patterns of three SRGs in the Mediterranean mussel *Mytilus galloprovincialis* during early developmental stages. Particularly, the spatial and temporal transcription patterns of *Dmrt 1-like* (*Dmrt-1L*), *Sox-H*, and *Fox-L2*—which have been identified as tightly linked to primary SD by analyses in **Chapter 4** and in previous works, will be investigated. By also including the analysis of the germline marker *Vasa/Vasa*, this chapter will provide novel insights into the mechanisms of SD and primordial germ cell (PGC) specification. Transcription patterns will be investigated through computational differential gene expression (DGE) analyses and mRNA *in-situ* hybridization chain reaction (HCR); the expression pattern of *Vasa/Vasa* will be investigated also through immunolocalization.

Overall, this PhD project aims to adopt a multi-layered and integrative approach that combines evolutionary genomics, gene expression analyses, and comparative biology to explore SD in bivalves. Bivalves represent a relatively underexplored group, and given the remarkable diversity of their SD processes, require a strong evolutionary perspective to decipher the mechanism. Here, the integration of genome-wide molecular evolution analysis with gene expression studies provides a novel framework for understanding how SRGs, such as those belonging to the DSFG families, contribute to SD and sexual differentiation. This work also benefits from cross-species comparisons, which places bivalve SD within a broader evolutionary context, allowing for the identification of commonalities and unique traits in sex-determining pathways across taxa. Moreover, by investigating the expression patterns of three SRGs during early development in *M. galloprovincialis*, this project addresses a critical gap in the understanding of how these genes may regulate the sexual process, as to date bivalve SD has been investigated mostly in adult life stages. Through a comprehensive and comparative methodology, the project promises to provide a first reference broad-scale evolutionary resource for bivalve SD, also pushing forward the boundaries of reproductive and evolutionary biology in non-model species.

Chapter 3

Bivalves as emerging model systems to study the mechanisms and evolution of sex determination: a genomic point of view

Filippo Nicolini^{1,2}, Fabrizio Ghiselli¹, Andrea Luchetti¹, Liliana Milani¹

¹*Department of Biological, Geological and Environmental Science, University of Bologna, Bologna (BO), Italy.*

²*Fano Marine Center, Fano (PU), Italy.*

Published in: 2023, *Genome Biology and Evolution*, 15(10):evad181. doi: [10.1093/gbe/evad181](https://doi.org/10.1093/gbe/evad181)

Abstract. Bivalves are a diverse group of molluscs that have recently attained a central role in plenty of biological research fields, thanks to their peculiar life history traits. Here we propose that bivalves should be considered as emerging model systems also in sex-determination studies, since they would allow to investigate: (i) the transition between environmental and genetic sex determination, with respect to different reproductive backgrounds and sexual systems (from species with strict gonochorism to species with various forms of hermaphroditism); (ii) the genomic evolution of sex chromosomes, considering that no heteromorphic sex chromosomes are currently known and that homomorphic sex chromosomes have been identified just in few species of scallops; (iii) the putative role of mitochondria at some level of the sex determination signaling pathway, in a mechanism that may resemble the cytoplasmatic male sterility of plants; (iv) the evolutionary history of sex-determination related gene families with respect to other

animal groups. In particular, we think that this last topic may lay the foundations for expanding our understanding of bivalve sex determination, as our current knowledge is quite fragmented and limited to few species. As a matter of fact, tracing the phylogenetic history and diversity of sex-determination related gene families (such as the Dmrt, Sox and Fox genes) would allow to perform more targeted functional experiments and genomic analyses, but also fostering the possibility of establishing a solid comparative framework.

Significance. In this perspective, we provide an examination of the phylogenetic diversity of Dmrt genes, a sex-determination related gene family, to address the importance of bivalves in sex determination studies. By analyzing their taxonomic distribution and sequence diversity, we show how such a comparative study may set a common ground plan to settle down targeted functional experiments and essays. This kind of approach should be applied more extensively in future studies, especially when dealing with understudied organisms.

Bivalves are the second largest clade in molluscs, counting more than 18,000 species ([Catalogue of Life](#), accessed on 16/12/2022) distributed at all depths and in all marine environments, as well as in some freshwater habitats. Thanks to their high diversity and peculiar biological features, they have been proposed as promising model organisms for investigating a wide array of biological, ecological, and evolutionary issues, from mitochondrial biology and evolution to the physiological plasticity under fluctuating environmental conditions (**Milani and Ghiselli, 2020; Ghiselli et al., 2021**). In this context, bivalves may serve as a compelling model system to investigate the evolution and characteristics of sex determination (SD) as well, thanks to the diversity of their reproductive modes and genomic features. Nonetheless, this research field has been largely overlooked and many aspects of bivalve reproductive biology remain uncharacterized. In this perspective, we address the topic by first examining the relevant questions that bivalves may help to answer regarding processes and patterns of SD, and then providing a case study in the field of comparative genomics.

3.1 Open yet inspiring topics in bivalve sex determination

Despite the socio-economic and scientific importance of bivalves, the knowledge concerning the genetic and molecular bases of their SD system is quite limited and its study has been mostly neglected. Yet, bivalves may constitute a novel model system in SD studies that is as intriguing and valuable as other well-established models, such as vertebrates, insects and plants (**of Sex Consortium, 2014**), as they may provide complementary perspectives in many aspects of SD evolutionary studies. Topics such as (i) the transition between environmental and genetic SD, (ii) the evolution of sex chromosomes, (iii) the mito-nuclear interaction, and (iv) the evolution of SD related genes, can largely benefit from the integration with bivalve studies. But many others are likely to emerge as research in the field progresses.

3.1.1 Transitions between environmental and genetic sex determination

Clues from several works seem to suggest that both genetic and environmental factors are involved in bivalve SD, thus implying that a mixed system may exist (reviewed in **Breton et al., 2018**). The traditional dichotomy between environmental sex determination (ESD) and genetic sex determination (GSD) seems inapplicable in most bivalve species, where ESD and GSD rather represent the two ends of a continuum of mixed and plastic conditions. A weak

distinction between ESD and GSD is also found in amphibians, reptiles and teleost fish, three clades in which environment-dependent SD has been largely studied. Here, the interaction—or even the transition—between the two sexual systems have been reported in many species, suggesting that sex-determining mechanisms can be extraordinary plastic (**Bachtrog et al., 2014; Capel, 2017**). Adding a representative and diverse group of Lophotrochozoa (Protostomia) to those vertebrate taxa, can widely expand the comparative framework of the investigation, allowing to better understand the evolution of SD as a whole. In bivalves, ESD has been studied mostly in oysters, where hermaphroditic species show an effect of temperature on SD (reviewed in **Breton et al., 2018; Fig. 3.1**). Oysters may indeed constitute a prolific model to examine how the SD pathways are shaped in the presence of different initial triggers and highly dynamic reproductive backgrounds. In fact, various sexual systems can be found in oysters, such as (i) strictly gonochoric population, (ii) the coexistence of simultaneous hermaphroditic with strictly gonochoric individuals in the same population, (iii) the possibility of sex change according to environmental conditions, and (iv) the presence of both parasitic dwarf males and free-living males in the same species (**Collin, 2013**). Consequently, oysters may be extremely useful to understand how epigenetic control is involved in sex change, how gene regulatory networks can sustain the occurrence of different hermaphroditic conditions within gonochoric populations, and whether certain SD systems are more labile than others (**Abbott, 2011**).

3.1.2 Evolution of sex chromosomes

So far, heteromorphic sex chromosomes (HeSCs)—i.e., sex chromosomes showing strong morphological differentiation, have never been observed in bivalves (**Breton et al., 2018**), while the first evidence of homomorphic sex chromosomes (HoSCs)—i.e., sex chromosomes showing little or no differentiation, comes from a very recent study on several scallop species, where a non-homologous origin of the SD system has been proposed for different subfamilies (**Han et al., 2022; Fig. 3.1**). Theory predicts that, once originated, sex chromosomes (SCs) will eventually turn into HeSCs, because of the recombination arrest in the sex-determining region (**Bachtrog et al., 2014; Beukeboom and Perrin, 2014; Han et al., 2022**). Nonetheless, HoSCs are much more widespread in the animal kingdom than expected, sometimes also being of ancient age (**Bachtrog et al., 2014; Han et al., 2022**).

Species from the order Pectinida may thus be useful to investigate what determines the

long-term maintenance of HoSCs and which genomic architectures and molecular dynamics prevent HeSCs from evolving in bivalves. Additionally, they may be taken as model systems to investigate the origin of SCs in relation to the sexual systems and the route by which molecular pathways have been reprogrammed in the transition between different SD mechanisms (**Han et al., 2022**).

Researchers have been addressing this topic mainly in snakes, ratites and sturgeons (**Bachtrog et al., 2014; Han et al., 2022** and references therein). Though, scallops currently hold the oldest HoSC pairs, which are dated back to about 350 million years. The system is thus of great importance to investigate the role of sex-biased gene expression and selection forces in the long-term stability of SCs (**Han et al., 2022**), as well as the intertwining between SD systems.

3.1.3 Mito-nuclear interactions

An additional pivotal topic in bivalve biology, tentatively connected to SD, regards the doubly uniparental inheritance (DUI) of mitochondria, a process in which two highly divergent mitochondrial genomes are transmitted uniparentally through the maternal and paternal lineages, respectively through eggs and sperm. This process, which has been reported in more than a hundred bivalve species from five different orders (**Fig. 3.1; Gusman et al., 2016; Capt et al., 2020**), has been proposed to interact with the major nuclear pathways that primarily establish the sexual identity, in a way that can resemble the cytoplasmatic male sterility (CMS) of plants (**Ghiselli et al., 2013; Breton et al., 2022**). In CMS, specific mitochondrial chimeric open reading frames (ORFs) cause the pollen to be sterile, while certain nuclear loci act in counterbalance to restore male fertility when occurring in the same individual. This Red-Queen scenario, in which balancing selection shapes the evolution of both CMS and restorer-of-fertility genes and keeps the two sexes viable, has been also hypothesized to be acting on bivalve DUI species (**Ghiselli et al., 2013; Xu, Iannello, et al., 2022**), where additional and effectively-transcribed ORFs have been observed in both the male-inherited and female-inherited mitochondrial lineages (**Milani et al., 2013, 2014**).

Clearly, if a functional interplay between DUI and SD in bivalves is proven, this will provide new research questions regarding not only bivalve biology itself but also broader evolutionary topics (e.g., are there any converging trait between DUI and CMS systems? What is the degree of plasticity of such mitochondria-related SD systems? Are mitochondria-related SD systems

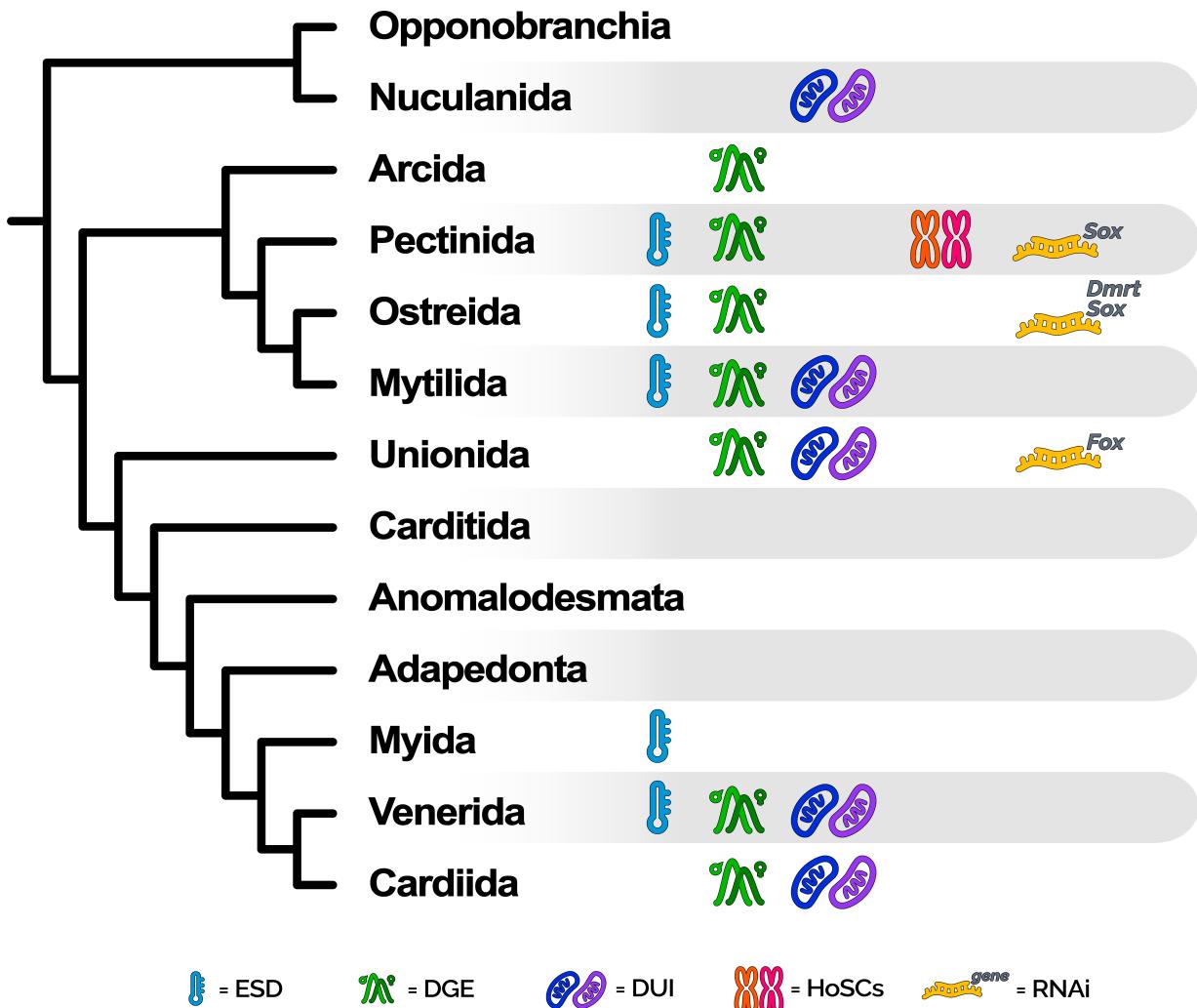


Figure 3.1 – Graphical summary of the available knowledge and experiments concerning the genetic basis of SD in bivalves, at the level of major taxonomic orders (as reported in WoRMS; accessed before or on 14/03/2023). For each bivalve clade it is reported: (i) the availability of records of ESD; (ii) the availability of differential gene expression (DGE) experiments specifically intended to investigate sex-biased or sex-specific genes; (iii) whether the DUI of mitochondria has been reported in at least one species; (iv) whether HoSCs have been identified in at least one species; (v) the availability of RNA interference (RNAi) experiments for genes belonging to the Dmrt, Sox, and Fox gene (DSFG) families. The phylogenetic tree on the left has been drawn on the basis of the most widely accepted topology for bivalves, according to analyses based on nuclear markers and morphological data. The tips of the tree correspond to major bivalve orders, except for Opponobranchia and Anomalodesmata, which represent higher-level taxonomic ranks. References for the availability of data and experiments can be found throughout the main text.

more widespread in eukaryotes than currently thought?).

3.1.4 Evolution of sex-determination related genes

Considering this intricate scenario of SD mechanisms and the wide diversity of bivalves, in the last years many differential transcription analyses have been performed on several species with the attempt to identify the most probable sex-determination related genes (SRGs; e.g., **Milani et al., 2013; Zhang et al., 2014; Chen et al., 2017; Capt et al., 2018; Shi et al., 2018; Fig. 3.1**). Interestingly, certain genes consistently emerged across different bivalve species as being substantially more transcribed in one sex (sex-biased) or exclusively transcribed in one sex (sex-specific), suggesting their potential involvement in the SD pathway. These genes mainly belong to the *dsx* and *mab-3* related transcription factor (Dmrt), *Sry*-related HMG-box (Sox), and forkhead box (Fox) families, which play a role in various developmental processes (including the SD cascade) in most animals (**Marshall Graves and Peichel, 2010; Bachtrog et al., 2014; Beukeboom and Perrin, 2014**). Members of these three gene families are also included in the working model for the SD regulatory network proposed for the Pacific oyster *Crassostrea gigas* by **Zhang et al. (2014)**, in which: *CgSoxH* (which belong to the Sox family) promotes male gonad development by activating *CgDsx* (also known as *Dmrt 1-like [Dmrt-1L]*; which belong to the Dmrt family) and inhibiting *CgFoxL2* (which belong to the Fox family); *CgFoxL2*, when not inhibited by the pair *CgSoxH/CgDsx*, promotes female gonad development. Similarly, **Han et al. (2022)** appointed *FoxL2* as a putative SD gene in the two scallop species *Patinopacten yessoensis* and *Chlamys farreri*. If their pivotal role in SD of bivalves is confirmed, an evolutionary genomic analysis may help in better understanding why members of the above-mentioned gene families appear particularly prone to be recruited in the SD cascade also in distantly related species, as it is observed for *Dmrt1* and *Sox3* homologs in vertebrates (**Marshall Graves and Peichel, 2010; Bachtrog et al., 2014**; and the following section). Furthermore, considering the occurrence of mixed SD systems in bivalves, DSFGs may provide new perspectives on the influence of different environmental cues on the molecular evolution of animal SRGs. However, to date, experiments have been limited to molecular cloning, differential transcription, and tissue localization of such genes (**Liang et al., 2019; Sun et al., 2022**), while only a few have directly investigated their biological functions in bivalves, for example through post-transcriptional silencing of target mRNAs (RNAi; **Fig. 3.1**; e.g., **Liang et al., 2019; Wang et al., 2020; Sun et al., 2022**.

Overall, DSFGs are highly interesting targets to be investigated in the framework of bivalve SD and have indeed obtained much more attention than the study of SCs or the role of environmental cues. However, much work is still to be done in order to understand their function in the SD signaling pathway and their evolutionary history.

3.2 The case of the Dmrt gene family in bivalves

Among the SRG candidates identified in bivalves, Dmrt genes (named after *doublesex* [*dsx*] from *Drosophila melanogaster* and *male abnormal-3* [*mab-3*] from *Caenorhabditis elegans*) are of particular interest. As a matter of fact, in vertebrates, besides their role in placode neurogenesis and somite patterning (reviewed in **Mawaribuchi et al., 2019**), Dmrt genes are also involved in the development of male gonads and the maintenance of the testicular function (**Sun et al., 2022**). Their role in the specification and organization of male sexual characters seems indeed to be common across Metazoa, suggesting that a similar function may have been already present in the Bilateria common ancestor (**Kopp, 2012; Beukeboom and Perrin, 2014**).

The first attempts to dig inside the phylogenetic history and diversity of bivalve Dmrt genes have been provided by **Li et al. (2018)** and **Evensen et al. (2022)**: besides retrieving all the canonical genes (i.e., *Dmrt2*, *Dmrt3* and *Dmrt4/5*), their inferences brought to light a monophyletic Dmrt group (named *Dmrt-1L*) which appears to be private to molluscs and present in several bivalve species. The *Dmrt-1L* monophyletic group is confirmed also when expanding the analysis by mining genomes from a wider range of bivalve taxa (**Tab. 3.1; Fig. 3.2A**), suggesting that *Dmrt-1L* genes are widespread in bivalves and were likely present in their common ancestor (**Evensen et al., 2022**). In particular, *Dmrt-1L* genes can be successfully retrieved in species of the orders Mytilida, Ostreida, Pectinida, Unionida, and from *Scapharca broughtonii* (Arcida), while the opposite holds for Venerida, *Sinonovacula constricta* (Adapedonta), and *Dreissena* spp. (Myida; **Fig. 3.2B**). Clearly, the absence of *Dmrt-1L* genes demands further investigations, as it may derive from errors in genome assembly and annotations.

The present analysis also supports a higher amino acid sequence divergence of the *Dmrt-1L* orthology group with respect to the other Dmrt orthology groups (**Fig. 3.2C**), which may be explained by a higher rate of sequence evolution related to their sex-biased expression in certain species (**Zhang et al., 2014; Shi et al., 2015; Li, Zhang, et al., 2018; Evensen et al.,**

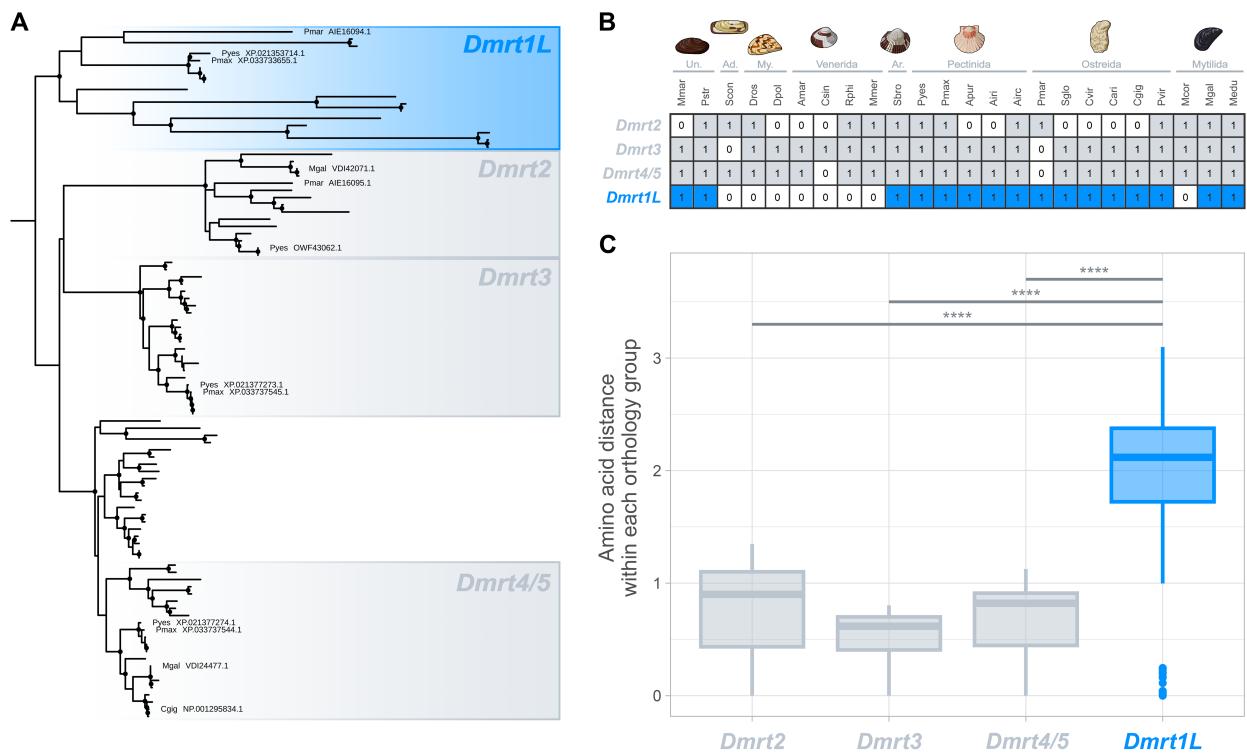


Figure 3.2 – Phylogenetic tree (A) and taxonomic distribution (B) of Dmrt genes in bivalves, and comparison of amino acid pairwise distances within *Dmrt-1L* and the other Dmrts (C). (A) Dmrt orthologs from bivalve genome assemblies were obtained with HMMsearch (HMMER toolkit; **Eddy, 2011**) with the Pfam HMM profile of the DM domain (PF00751). Amino acid alignment was obtained with MAFFT-DASH (**Rozewicki et al., 2019**), and manually inspected to remove poorly aligning sequences, and trimmed with trimAI (gap threshold of 60 %; **Capella-Gutiérrez et al., 2009**). The phylogenetic analysis was carried out using IQ-TREE 2 (**Minh et al., 2020**) with default parameters. Nodes with bootstrap values greater than 84 are marked with filled black circles. The tree was rooted according to **Evensen et al., 2022**. Dmrt genes analysed by **Evensen et al., 2022** were used as reference to annotate the various orthology groups, and accession numbers are reported in the tree. The phylogenetic tree with all annotated tips and nodes can be accessed on supplementary material online. (B) Taxonomic distribution of identified Dmrt genes in bivalve genomes. Orders as reported in WoRMS (accessed before or on 14/03/2023) and in **Fig. 3.1** are specified. (C) Pairwise amino acid distances were computed for amino acid sequences within each Dmrt orthology group identified in the tree, with the R package ‘phangorn’ (**Schliep, 2011**) under the JTT substitution model. After checking for normality with the Shapiro-Wilk test ($W = 0.88544$, $p < 2.2 \times 10^{-16}$) and for group effect with the Kruskal-Wallis test ($p < 2.2 \times 10^{-16}$), the pairwise Wilcoxon rank-sum test was used to compare the distributions of pairwise amino acid distances of *Dmrt-1L* and the other Dmrts. Horizontal bars mark the significative results with $p < 2.2 \times 10^{-16}$ (****; Bonferroni correction for multiple test was applied). The list of genome assemblies used for these analyses and species identifiers can be found in **Tab. 3.1**. Un.: Unionida; Ad.: Adapedonta; My.: Myida; Ar.: Arcida.

2022). This is consistent with what has been already observed for the SRGs *Dmrt1* and *dsx* in vertebrates and *Drosophila*, respectively (e.g., **Bewick et al., 2011; Baral et al., 2019**). In fact, sex-biased genes (including SRGs) often tend to evolve faster than unbiased genes at the level of protein sequences, either when considering male-biased (reviewed in **Parsch and Ellegren, 2013; Grath and Parsch, 2016**) or female-biased genes (e.g., **Papa et al., 2017; Ghiselli et al., 2018**). Another possible explanation for the higher amino acid divergence of *Dmrt-1L* genes may lie on their expression breadth, that is, genes with a narrow tissue-specific expression tend to evolve faster than more ubiquitous genes (**Parsch and Ellegren, 2013; Xu, Martelossi, et al., 2022**). As a matter of fact, *Dmrt-1L* genes have been found to be significantly more transcribed in the gonadic tissue (particularly in testes) in *P. yessoensis* (**Li, Zhang, et al., 2018**) and *C. gigas* (**Yue et al., 2021**).

Understanding the role and molecular interactions of *Dmrt-1L* genes in bivalve SD and gonad development would greatly enhance the possibility of outlining the evolutionary causes and consequences of their high amino acid divergence (**Fig. 3.2C**), for example by linking the molecular evolution to the degree of pleiotropy. However, most of our knowledge on *Dmrt-1L* biology is currently limited to the temporal and tissue localization of transcripts in a few species of bivalves (e.g., **Li, Zhang, et al., 2018; Yue et al., 2021**). In fact—apart from the work by **Sun et al. (2022)**, which confirmed the role of *Dmrt-1L* in the gonad development of *C. gigas* through non-invasive RNAi and found that the knocked-down phenotype results in size reduction of male gonads—no other experiments intended to elucidate the function of *Dmrt-1L* genes in bivalves have been carried out so far (**Fig. 3.1**). This clearly hinders any possible integration between molecular data with functional assays. If the role of *Dmrt-1L* as major sex determinants was confirmed, bivalves would become an intriguing clade in which investigate why, in Metazoa, certain genes (namely, the Dmrt gene family) appear particularly prone to being recruited at the top of the SD cascade. To date, this phenomenon has been widely examined in vertebrates, where *Dmrt1* genes have independently gained a primary role in male SD in fish, amphibians, and birds, and are considered candidate sex-determining genes also in monotreme mammals (**Marshall Graves and Peichel, 2010; Beukeboom and Perrin, 2014; Mawaribuchi et al., 2019**). Bivalves may provide an alternative evolutionary scenario to study the selective forces and molecular modifications that support Dmrt genes in repeatedly taking over the SD process. In fact, since *Dmrt-1L* genes seem to be restricted to molluscs (**Fig. 3.2A**), it would be intriguing to clarify if the putative involvement in the SD cascade of

extant bivalve species is the result of shared ancestry or convergent evolution, which would establish a study system for the evolution of Dmrt genes parallel to that of vertebrates (see **Capel, 2017**).

Obviously, *Dmrt-1L* should not be expected to be the sole sex-determining gene (SDG). In fact, *Fox-L2* has already been appointed as the female SDG in *P. yessoensis* and *C. farreri* (**Han et al., 2022**). Consequently, we should expect that other primary genetic determinants exist, consistently with the extremely high species diversity of the clade. Thus, bivalves may additionally serve as a valuable model system to study how genes from different families take over the SD cascade and are shaped by selection.

3.3 Conclusions: bivalves as new models in the study of sex determination

SD is undoubtedly a fascinating biological and evolutionary topic as much as it is challenging to investigate. Our understanding of the causes and consequences of the SD mechanism diversity strongly relies on the study of different systems and non-model model organisms (**Bachtrög et al., 2014; Milani and Ghiselli, 2020**), which provide the foundation for depicting a comprehensive evolutionary and comparative framework in which new and coherent research perspectives can be grounded.

In recent years, bivalves have been achieving growing importance in many fields of biology, from ecology to genomics, and from environmental biomonitoring to mitochondrial studies (**Milani and Ghiselli, 2020; Ghiselli et al., 2021**), but they can be a valuable model to address also SD studies. The diversity of their life history traits provides indeed a challenging, yet extremely fascinating framework, to put the SD processes into an evolutionary context.

Bivalves can help us explain how ESD and GSD interplay with each other in response to the environmental conditions, as a mixed system of both has been proposed to act in the establishment of bivalve sexual identity (reviewed in **Breton et al., 2018**). Moreover, the occurrence of the many existing variants of hermaphroditism and gonochorism even in closely related species, or within the same population, strongly suggests that the basic SD pathway (whether genetic, environmental, or mixed) should be plastic enough to sustain the existence of individuals of both sexes, thus providing the opportunity to study how SD gene regulatory

networks are shaped and selected throughout evolution and how epigenetic regulation may influence SD. The unique DUI system further poses an undeniable challenge in SD studies since it may represent an SD-linked mechanism which relies on the non-nuclear portion of the genome and may unfold many new research paths (**Milani and Ghiselli, 2020; Ghiselli et al., 2021**). Nonetheless, much of the research effort on bivalve SD has been devolved to specific groups of socio-economic importance, such as Mytilida, Ostreida, Pectinida, and Unionida, while the other lineages of the bivalve phylogeny have been neglected (**Fig. 3.1**). Our understanding of the SD processes of bivalves is thus restricted and is mainly lacking a broad comparative framework in which to draw comprehensive evolutionary inferences.

Genes from the DSFG families, which are involved in SD also in other Metazoa, may be considered excellent genomic targets to study the processes and patterns of molecular evolution in sex-biased genes, as well as of the recurrent recruitment of genes in the SD cascade. Also, identifying the major genetic regulators of SD in bivalves would burst the functional study of the interaction between ESD and GSD, by providing genetic targets that can be manipulated through RNAi and/or genome editing techniques to understand the role of environmental cues in SD. In the same way, knowing the main genetic actors of SD would allow researcher to identify SCs not only on the basis of *in-silico* techniques (such as k-mer based or single nucleotide polymorphism [SNP] methods) but also by less-expensive wet lab protocols (such as fluorescence mRNA *in-situ* hybridization [mRNA-ISH] on metaphase chromosome plates). Furthermore, it would help to understand whether and how the mitochondrial additional ORFs of DUI species interact with the SD system, by performing thorough gene expression essays.

In conclusion, we strongly urge researchers to invest more resources in the integrative study of bivalve SD to unravel the many underlying mechanisms and expand our understanding of this biological process. Given our limited knowledge in the field, one of the first routes that should be undertaken may rely on the comparative study of SRGs of bivalves from a genomic perspective, as this kind of data is nowadays growing at a rate faster than ever. Establishing such a genomic ground plan for understudied organisms will in fact allow researchers to develop evolutionary-aware experiments with better selected genetic targets.

Table 3.1 – List of bivalve genomes from which Dmrt genes have been extracted. For each species, the accepted name and the most-common synonym (in parentheses) are reported. NCBI accession numbers are provided, when available, as well as BUSCO scores of the predicted proteomes against the ‘metazoa_odb10’ dataset (Manni et al., 2021).

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<i>Anadara (Scapharca) broughtonii</i>	Sbro	Arcida	Chromosome	C:91.2% [S:85.6%,D:5.6%] F:2.6% M:6.2%	Bai et al., 2019	–
<i>Simonovacula constricta</i>	Scon	Adapedonta	Chromosome	C:92.5% [S:80.4%,D:12.1%] F:3.4% M:4.1%	Ran et al., 2019	GCA_007844125.1
<i>Dreissena polymorpha</i>	Dpol	Myida	Chromosome	C:86.9% [S:75.1%,D:11.8%] F:6.4% M:6.7%	McCartney et al., 2022	GCA_020536995.1
<i>Dreissena rostriformis</i>	Dros	Myida	Scaffold	C:75.2% [S:73.2%,D:2.0%] F:15.2% M:9.6%	Calcino et al., 2019	GCA_007657795.1
<i>Mytilus unguiculatus (coruscus)</i>	Mcor	Mytilida	Chromosome	C:80.0% [S:79.1%,D:0.9%] F:7.7% M:12.3%	Yang et al., 2021	GCA_017311375.1
<i>Mytilus edulis</i>	Medu	Mytilida	Scaffold	C:83.7% [S:64.5%,D:19.2%] F:5.2% M:11.1%	Corrochano-Fraile et al., 2022	GCA_017397895.1
<i>Mytilus galloprovincialis</i>	Mgal	Mytilida	Scaffold	C:80.3% [S:47.5%,D:32.8%] F:8.8% M:10.9%	Gerdol et al., 2020	GCA_900618805.1

Tab. 3.1 continued from previous page

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<i>Perna viridis</i>	Pvir	Mytilida	Scaffold	C:99.4% [S:99.0%,D:0.4%] F:0.2% M:0.4%	Inoue et al., 2021	GCA_018327765.1
<i>Magallana (Crassostrea) ariakensis</i>	Cari	Ostreida	Chromosome	C:94.6% [S:90.9%,D:3.7%] F:0.9% M:4.5%	Li et al., 2021	GCA_020567875.1
<i>Magallana (Crassostrea) gigas</i>	Cgig	Ostreida	Chromosome	C:98.5% [S:67.6%,D:30.9%] F:0.3% M:1.2%	Penaloza et al., 2021	GCF_902806645.1
<i>Crassostrea virginica</i>	Cvir	Ostreida	Chromosome	C:98.1% [S:58.6%,D:39.5%] F:0.3% M:1.6%	Gómez-Chiarri et al., 2015	GCF_002022765.2
<i>Saccostrea glomerata</i>	Sglo	Ostreida	Scaffold	C:88.9% [S:85.3%,D:3.6%] F:5.1% M:6.0%	Powell et al., 2018	GCA_003671525.1
<i>Argopecten irradians concentricus</i>	Airc	Pectinida	Scaffold	C:94.8% [S:93.9%,D:0.9%] F:3.7% M:1.5%	Liu et al., 2020	GCA_004382765.1
<i>Argopecten irradians irradians</i>	Airi	Pectinida	Scaffold	C:94.8% [S:93.9%,D:0.9%] F:3.7% M:1.5%	Liu et al., 2020	GCA_004382745.1
<i>Argopecten purpuratus</i>	Apur	Pectinida	Scaffold	C:89.2% [S:88.5%,D:0.7%] F:5.0% M:5.8%	Liu et al., 2020	–

Tab. 3.1 continued from previous page

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<i>Pecten maximus</i>	Pmax	Pectinida	Chromosome	C:98.5% [S:74.7%,D:23.8%] F:0.4% M:1.1%	Kenny et al., 2020	GCF_902652985.1
<i>Mizuhopecten (Patinopecten) yessoensis</i>	Pyes	Pectinida	Scaffold	C:98.6% [S:75.2%,D:23.4%] F:0.4% M:1.0%	Wang, Zhang, et al., 2017	GCF_002113885.1
<i>Margaritifera marginifera</i>	Mmar	Unionida	Scaffold	C:92.6% [S:82.3%,D:10.3%] F:3.2% M:4.2%	Gomes-dos-Santos et al., 2021	GCA_015947965.1
<i>Potamilus streckersoni</i>	Pstr	Unionida	Scaffold	C:74.7% [S:73.8%,D:0.9%] F:7.0% M:18.3%	Smith, 2021	GCA_016746295.1
<i>Calyptogena (Archivesica) marissinica</i>	Amar	Venerida	Chromosome	C:82.0% [S:80.0%,D:2.0%] F:6.1% M:11.9%	Ip et al., 2021	GCA_014843695.1
<i>Cyclina sinensis</i>	Csin	Venerida	Scaffold	C:94.0% [S:83.8%,D:10.2%] F:1.9% M:4.1%	Wei et al., 2020	GCA_012932295.1
<i>Mercenaria mercenaria</i>	Mmer	Venerida	Chromosome	C:95.4% [S:70.9%,D:24.5%] F:0.5% M:4.1%	Song et al., 2021	GCF_014805675.1
<i>Ruditapes philippinarum</i>	Rphi	Venerida	Chromosome	C:83.4% [S:74.5%,D:8.9%] F:8.8% M:7.8%	Xu, Martelossi, et al., 2022	GCA_026571515.1

3.4 Acknowledgments

The authors are extremely thankful to Sofía Blanco González from the University of Vigo for her willingness to engage in discussions and for genuinely sharing her opinion on this work.

3.5 Data Availability

Analyzed data and R scripts used to generate plots can be accessed in supplementary material online deposited at the following GitHub repository: [filonico/bivalve_sex_perspective](https://github.com/filonico/bivalve_sex_perspective).

Chapter 4

Identification of putative sex-determination related genes in bivalves through comparative molecular evolutionary analyses

Filippo Nicolini^{1,2}, Mariangela Iannello¹, Giovanni Piccinini¹, Sergey Nuzhdin³, Fabrizio Ghiselli¹, Andrea Luchetti¹, Liliana Milani¹

¹*Department of Biological, Geological and Environmental Science, University of Bologna, Bologna (BO), Italy.*

²*Fano Marine Center, Fano (PU), Italy.*

³*Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, USA.*

Manuscript in preparation.

4.1 Introduction

In sexually reproducing organisms, the modes of sex determination (SD), i.e., the process by which the male or female identity of an organism (or of the gonadic tissue) is established, is highly diverse, ranging from strictly genetic systems to environmentally-dependent processes (**Haag and Doty, 2005; Uller and Helanterä, 2011; Bachtrog et al., 2014; Beukeboom and Perrin, 2014**). Characterising the molecular basis of SD is crucial for understanding not only reproductive biology but also the evolutionary pressures shaping these systems (**Wilkins, 1995; Ellegren and Parsch, 2007; Grath and Parsch, 2016; Nicolini, Ghiselli, et al., 2023**), as sex-determination related genes (SRGs), including primary sex-determining genes (SDGs), are those responsible for the phenotypic differences of males and females, thanks to their sex-biased expression and interactions (**Ellegren and Parsch, 2007; Beukeboom and Perrin, 2014; Grath and Parsch, 2016**). One key aspect of SRGs is that they often exhibit accelerated rates of sequence evolution, due to their involvement in sex-related traits and reproduction. This represents the effects of sexual and/or adaptive selection, which act in sex-biased genes and produce high-divergent proteins at the interspecific level (**Civetta and Singh, 1998; Ellegren and Parsch, 2007; Meisel, 2011; Grath and Parsch, 2016**). Rapid sequence evolution is known for *Sex-determining region of chromosome Y (Sry)* of eutherians (**Pamilo and O’Neill, 1997; Mawaribuchi et al., 2012**), *dsx and mab-3 related gene W (Dm-W)* of the African clawed frog *Xenopus laevis*, and *dsx and mab-3 related gene Y (Dmy)* of the medaka fish *Oryzias latipes* (**Mawaribuchi et al., 2012**), all of which are master SDGs, that is, genes whose expression is primarily responsible for the establishment of the sexual fate of the organism. Evolution under episodic diversifying selection has been detected also in *Drosophila* for genes involved in the SD cascade (e.g., *Sex-lethal [Sxl]*, *transformer [tra]*, and *doublesex [dsx]*), in correspondence with its establishment in the genus common ancestor (**Mullon et al., 2012; Baral et al., 2019**); though, rapid sequence evolution seems to not be concerning extant amino acid sequences (**Haerty et al., 2007; Baral et al., 2019**), as they are globally evolving under purifying selection, especially in their catalytic domain (**Mullon et al., 2012; Baral et al., 2019**). Concerning the *dsx* genes, higher rates of nucleotide and amino acid sequence evolution can be however observed for male-specific regions, if compared to female-specific and oligomerization regions (**Baral et al., 2019**).

While SD has been extensively studied in model organisms, like mammals, insects, and

nematodes, comparatively little is known about the molecular ground plans in non-model organisms. A remarkable example of this is represented by bivalve molluscs, which exhibit a wide variety of reproductive strategies and sexual systems (**Breton et al., 2018**). Notwithstanding the considerable importance in the human socio-economic landscape (reviewed in **Haszprunar and Wanninger, 2012**; **Gomes-dos-Santos et al., 2020**), the study of SD mechanisms in bivalves has been hampered by the striking divergence among species (**Li et al., 2022**), and thus largely overlooked and limited to few case studies (**Breton et al., 2018; Nicolini, Ghiselli, et al., 2023**). So far, no master SDG has been unambiguously identified, and the only working hypothesis on the functioning of the SD gene regulatory network is available for the Pacific oyster *Crassostrea gigas* (now *Magallana gigas*; **Zhang et al., 2014**). Nonetheless, the field still lacks both a robust functional investigation and an evolutionary framework in which to place the current knowledge (**Nicolini, Ghiselli, et al., 2023**). As a matter of fact, major efforts have been dedicated to identify sex-biased genes through differential gene expression (DGE) analyses (e.g., **Milani et al., 2013; Teaniniuraitemoana et al., 2014; Zhang et al., 2014; Capt et al., 2018; Afonso et al., 2019**), but very few have leveraged cutting-edge techniques to investigate their actual role in SD and/or gonad differentiation and development (e.g., **Liang et al., 2019; Sun et al., 2022**).

Components of the Dmrt, Sox, and Fox gene (DSFG) families are notoriously known as key actors in several developmental processes across Metazoa (**Benayoun et al., 2011; Matson and Zarkower, 2012; Sarkar and Hochedlinger, 2013; Mawaribuchi et al., 2019**), including SD in certain clades: the aforementioned *Dm-W*, *Dmy*, and *dsx* all belong to the *dsx* and *mab-3* related transcription factor (Dmrt) gene family, while *Sry* belongs to the *Sry*-related HMG-box (Sox) gene family; *Fox-L2*, which takes part in most of the vertebrate SD processes as a downstream effector of the female pathway, belongs to the forkhead box (Fox) gene family. Members of the DSFGs have been identified as putative SRGs also in bivalves, thanks to both DGE analyses and mRNA *in-situ* hybridization (mRNA-ISH; e.g., **Naimi et al., 2009; Li, Zhang, et al., 2018; Liang et al., 2019; Yue et al., 2021**), suggesting that their role in morphological and sexual development is maintained also in the clade. However, the clear role of DSFGs has yet to be elucidated, probably as a consequence to the lack of (i) a systematic classification of the families and (ii) a comprehensive understanding of their evolutionary history.

In order to overcome such limitations, the present study aims to perform a thorough inves-

tigation of the DSFG families in bivalves, with the attempt to provide a high-quality resource to be used as a reference for future studies. Through the analysis of more than 40 annotated bivalve genomes and transcriptomes, we aim (i) to describe the complete set and evolutionary history of DSFGs in bivalves by means of phylogenetic inferences, manual curation, and orthology prediction; furthermore, we aim (ii) to identify DSFGs potentially involved in bivalve SD by investigating their sequence evolution in a genome-wide context. As a matter of fact, our hypothesis is that, if any of the DSFGs is directly involved in SD (i.e., is a SDG), then we should expect it to be experiencing a higher rate of sequence evolution, as already found in previous studies (**Pamilo and O'Neill, 1997; Mawaribuchi et al., 2012**) and discussed earlier; this characteristic, in turn, would be reflected in a high diversity of the extant amino acid sequences across the bivalve clade. To assess the robustness and reliability of our approach, we additionally applied our pipeline to two non-bivalve datasets, composed of mammal and *Drosophila* species, respectively (hereon referred to as the ‘mammal dataset’ and the ‘fruit fly dataset’). By choosing two clades for which SD is well characterised, we wanted to compare our results with those obtained on taxa for which a deeper and more detailed knowledge is available. Particularly, mammals and *Drosophila* provide two different frameworks to study the patterns of molecular evolution in SDGs: the former is a system where SD is completely genetic (i.e., the development into a male or into a female is triggered by the up- or down-regulation of *Sry* in undifferentiated gonads, respectively), while the latter is a system where SD is chromosomal, thus lacks a master SDG (the sexual fate of the individual is determined by the ratio between autosomal and X chromosomes). Hence, they represent opposing control datasets to be compared to bivalves, as it is expected that a higher rate of sequence evolution concerns only master SDGs (i.e., the top regulatory part of the SD cascade), but not also the downstream genes (i.e., the bottom effectors). If our method is robust, we should thus expect that, (i) in the mammalian dataset *Sry* is detected as rapidly-evolving (**Pamilo and O'Neill, 1997; Mawaribuchi et al., 2012**), while (ii) in the fruit fly dataset no gene among those working within the sex-determining cascade (including *dsx*) is evolving at a higher pace (**Haerty et al., 2007; Mullon et al., 2012; Baral et al., 2019**). By testing the performance of the pipeline in mammals and fruit flies, we were able to assess the reliability of results in bivalves.

This work offers novel insights into the evolutionary dynamics of SRGs and contributes a valuable genomic resource for understanding SD in bivalves, one of the most ecologically and economically important groups of marine organisms. Particularly, here we provide the first

extensive phylogenetic-based classification of DSFGs in bivalves, covering many species from the major bivalve orders, along with a comprehensive investigation of their sequence evolution.

4.2 Materials and methods

4.2.1 Dataset of bivalve annotated genomes and transcriptomes

Annotated genome assemblies of bivalves were obtained from various publicly available resources, while reference genome assemblies for gastropods and cephalopods were downloaded from NCBI (**Tab. S1**). Isoforms were removed from genome annotations using a perl script from the AGAT toolkit (v0.8.0; **Dainat et al., 2022**). Concerning *Sinonovacula constricta* (Adapedonta), the nucleotide coding sequence fasta file was not available for download. To avoid excluding the species from our analyses, the file was generated in-house by mapping the annotated protein sequences on the reference genome using miniprot (v0.13-0; **Li, 2023**). Then, the corresponding nucleotide sequences were extracted using AGAT.

In order to provide an extensive identification of SRGs also for underrepresented bivalve orders (mainly belonging to the Heterodonta clade), 14 additional species represented by sequenced transcriptomes were included in the analyses. Assembled and annotated transcriptomes were obtained from **Piccinini et al., 2021** and **Iannello et al., 2023**. The resulting set of annotated genomes and transcriptomes (hereafter referred to as the ‘comprehensive set’) was checked for completeness using BUSCO with the Metazoa reference dataset (v5.2.2; **Manni et al., 2021**).

4.2.2 Identification and classification of Dmrt, Sox and Fox genes in bivalves

Members of DSFG families were retrieved in the comprehensive set with HMMsearch from the HMMER package (v3.3.2; **Eddy, 2011**). The signature catalytic domains of each family were used as queries. Specifically, hidden Markov model (HMM) profiles were built after the Pfam databases for the *dsx* and *mab-3* (DM) domain (PF00751), the high mobility group (HMG) box (PF00505) and the forkhead domain (PF00250) to retrieve members of the DSFG families, respectively. The e-value for both the per-target and the per-domain inclusion threshold was set to 10^{-5} .

Obtained hits were then annotated using (i) the PANTHER HMM standalone sequence

scoring against the PANTHER library v18.0 and (ii) RPS-BLAST (v2.5.0+) against the Conserved Domain Database (CDD; pre-compiled version, downloaded from <ftp.ncbi.nih.gov> on 09/11/23). In both cases, hits with an e-value of 10^{-5} were retained. Genes which were correctly annotated by both systems (on the basis of the PANTHER gene family and CDD domain identifiers; **Tab. S2**) were kept for subsequent analyses.

DSFGs from *Homo sapiens*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (**Tab. S3**; hereafter referred to as ‘reference species’) were retrieved from NCBI and were used as reference genes for annotation (see below). Classification and nomenclature of each family was retrieved from: **Mawaribuchi et al. (2019)** for Dmrt genes; **Phochanukul and Russell (2010)** and **Sarkar and Hochedlinger (2013)** for Sox genes; **Mazet et al. (2003)** for Fox genes.

The alignments of mollusc and reference DSFGs were guided by the aforementioned Pfam HMM profiles and performed with Clustal Omega (v1.2.3; **Sievers et al., 2011**), then trimmed with trimAl (v1.4.rev15; **Capella-Gutiérrez et al., 2009**) with a gap threshold of 40 %. Resulting alignments were manually inspected to remove sequences with incomplete catalytic domains, then aligned and trimmed again as before. Phylogenetic trees were inferred using IQ-TREE (v2.1.4-beta COVID-edition; **Minh et al., 2020**) with automatic model selection (**Kalyaanamoorthy et al., 2017**), 1,000 bootstrap replicates and 5 independent runs. The phylogenetic tree of Dmrt genes was midpoint rooted, as no clear homology relationship has been found with other gene families or zinc-finger proteins so far (**Wexler et al., 2014**). Phylogenetic trees of Sox and Fox gene families were rooted using two fungi mating protein A (Mat-A) sequences (XP_62685912.1, CCD57795.1) and two Amoebozoa forkhead-like domains (XP_004368148.1, XP_004333268.1), respectively (**Nakagawa et al., 2013**; **Heenan et al., 2016**). The rooting was performed with Gotree (v0.4.5; **Lemoine and Gascuel, 2021**). To identify and annotate bivalve homology groups within each gene family, we employed a species overlap algorithm followed by a Markov clustering algorithm (MCL) weighted by node supports as implemented in Possvm (v1.2; **Grau-Bové and Sebé-Pedrós, 2021**). DSFGs from *H. sapiens*, *D. melanogaster*, and *C. elegans* were used as reference annotation.

In order to better establish the orthology relationships among ambiguous groups of Dmrt and Fox genes, we run a series of other phylogenetic reconstructions (see **Section 4.4**), by using the same pipeline as before. In the case of *Fox-Y* genes, we also employed Fox gene sequences from the sea urchin *Strongylocentrotus purpuratus*, as given by **Tu et al. (2006)**. All

the phylogenetic trees were plotted using the R package ‘ggtree’ (**Yu et al., 2017**).

4.2.3 Sequence diversity of bivalve single-copy orthogroups

As a metrics to measure the sequence diversity of bivalve DSFGs, and test whether those putatively involved in SD show higher values than other genes, we employed the amino acid sequence divergence (AASD). As a matter of fact, this metric is fast and straightforward to obtain, as it only requires the amino acid alignment and the corresponding best-fit substitution mode.

To this purpose, we produced amino acid alignments of bivalve single-copy orthogroups (SCOs) and built the distribution of their median AASD. Specifically, we assembled a second dataset (hereafter referred to as the ‘reduced bivalve dataset’) which includes, for each bivalve genus, only the best genomes and transcriptomes in terms of either BUSCO scores (on the ‘metazoan_odb10’ dataset; **Manni et al., 2021**) or assembly statistics (**Tab. S1**), in order to reduce computational time. *Archivesica marissinica* (now *Calyptogena marissinica*) and *Saccostrea glomerata* were also removed, as their annotated coding sequences contain many stop codons, which prevent accurate amino acid guided alignments. Genes were clustered in orthologous groups using OrthoFinder (v2.5.5; **Emms and Kelly, 2019**) with DIAMOND ultra-sensitive and default parameters. Resulting orthogroups were splitted into SCOs using DISCO (v1.3.1; **Willson et al., 2022**), and orthogroups with at least 17 species (50 % of the species included in the bivalve reduced dataset) were retained. Amino acid and nucleotide sequences of SCOs were then aligned using Clustal Omega as implemented in TranslatorX (v1.1; **Abascal et al., 2010**), and jointly trimmed using trimAl with a gap threshold of 40 % and the removal of spurious sequences (`-resoverlap 50 -seqoverlap 50`). Eventually, orthogroups containing (i) internal stop codons, (ii) with less than 17 species left (50 % of the species included in the bivalve reduced dataset), or (iii) containing DSFGs were removed from downstream analyses. The best amino acid substitution model was inferred for each trimmed alignment using ModelFinder as implemented in IQTREE2 (model search was restricted to matrices accepted by the ‘phangorn’ R library; i.e., Blosum62, cpREV, Dayhoff, DCMut, FLU, HIVb, HIVw, JTT, JTDCMut, LG, mtART, mtMAM, mtREV, mtZOA, rtREV, VT, WAG) and the corresponding pairwise amino acid distances were computed with the function ‘dist.ml’ from the ‘phangorn’ R package (**Schliep, 2011**). The same pipeline was also employed to obtain pairwise

amino acid distances for each DSFG single-copy orthologous group. We decided to employ the pairwise amino acid distance instead of the tip-to-tip phylogenetic distance (which accounts for a more comprehensive evolutionary signal) in order to save computational time. However, to check whether the two metrics were comparable to each other, we randomly selected 200 SCOs (including orthogroups from the DSFGs) and computed the maximum likelihood (ML) trees using IQTREE2, with ModelSelection restricted as before. Then, the tip-to-tip pairwise distances were obtained with the R package ‘adephylo’ (**Jombart and Dray, 2010**).

The distribution of amino acid distances was then built after the median values of pairwise distances of each SCO, and genes were categorised accordingly into three groups: Group 1, consisting of genes from the 1 % upper quantile of the distribution; Group 2, consisting of genes between the 1 and 5 % upper quantiles; and Group 3, consisting of all the remaining genes. Group 1 and Group 2 genes will be referred to as ‘highly-divergent genes’.

4.2.4 Mammals and *Drosophila* spp. as test datasets

To validate our approach for the study of bivalve SRG molecular evolution, we run the same analysis on two additional datasets, consisting of reference genomes of mammals and *Drosophila* species (**Tab. S4** and **S5**, respectively), whose sex-determining mechanisms are well studied and characterised. As a matter of fact, despite it is well known that SDGs tend to evolve faster than genes not involved in SD, the hypothesis has never been tested extensively across the entire phylogenetic diversity of a group: molecular evolution of SDGs and SRGs has mainly been tested on single/pairs of species or inside the boundaries of taxonomic genera (**Stothard and Pilgrim, 2003; Haerty et al., 2007; Mank et al., 2007; Mullon et al., 2012; Papa et al., 2017; Ghiselli et al., 2018**). For both mammals and fruit flies, annotated genomes were downloaded from NCBI using the command-line tool ‘datasets’, then processed using the same pipeline and scripts as before (**Fig. 4.1**).

4.2.5 GO-term enrichment

After having obtained the distributions of AASD in the three datasets (reduced bivalves, mammals, and fruit flies) and having sorted SCOs genes up into 3 groups (Group 1, Group 2, and Group 3), we performed a gene ontology (GO) enrichment analysis of genes from Group 1 and genes from Group 1 + Group 2. To do so, we firstly selected one gene per SCO, giving priority

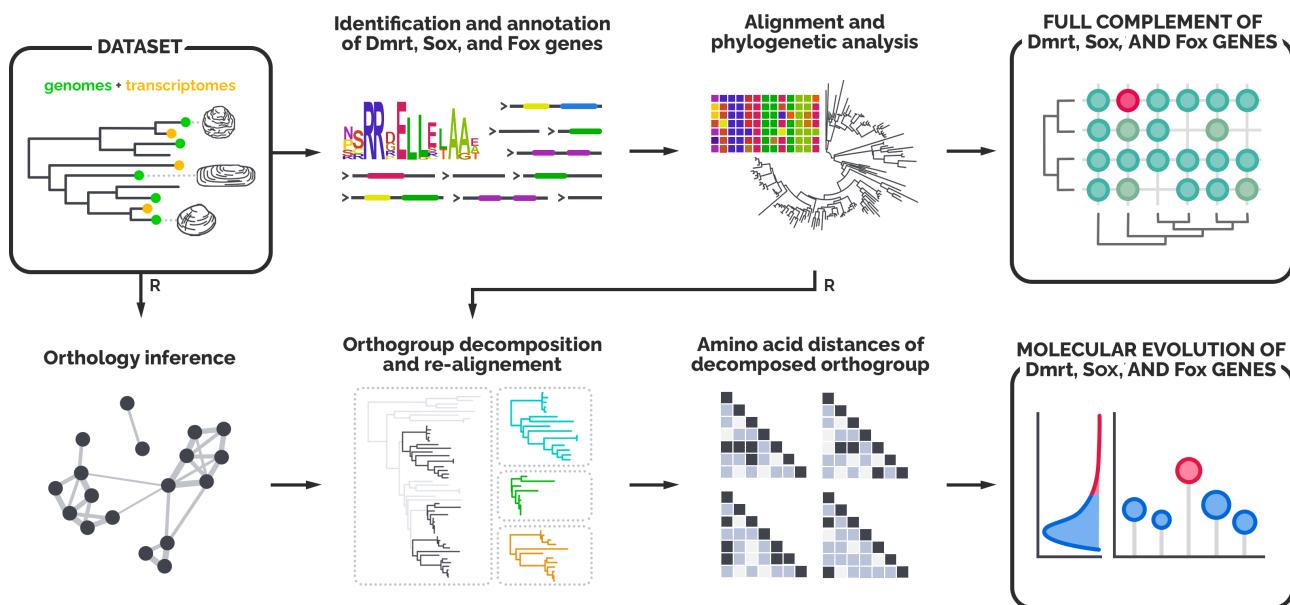


Figure 4.1 – Workflow of the analyses for the bivalve dataset. Starting from a set of both genomes and transcriptomes covering a great portion of bivalve taxonomic diversity, we first characterized the entire complement of DSFG genes (upper row). In particular, we used sequence annotation and phylogenetic tools to obtain reliable sequences and filter out any putative mis-assembled or mis-annotated sequence. Afterwards, we built a reduced set of transcriptomes and genomes (the reduced bivalve dataset, where we minimized the redundancy of congeneric species) from which to draw the molecular evolution patterns of orthologous genes (bottom row). In particular, after having obtained SCOs, we calculated the amino acid distances within each orthogroup and then built the distribution of median values. The same pipeline was also employed for the mammal and the fruit fly datasets, with just two minor differences: the starting dataset was composed of only genomes, and that the reduction step (R) was not necessary.

to few chosen species: (i) for bivalves, we selected genes from *Pecten maximus*, or alternatively from *C. gigas*, *Hyriopsis bialata* (now *Unio delphinus*), *Tridacna squamosa*, and *Solen grandis*; (ii) for mammals, we selected genes from *H. sapiens*, or alternatively from *Bubalus bubalis*, *Panthera tigris*, *Camelus dromedarius*, and *Monodelphis domestica*; (iii) for fruit flies, we selected genes from *D. melanogaster*, or alternatively from *Drosophila hydei*, *Drosophila pseudoobscura*, and *Drosophila suzukii*. By doing so, we ensured that each SCO was represented by one gene. Afterwards, we annotated the obtained datasets with the corresponding GO terms using the OMA browser (accessed on 18/09/2024; **Altenhoff et al., 2024**). The GO-term enrichment of Group 1 genes and Group 1 + Group 2 genes was performed with the R package ‘topGO’ with the Fisher exact test (**Alexa and Rahnenführer, 2009**).

4.3 Results

4.3.1 Genomic and transcriptomic datasets

The complete bivalve dataset consists of 29 bivalve genomes, 14 bivalve transcriptomes, and 7 outgroup genomes (5 gastropods and 2 *Octopus* spp.; **Tab. S1**). BUSCO statistics for complete single-copy genes spanned from the 64.9 % in *Modiolus modiolus* to the 99.4 % of *Perna viridis*, with a median value of 94.7 %. We were able to get at least one representative species for 11 different bivalve orders, covering a good proportion of the phylogenetic diversity of the clades Pteriomorpha, Palaeoheterodonta, and Imparidentia, and thus building the most extensive genomic and transcriptomic dataset for bivalve comparative analyses so far (**Tab. S1**). Unfortunately, no genomes or transcriptomes for Protobranchia, Archiheterodonta, and Anomalodesmata were available at the time of the project, thus we were not able to include any of those clades in our analysis. The reduced bivalve dataset (used for the orthology inference and the molecular evolution analysis; **Fig. 4.1**) consists instead of 36 genomes and transcriptomes (**Tab. S1**), and was built to retain just one species for each taxonomic genera.

The mammal dataset consists of 32 species and 1 outgroup (*Gallus gallus*, Aves; **Tab. S4**), and covers 12 major orders, while the fruit fly dataset consists of 17 species and 1 outgroup (*Anopheles gambiae*, Culicidae; **Tab. S5**), and covers 2 *Drosophila* subgenera (i.e., *Drosophila* and *Sophophora*). BUSCO statistics for complete single-copy genes were generally higher than those of bivalves, with a median of 98.3 % for mammals and of 99.8 % for fruit flies (**Tab. S4** and **S5**).

4.3.2 The Dmrt, Sox, and Fox complements in bivalves

Our annotation pipeline managed to successfully identify and annotate DSFGs in bivalves, as proved by the same analysis in mammals and fruit flies (see **Section 4.3.4**).

We retrieved four main orthology groups of Dmrt genes in bivalves (**Fig. 4.2** and **S1**; **Tab. S6**), three corresponding to the groups present in the Bilateria common ancestor (*Dmrt-2*, *Dmrt-3*, and *Dmrt-4/5*; **Mawaribuchi et al., 2019**), and one additional group with no unambiguous ortholog among reference genes, and thus putatively specific to molluscs (named *Dmrt 1-like* [*Dmrt-1L*], as per **Li, Zhang, et al., 2018**; **Evensen et al., 2022**). The majority of identified Dmrt genes are present in single-copy in each species, but *Dmrt-4/5*s show a group-

specific expansion in Palaeoheterodonta and Heterodonta, while *Dmrt-1L* is completely absent from Heterodonta. The degree of missing data for *Dmrt* genes in bivalves is about 35 %, with *Dmrt-2* having the highest (~56 %) and *Dmrt-4/5* the lowest (~7 %; **Tab. S7**). The coupling of ubiquitin conjugation to endoplasmic reticulum degradation (CUE)-like DM-associated (DMA) domain has been annotated in most of the *Dmrt-3* and *Dmrt-4/5* genes, while an additional DM domain has been annotated in *Dmrt-1L* genes in Mytilida and the gastropod *Pomacea canaliculata* (**Tab. S6**). Concerning *Sox* genes, we retrieved six main orthology groups, none of which is restricted to molluscs or bivalves (**Fig. 4.2** and **S2**; **Tab. S6**). Five *Sox* groups (*Sox-B1/2*, *Sox-C*, *Sox-D*, *Sox-E*, and *Sox-F*) are those traditionally considered to be present in the Bilateria common ancestor (**Phochanukul and Russell, 2010**), while one has been identified outside mammals only recently (*Sox-H*, or *Sox-30*; **Han et al., 2010**). *Sox-B2* and *Sox-B1* have been grouped in the same clade, as in our phylogenetic reconstruction the former results in a paraphyletic group with the latter (**Fig. S2**), despite being traditionally recognised as a separate paralogy group in humans, fruit flies, and nematodes. The degree of missing data for *Sox* genes in bivalves is ~8 %, with *Sox-H* having the highest (~21 %) and *Sox-B1/2* and *Sox-C* both having no missing genes (**Tab. S7**). The *Sox* N-terminal signature domain was annotated for *Sox-E* genes (**Tab. S6**). Concerning *Fox* genes, we retrieved 27 main orthology groups (**Fig. 4.2** and **S3**; **Tab. S6**), two of which are specific to molluscs (*Fox-OG13/NA*, *Fox-OG16/NA*). Additionally, other potential mollusc-specific *Fox* groups have been identified, but these have been excluded from the final orthology analysis as they are present in less than half of bivalve species (see **Section 4.2**; **Tab. S6**). The two major *Fox* gene subgroups, Group I (monophyletic, specific to Metazoa; includes *Fox-A*, *Fox-B*, *Fox-C*, *Fox-D*, *Fox-E*, *Fox-F*, *Fox-G*, *Fox-H*, *Fox-L1*, *Fox-L2*, *Fox-Q2*) and Group II (paraphyletic, specific to Opisthokonta; includes *Fox-O*, *Fox-P*, *Fox-J2*, *Fox-J1*, *Fox-K*, *Fox-N2/3*, *Fox-N1/4*; **Larroux et al., 2008**), have been recovered, including the four *Fox* genes that were present in the Bilateria common ancestor (*Fox-C*, *Fox-F*, *Fox-L1*, and *Fox-Q1*; **Shimeld et al., 2010**). Two putative lineage-specific expansions have been recovered for *Fox-OG28/NA*, one regarding *Mytilus* spp. and one regarding the two Myida species (**Fig. 4.2**; **Fig. S3**). The degree of missing data for *Fox* genes in bivalves is about 22 %, with *Fox-H* having the highest (~42 %) and *Fox-J1* having no missing genes (**Tab. S7**). The forkhead-associated (FHA) domain was annotated for *Fox-K* genes, the *Fox-P* coiled-coil signature domain was annotated for *Fox-P* genes, while both the forkhead N- and C-terminal signature domains were annotated for *Fox-A* genes (**Tab. S6**). Regarding bivalve

species, the amount of missing data greatly differs between genomes and transcriptomes, with a mean of about 9 % and about 45 %, respectively. *Argopecten irradians concentricus*, *Mytilus coruscus* (now *unguiculatus*), and *P. maximus* have no missing data, while *Loripes orbiculatus* has the highest proportion (~64 %; **Tab. S7**).

4.3.3 Amino acid sequence divergence of Dmrt, Sox, and Fox genes in bivalves

In the reduced bivalve dataset, OrthoFinder collectively analysed >1.2 M genes distributed in 34 species. 89.4 % of these genes were placed in orthogroups, while 10.6 % were not. The number of retrieved SCOs is 5, which is drastically low but can be explained considering the mixed nature of the dataset, that is, including both genomes and transcriptomes with highly different BUSCO scores (**Tab. S1**). In order to be able to analyse a greater number of genes, we decomposed OrthoFinder orthogroups using DISCO and eventually obtained ~11 k SCOs with at least 50 % of the species. By running the same pipeline on DSFGs, we included in the AASD analysis 32 SCOs (**Fig. 4.2**) out of 33 initial Possvm-identified groups (*Fox-H* didn't meet the species occupancy threshold; **Fig. 4.3**).

From the distribution of median AASD, 112 genes were assigned to Group 1 (1 % upper quantile), 447 to Group 2 (5 % upper quantile), and 10.603 to Group 3. Most of the DSFGs (29/32) fell in Group 3 (**Fig. 4.3**), which means they have a median AASD comparable to the vast majority of other genes in bivalves (median level of the genomes). Just *Dmrt-1L*, *Sox-H*, and *Sox-F* showed higher divergences, and have been accordingly placed in Group 2. Overall, pairwise AASD proved to be a good approximation of the tip-to-tip distances ($R = 0.84, p < 2.2 \times 10^{-16}$, calculated on 200 randomly-selected trees; **Fig. 4.3C**), while it showed no influence from the alignment length ($R = 0.11$) or the number of represented species ($R = -0.23$; **Fig. 4.3D** and **4.3E**). Genes from Group 1 and Group 2 are strongly involved in cellular regulatory processes (such as those related to the metabolism of nucleic acids, proteins, and other macromolecules), but also in development and response to external stimuli, as shown by the GO-term enrichment analysis (**Tab. 4.1** and **S10**).

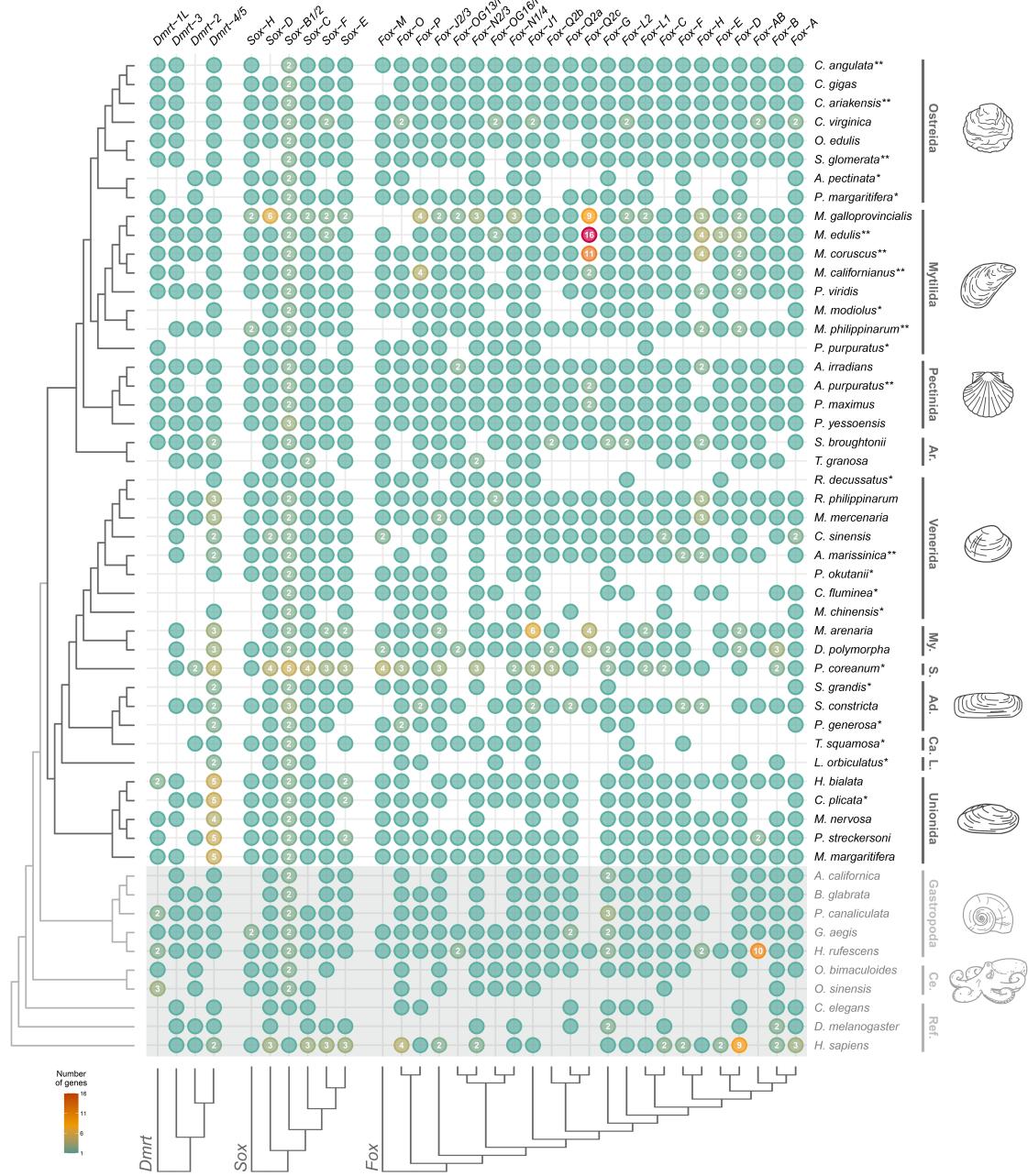


Figure 4.2 – DSFG complement in bivalves and their outgroups. Presence/absence of genes in various species are indicated by filled circles. Numbers inside each circle specify genes with 2 or more copies. The shaded area highlights non-bivalve species, belonging either to other molluscs or to the references. The phylogenetic tree of analyzed species, as inferred from literature, is shown on the left, while major taxonomic groups are reported on the right. Species represented by transcriptomic data are marked with an asterisk ('*'), and species not present in the reduced bivalve dataset are marked with two asterisks ('**'; see main text and **Fig. 4.1**); note that the two categories do not overlap. DSFG trees are shown on the bottom (full trees can be found in **Fig. S1–S3**). Full species names, along with all assembly and taxonomic information, can be found in **Tab. S1**. Ad.: Adapedonta; Ar.: Arcida; Ca.: Cardiida; Ce.: Cephalopoda; L.: Lucinida; My.: Myida; Ref.: reference genes; S.: Sphaeriida.

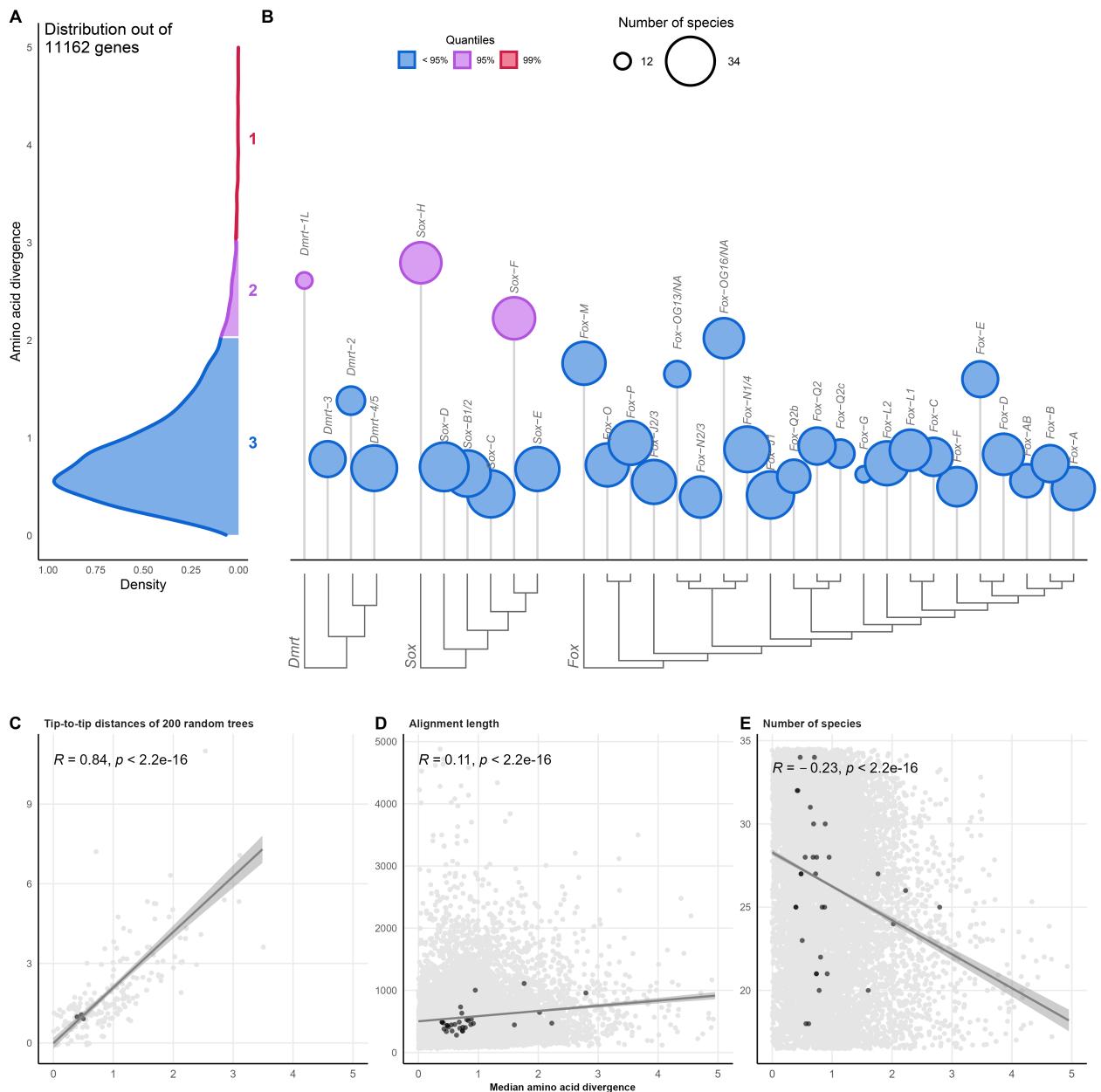


Figure 4.3 – Distribution of AASD of single-copy orthogroups in bivalves (A), including DSFGs (B), and their correlations with tip-to-tip distances (C), alignment lengths (D), and number of species (E). The distribution of AASD has been computed on the median values of pairwise distances of >11k SCOs from the reduced bivalve dataset (see main text and **Fig. 4.1**). Genes have been divided according to their median AASD value into three different groups, which are indicated by different colors and increasing numbers (Groups 1, 2, and 3). Circle heights of DSFGs show the median value of their AASD, while the size indicates the number of represented species. DSFG trees are shown on the bottom (full trees can be found in **Fig. S1** and **S3**). Darker points in C–E indicate DSFG SCOs. The correlation between the amino acid distance and the tip-to-tip distance has been computed on 200 randomly-selected orthogroups.

Table 4.1 – Top enriched GO terms for highly-divergent genes of bivalves, mammals, and *Drosophila*. The extended version of the table, which includes also the expected number of annotated genes per GO term and all the other enriched GO terms, can be accessed in [Tab. S10](#).

Dataset	GO.ID	Term	Annotated genes	Significant genes	Classic genes	Fisher
Bivalvia	GO:0060255	regulation of macromolecule metabolic process	737	59	0.0453	
	GO:0080090	regulation of primary metabolic process	673	53	0.0182	
	GO:0019219	regulation of nucleobase-containing compound metabolic process	541	41	0.0239	
	GO:0006351	DNA-templated transcription	571	39	0.0377	
	GO:0032774	RNA biosynthetic process	579	39	0.0449	
	GO:0051252	regulation of RNA metabolic process	517	37	0.0272	
	GO:0006355	regulation of DNA-templated transcription	490	35	0.0375	
	GO:2001141	regulation of RNA biosynthetic process	491	35	0.0384	
	GO:0006950	response to stress	370	33	0.0195	
	GO:0032502	developmental process	261	27	0.0445	
	GO:0006468	protein phosphorylation	345	23	0.0248	
	GO:0031325	positive regulation of cellular metabolic process	125	17	0.0080	
	GO:0010604	positive regulation of macromolecule metabolic process	151	17	0.0405	
	GO:0051172	negative regulation of nitrogen compound metabolic process	117	16	0.0081	
	GO:0051173	positive regulation of nitrogen compound metabolic process	137	15	0.0245	
	GO:0006310	DNA recombination	66	14	0.0009	
	GO:0048513	animal organ development	83	12	0.0409	
	GO:0010629	negative regulation of gene expression	78	11	0.0005	
	GO:0023051	regulation of signaling	133	11	0.0287	
	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	64	11	0.0364	
	GO:0009605	response to external stimulus	90	11	0.0454	

Tab. 4.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes	Significant genes	Classic Fisher
Bivalvia	GO:0044419	biological process involved in interspecies interaction between organisms	63	11	0.0476
	GO:0006955	immune response	1,297	145	0.0006
	GO:0098542	defense response to other organism	853	112	0.0207
	GO:0045087	innate immune response	647	82	0.0000
	GO:0001817	regulation of cytokine production	630	51	0.0466
	GO:0042742	defense response to bacterium	233	45	0.0000
	GO:0006954	inflammatory response	642	45	0.0174
	GO:0019221	cytokine-mediated signaling pathway	382	44	0.0000
	GO:0002250	adaptive immune response	342	44	0.0000
	GO:0001819	positive regulation of cytokine production	402	41	0.0272
	GO:0002697	regulation of immune effector process	308	37	0.0443
	GO:0042110	T cell activation	432	35	0.0256
	GO:0051607	defense response to virus	257	34	0.0000
	GO:0048232	male gamete generation	491	32	0.0226
	GO:0007283	spermatogenesis	478	31	0.0280
Mammalia	GO:0070661	leukocyte proliferation	273	29	0.0129
	GO:0002449	lymphocyte mediated immunity	221	29	0.0483
	GO:0070663	regulation of leukocyte proliferation	212	25	0.0187
	GO:0050727	regulation of inflammatory response	300	24	0.0024
	GO:0031349	positive regulation of defense response	240	24	0.0124
	GO:0002768	immune response-regulating cell surface receptor signaling pathway	177	22	0.0034
	GO:0050829	defense response to Gram-negative bacterium	66	17	0.0000
	GO:0071222	cellular response to lipopolysaccharide	164	17	0.0001

Tab. 4.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes	Significant genes	Classic Fisher
Mammalia	GO:0010466	negative regulation of peptidase activity	163	16	0.0004
	GO:0002429	immune response-activating cell surface receptor signaling pathway	164	16	0.0024
	GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	137	16	0.0124
	GO:0071706	tumor necrosis factor superfamily cytokine production	137	16	0.0124
	GO:0070665	positive regulation of leukocyte proliferation	132	16	0.0277
	GO:0045089	positive regulation of innate immune response	113	16	0.0322
	GO:0071356	cellular response to tumor necrosis factor	175	15	0.0022
	GO:0002695	negative regulation of leukocyte activation	148	15	0.0115
	GO:0002456	T cell mediated immunity	82	15	0.0161
	GO:0002705	positive regulation of leukocyte mediated immunity	113	15	0.0184
<i>Drosophila</i>	GO:0032680	regulation of tumor necrosis factor production	133	15	0.0326
	GO:0032640	tumor necrosis factor production	133	15	0.0326
	GO:0050866	negative regulation of cell activation	165	15	0.0405
	GO:0000819	sister chromatid segregation	140	11	0.0293
	GO:0070192	chromosome organization involved in meiotic cell cycle	54	9	0.0085
<i>Drosophila</i>	GO:0007131	reciprocal meiotic recombination	37	7	0.0007
	GO:0007143	female meiotic nuclear division	54	6	0.0227
	GO:0035967	cellular response to topologically incorrect protein	44	5	0.0333
	GO:0035966	response to topologically incorrect protein	47	5	0.0427
	GO:0007141	male meiosis I	13	4	0.0015
	GO:0140543	positive regulation of piRNA transcription	3	3	0.0001
	GO:0010526	retrotransposon silencing	8	3	0.0033
	GO:0007130	synaptonemal complex assembly	10	3	0.0067

Tab. 4.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes	Significant genes	Classic Fisher
<i>Drosophila</i>	GO:0030719	P granule organization	11	3	0.0089
	GO:0071218	cellular response to misfolded protein	12	3	0.0115
	GO:0051788	response to misfolded protein	12	3	0.0115
	GO:0007135	meiosis II	15	3	0.0217
	GO:0034508	centromere complex assembly	19	3	0.0409

4.3.4 Dmrt, Sox, and Fox genes, and amino acid sequence divergence in the test datasets

The DSFG datasets retrieved in mammals and fruit flies are far more complete than those in bivalves, and most of the already-recognised orthology groups have been identified.

In mammals, we retrieved 7 Dmrt orthology groups with about 3.1 % of missing data, 20 Sox orthology groups with about 8.1 % of missing data, and 42 Fox orthology groups with about 4.6 % of missing data (**Fig. S4A, S5, and S7; Tab. S8**). Of these, just *Sox-5* was not included in the subsequent AASD analysis, as it did not meet the 50 %-species occupancy threshold. OrthoFinder analysed about 650 M genes, and the number of SCOs used in the AASD analysis (thus resulting from the DISCO-based orthogroup decomposition pipeline) is >16 k (**Fig. 4.4A**). From the distribution of median AASD, 163 genes were assigned to Group 1, 649 to Group 2, and 15,355 to Group 3. Most of the DSFGs (66/68) fell in Group 3 (**Fig. 4.4B**), while *Sry* and *Fox-D4* showed higher divergences, and have been accordingly placed in Group 1 and Group 2, respectively. Highly-divergent genes show a strong enrichment in immune-related functions (such as innate and adaptive immune response, defence response to bacteria and viruses, lymphocyte metabolism, etc.), but also in reproductive processes (such as spermatogenesis; **Tab. 4.1** and **S10**).

Concerning *Drosophila*, we retrieved 4 Dmrt orthology groups with about 1.7 % of missing data, 7 Sox orthology groups with about 3.9 % of missing data, and 17 Fox genes with about 8.3 % of missing data (**Fig. S4B, S8, and S10; Tab. S9**). OrthoFinder analysed about 240 M, and the distribution of median AASD was built after >12k SCOS (**Fig. 4.4C**). 126 genes were assigned to Group 1, 501 to Group 2, and 11,880 to Group 3. All of the DSFGs have been used in the AASD analysis, but none of them have been placed in Group 1 or 2, that is, all the DSFGs in *Drosophila* have an AASD comparable to the median level of the genome (**Fig. 4.4D**). Genes of Group 1 and Group 2 show a GO-term enrichment in meiotic processes, such as chromosome/chromatid organisation, and retrotransposon silencing (**Tab. 4.1** and **S10**).

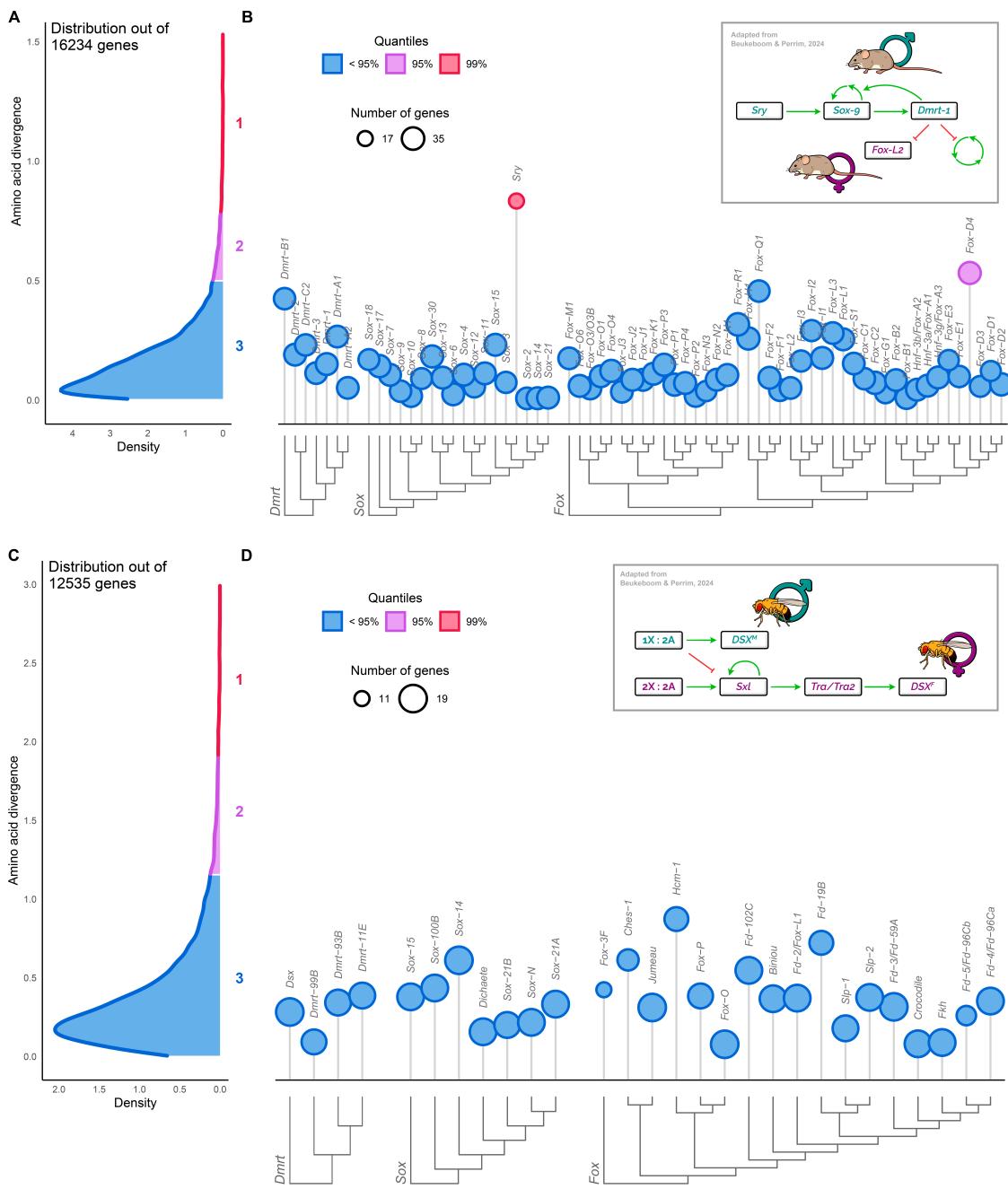


Figure 4.4 – Distribution of AASD of single-copy orthogroups in Mammalia (A) and *Drosophila* (C), including DSFG (B-D). The distributions of AASD in mammals and fruit flies have been computed on the median values of pairwise distances of over 16k and 12k SCOs, respectively. Genes have been divided according to their median AASD value into three different groups, which are indicated by different colors and increasing numbers (Groups 1, 2, and 3). Circle heights of DSFGs show the median value of their AASD, while the size indicates the number of represented species. DSFG trees are shown on the bottom (full trees can be found in Fig. S5–S7 for mammals and in Fig. S8–S10 for fruit flies). Insets: scheme of the sex-determination molecular pathways in *Mus musculus* and in *Drosophila melanogaster*, with shown the main genes involved (adapted from Beukeboom and Perrin, 2014). Green arrows indicate transcription activations, red arrows indicate transcription suppressions. X: sex chromosomes; A: autosomal chromosomes; *DSX^{M/F}*: *DSX* splicing variants present in males or females, respectively.

4.4 Discussion

4.4.1 A new manually-curated and phylogenetic-based reference dataset of Dmrt, Sox, and Fox genes in bivalves

The annotation and characterisation process of a gene family in a certain clade of organisms may harbour many overlooked challenges (**Vizueta Moraga et al., 2020**). For example, the presence of highly-conserved catalytic domains may hamper the correct identification of the components of a gene family because of insufficient phylogenetic signal, as it is the case for Hox and ParaHox genes and their homeobox motif (**Baldwin-Brown et al., 2018; Nicolini, Martelossi, et al., 2023**). Conversely, the components of dynamic gene families characterised by abrupt and sequential duplication events may be difficult to sort into separate groups. As a matter of fact, varying levels of sequence heterogeneity and gene copy numbers makes the inference of orthologous groups hard, as for certain clans of the P450 gene family (**Dermauw et al., 2020**). Regardless of the causes, having a solid and wide phylogenetic context in which to study gene duplications and losses, and orthology relationships, is crucial to overcome these difficulties. In the same way, manual curation and visual inspection of multiple sequence alignments, phylogenetic trees, and gene structures (in terms of domain composition, start and stop codons, and other feature representations) is helpful, despite being time-demanding and possibly low reproducible. In this study, we characterised the full complement of DSFGs in the vast class of bivalves, by leveraging sequence domain annotation, phylogenetics, and manual curation of the dataset. Our aim was to obtain the most reliable gene complements as possible, combined with a vast taxonomic dataset, a solid phylogenetic inference, an openly-available dataset of gene sequences, and a reproducible pipeline for the annotation of gene identity. By doing so, we want to provide a reliable resource for future studies of DSFGs, either focused on bivalves or generally in Metazoa.

Concerning the Dmrt gene family, we identified orthologs of the vertebrate *Dmrt-2*, *Dmrt-3*, and *Dmrt-4/5* (or *A1/A2*; **Fig. 4.2** and **S1**; **Tab. S6**), which are also expected to have been present in the Bilateria common ancestor (**Mawaribuchi et al., 2019**). **Wang et al. (2023)** found that *Dmrt-4/5* is duplicated in *Mercenaria mercenaria* and *Cyclina sinensis* (Venerida), and in *Dreissena polymorpha* (Myida), and we confirm this result by tracing back the duplication event to the split between Palaeoheterodonta (here represented by Unionida) and Heterodonta (here

represented by Venerida, Myida, Sphaeriida, Adapedonta, Cardiida, and Lucinida; **Fig. 4.2**). Furthermore, we confirm *Dmrt-1L* to be present in many bivalve species (mainly belonging to the Ostreida, Pectinida, Mytilida, and Unionida orders; **Fig. 4.2**), as well as in gastropods and *Octopus*. Though, our phylogenetic analysis did not retrieve any unambiguous orthology relationship among *Dmrt-1L* and either vertebrate *Dmrt-1* or *Drosophila dsx* genes, as instead it was proposed in previous works (Li, Zhang, et al., 2018; Evensen et al., 2022). As a matter of fact, the amino acid sequence of the *Dmrt-1L* DM domain does not recall that of any other Dmrt gene. Furthermore, it must be considered that various phylogenetic analyses have recovered both *Dmrt-1* and *dsx* genes to be restricted to vertebrates and arthropods, respectively (Wexler et al., 2014; Mawaribuchi et al., 2019; Panara et al., 2019), that is, they do not have any direct ortholog outside their relative clades. Thus, if *Dmrt-1L*, *dsx*, and *Dmrt-1* are true orthologs, their origin would need to be placed at least in the Bilateria common ancestor, which seems however to be not the case. All considered, we thus confirm that *Dmrt-1L* is not orthologous to *Dmrt-1* and *dsx* and is rather a mollusc-specific gene (Evensen et al., 2022). The monophyly of the group is not supported by the phylogenetic tree inferred with Dmrt genes from molluscs and the reference species (**Fig. S1**); though, it is recovered when analysing just genes from mollusc species (**Fig. S11**). To this regard, we speculate that in our analysis, the difficulty in obtaining the monophyly of *Dmrt-1L* genes may have arisen primarily because of the many *C. elegans*-restricted genes (**Tab. S3**), which are placed among the other bivalve genes (**Fig. S1**), but also because of the high AASD of *Dmrt-1L* genes (see **Section 4.4.2**), which hampers a straight-forward phylogenetic reconstruction. Furthermore, our broad-context analysis allowed us to identify some cases of incorrect gene identification in bivalves, which have arisen because of erroneous or ambiguous annotations in previous works, as a result of limited datasets or analyses. For example, the scallop-specific cluster of Dmrt genes retrieved by Wang et al. (2023) rather belongs to the *Dmrt-1L* group, while the classification of Dmrt genes in *Crassostrea* species provided by Zeng et al. (2024) needs to be revised following the one of this work (*Dmrt-1* genes *sensu*-Zhang are *Dmrt-4/5*; *Dmrt-2* genes *sensu*-Zhang are *Dmrt-3*; *Dmrt-3* genes *sensu*-Zhang are *Dmrt-1L*; hence, *Crassostrea* species do not have *Dmrt-2* genes).

For what concerns the Sox gene family, bivalves (or molluscs) do not show any major clade-restricted gene, as only the five Bilateria-specific Sox groups (*Sox-B1/2*, *Sox-C*, *Sox-D*, *Sox-E*, and *Sox-F*) and *Sox-H* have been identified (**Fig. 4.2** and **S2**; **Tab. S6**), in accordance with

previous findings (**Evensen et al., 2022; Wang and Nie, 2024; Yu et al., 2017**). *Sox-B1/2* is clearly made up of two subgroups (i.e., *Sox-B1* and *Sox-B2*), as expected, but their respective identity could not be unambiguously established, as *Sox-B1/2* genes of reference species do not form separate clusters (**Fig. S2**). Even when inferring the phylogenetic tree only of components of the *Sox-B1/2* group from molluscs and reference species, the identity can not be properly established (**Fig. S12**).

Compared to Dmrt and Sox genes, the Fox gene family appears as the most dynamic in terms of gene presence/absence, as already shown by other works (**Wu et al., 2020; Schomburg et al., 2022; Seudre et al., 2022**). Our phylogenetic analysis successfully recovered Group I and Group II of Fox genes (**Larroux et al., 2008**), which include the four Fox genes that were present in the Bilateria common ancestor (*Fox-C*, *Fox-F*, *Fox-L1*, and *Fox-Q1*; **Fig. 4.2** and **S3**; **Tab. S6**; **Shimeld et al., 2010**). To our knowledge, this is the first broad-taxonomic identification and classification of Fox genes in bivalves, as up to now they have been systematically characterised only in *C. gigas* (**Yang et al., 2014**), *Patinopacten yessoensis* (now *Mizuhopecten yessoensis*; **Wu et al., 2020**), and *Ruditapes philippinarum* (**Liu et al., 2024**). Firstly, our analysis confirms the absence in molluscs of *Fox-I*, *Fox-Q1*, *Fox-R*, *Fox-S* (**Fig. S3**), which are in fact thought to have emerged with the diversification of deuterostomes or vertebrates (**Yang et al., 2014; Wu et al., 2020; Schomburg et al., 2022; Seudre et al., 2022**). Furthermore, we have found many Fox groups that appeared as mollusc-specific and/or still-unnamed at a first analysis (*Fox-OG2/NA*, *Fox-OG39/NA*, *Fox-OG15/NA*, and *Fox-OG28/NA*; **Tab. S6**). However, a more in-depth investigation revealed a different scenario. *Fox-OG2/NA* appears close to the human *Fox-M* gene in the phylogenetic tree, but they do not form a monophyletic group (**Fig. S3**). However, by comparing *Fox-OG2/NA* sequences and phylogenetic tree with those analysed by **Yang et al. (2014)**, **Wu et al. (2020)**, **Schomburg et al. (2022)**, and **Seudre et al. (2022)**, it appears clear that this group of Fox genes is indeed *Fox-M*. However, our analysis has failed to retrieve a monophyletic relationship among bivalve and human *Fox-M* genes, even when inferring a tree with just *Fox-J2*, *Fox-M*, *Fox-O*, and *Fox-P* complements (**Fig. S13**), which belong to the same Fox group. Regarding the *Fox-OG39/NA* group, it does not have any homolog in reference species (**Fig. S3**) but is found to belong to the *Fox-AB* group by sequence comparison with previous works (**Yang et al., 2014; Wu et al., 2020; Seudre et al., 2022**). *Fox-AB* was formerly described only in the sea urchin *S. purpuratus* and the lancelet *Branchiostoma floridae* (**Tu et al., 2006; Yu et al., 2008**), but was later identified also in several Spiralia lineages,

including molluscs (e.g., **Yang et al., 2014; Wu et al., 2020; Seudre et al., 2022**). A similar situation concerns *Fox-OG15/NA* and *Fox-OG28/NA*, which again could not be named based on orthology relationships with the reference species genes (**Fig. S3**), but actually represent two lineage-specific expansions of the *Fox-Q2* group (named *Fox-Q2b* and *Fox-Q2c*), as already appointed in previous studies (**Yang et al., 2014; Wu et al., 2020**). This observation fits within the wider context of the *Fox-Q2* group expansion in Bilateria and, particularly, in Spiralia, that led to remarkable differences in their gene copy numbers across various clades (**Seudre et al., 2022**). Two additional Fox genes have been previously identified in bivalves, and were named *Sox-Y* and *Sox-Z* (**Yang et al., 2014; Wu et al., 2020**). In our analysis, these Fox groups were at first identified as *Fox-OG13/NA* and *Fox-OG16/NA*, after sequence comparison of Fox genes from *C. gigas* and *P. yessoensis*. On one hand, *Fox-Y* was firstly identified in *S. purpuratus* (**Tu et al., 2006**) and only recently in a few bivalve species (**Yang et al., 2014; Wu et al., 2020**). However, when analysing bivalve and *S. purpuratus* Fox genes, we failed in retrieving such a clear orthology relationship, as *S. purpuratus* *Fox-Y* does not fall within the phylogenetic range of bivalve *Fox-OG13/NA*, which contains the supposed *Fox-Y* orthologs (**Fig. S14**). Also, the forkhead domains of *Fox-OG13/NA* genes were annotated as ‘forkhead domain P’ (**Tab. S6**). On the other hand, *Fox-Z* was firstly identified in bivalves and in several other protostomes, thanks to a phylogenetic work including the brachiopod *Lingula unguis*, the annelid *Capitella teleta*, the scorpion *Centruroides sculpturatus*, and the centipede *Strigamia maritima* (**Wu et al., 2020**). However, later works have not recovered this Fox gene group, even when analysing annelids (**Seudre et al., 2022**) and panarthropods (**Schomburg et al., 2022**) in a more dedicated effort. In this case, the forkhead domains were annotated as either a generic ‘forkhead domain’ or a ‘forkhead domain Q2’ (**Tab. S6**). All considered, we argue that bivalves possess two additional Fox groups (here *Fox-OG13/NA* and *Fox-OG16/NA*; **Fig. 4.2** and **S3**; **Tab. S6**) which are shared with other mollusc species, as revealed also by other authors. However, given the discordant results of the phylogenetic hypothesis and domain annotation, we think that a more thorough investigation on their orthology relationships with Fox genes from other Metazoa is needed, and thus we chose to not employ their former names *Fox-Y* and *Fox-Z*.

Besides the DSFG groups discussed so far, it must be also considered that many orphan genes have been identified (**Fig. S1** and **S3**; **Tab. S6**). For example, **Wu et al. (2020)** identified a duplication event of *Fox-H* genes in *C. gigas*, which has been recovered also in our analysis for

the entire Ostreida clade (*Fox-OG36/NA*; **Fig. S3**). Similarly, a gene orthology group putatively specific to Pteriomorphia has been identified among Sox genes (*Sox-OG1/NA*). Of course, these genes deserve as much attention as their widely-distributed paralogs, as they may constitute true group-specific expansions and may play fundamental roles in some biological processes. However, they have not been discussed here or included in **Fig. 4.2** for clarity purposes, but they are freely available in supplementary materials.

Overall, our analysis clearly shows the importance of adopting a wide-angle approach when characterising the members of a gene family, especially for large ones such as the Fox genes (**Schomburg et al., 2022**). As a matter of fact, the presence of duplication events and orphan genes needs to be addressed with a broad taxonomic dataset, in order to account for possible mis-annotations, gene phylogenetic mis-placements, and sequence heterogeneity. Additionally, many reference species need to be included for the gene identification process, in order to consider distantly-related genes and obtain a solid annotation. Our gene annotation pipeline also resulted to be very solid, even with non-model organisms and sub-optimal genomic and transcriptomic resources, as they are those of bivalves. As a matter of fact, by running the same pipeline on two additional datasets composed of mammal and fruit fly genomes, we were able to obtain high-quality orthology groups in accordance with previous knowledge on the clades (**Fig. S5 and S10; Tab. S8 and S9**), with little or no manual curation needed. Furthermore, this represents also the first broad analysis of DSFGs in both mammals and fruit flies, as so far attention has been mainly dedicated to single well-studied organisms or little clades (e.g., **Jackson et al., 2010**).

4.4.2 High amino acid sequence divergence identifies putative sex-determining genes

Sex-biased genes tend to evolve more rapidly than unbiased genes at the level of their protein sequences. Accelerated rates have been observed in both male-biased genes (reviewed in **Parsch and Ellegren, 2013; Grath and Parsch, 2016**) and female-biased genes (e.g., **Papa et al., 2017; Ghiselli et al., 2018**), but also in SRGs and primary SDGs (**O’Neil and Belote, 1992; Whitfield et al., 1993; de Bono and Hodgkin, 1996**). For example, it has been shown that *Dm-W*, *Dmy*, and *Sry* (which are SDGs in the African clawed frog *X. laevis*, in the medaka fish *O. latipes*, and in eutherians, respectively) all have higher substitution rates than their paralogues (*Dmrt-1* for

Dm-W and *Dmy*, *Sox-3* for *Sry*), particularly when considering their DNA-binding domains (**Mawaribuchi et al., 2012**). Similarly, both a burst of positive selection and a relaxation of purifying selection has been detected in *Drosophila Sxl* in correspondence with its recruitment at the top of the sex-determining cascade. The same signs of relaxed purifying selection have been found in the downstream targets of *Sxl*, that is, *tra* and *dsx*, despite no evidence of positive selection has been detected (**Mullon et al., 2012**).

Considering these shared features of SRGs and SDGs, we decided to look for signs of accelerated sequence evolution in DSFGs of bivalves, in order to evaluate if any of them could be associated with SD by employing the tools of molecular evolution. However, we wanted to analyse patterns of sequence evolution not only among putative SRGs and their close paralogs (as done for Dmrt genes in **Chapter 3**), but also considering the genomic context in which these genes evolve. In fact, our aim was to check whether higher rates of sequence evolution of SRGs hold true also when compared to other genes not involved in SD and not belonging to the same gene family. To do so, we obtained the AASD median values of >11 k SCOs from bivalve genomes (**Fig. 4.3A**), in order to build a statistical distribution to be used as a reference: if SRGs/SDGs (in this case, DSFGs) truly evolve faster than other genes, we may expect them to fall within the 5% (or even 1%) upper quantile of the distribution (**Fig. 4.3B**), i.e., within highly divergent genes (Group 1 and Group 2 genes of the distribution; see **Section 4.2**). We chose to use the AASD as a metric of sequence evolution (instead of the tip-to-tip distances of phylogenetic trees, which account for more comprehensive evolutionary models) in order to save computational time. As a matter of fact, the AASD median values proved to be a good approximation of the tip-to-tip median distances in 200 randomly-selected genes (**Fig. 4.3C**; $R = 0.84, p < 2.2 \times 10^{-6}$).

Among DSFGs, three fell within the 5% upper quantile, namely *Dmrt-1L*, *Sox-H*, and *Sox-F*. Interestingly, *Dmrt-1L* and *Sox-H* have been already proposed to be involved in the male SD pathway of *C. gigas* (**inset in Fig. 4.3B**; **Zhang et al., 2014**), on the basis of DGE analyses. Specifically, *Sox-H* (*CgSoxH*) would play a major role in *C. gigas* SD, by interacting with *Dmrt-1L* (*CgDsx*) and determining the onset of the male phenotype development; at the same time, both *Sox-H* and *Dmrt-1L* would inhibit *Fox-L2* (*CgFoxL2*), which instead is necessary to start the female phenotype development. *Dmrt-1L* and *Sox-H* have been appointed several other times to be involved in male-gonad development and differentiation, through DGE (e.g.,

Teaniniuraitemoana et al., 2014; Capt et al., 2018; Afonso et al., 2019), mRNA-ISH (e.g., Naimi et al., 2009; Li, Zhang, et al., 2018; Liang et al., 2019; Yue et al., 2021) and RNA interference (mRNA-ISH; Liang et al., 2019; Sun et al., 2022). Therefore, the high AASD of *Dmrt-1L* and *Sox-H* is coherent with previous works, strengthening their role as putative SRGs.

The relationship between high gene AASD and the involvement in SD is particularly enforced when looking at the patterns of AASD in the test datasets, which corroborates the solidity of our analysis: (i) from one side, in the mammal dataset—which represents a strictly genetic SD system, thus with a master and rapidly-evolving SDG, one of the genes from the 5% upper quantile of the distribution is *Sry* (**Fig. 4.4A** and **4.4B**), the male sex-determining gene in eutherians (**inset** in **Fig. 4.4B**); (ii) from the other side, in the fruit fly dataset—which represents a chromosomal SD system, thus without any expected difference in the rates of sequence evolution among SRGs, none of the DSFG exhibit significantly high AASD (**Fig. 4.4C** and **4.4D**), including the downstream effector *dsx* (**inset** in **Fig. 4.4D**). Also *Sxl* and *tra*, both involved in the SD pathway of *Drosophila* (**inset** in **Fig. 4.4D**) do not belong to the group of highly-divergent genes, as they have a mean amino acid divergence of about 0.09 and 0.9, respectively (Group 3; **Fig. 4.4D**). Therefore, it can be argued that both *Dmrt-1L* and *Sox-H* may not only be SRGs, but may participate in bivalve SD as primary SDGs, which is reflected in their high AASD, as it is observed for *Sry* in mammals. As a matter of fact, if they were involved in SD just as intermediate actors of the signalling cascade, then we should have not observed a high AASD, as *Drosophila Sxl*, *tra*, and *dsx* seem to suggest. Overall, these patterns of molecular evolution concerning SRGs and SDG are also supported by the way SD regulatory networks evolve. As a matter of fact, it has been proposed that the sex-determining cascades tend to arise and be established with a bottom-up mechanism (**Wilkins, 1995; Mullon et al., 2012; Beukeboom and Perrin, 2014; Capel, 2017**). This means that the regulative relationships among genes at the bottom of the cascade are settled up prior to the regulative relationships among genes at the top and, consequently, upstream regulators are progressively recruited to fine-tune diverse SD signals. These evolutionary patterns eventually produce gene-regulatory networks in which the divergence of the upstream triggers is higher than that of downstream effectors, in terms of both identity and sequence composition (**Beukeboom and Perrin, 2014**). This mechanism has been proposed for *Drosophila* species (**Mullon et al., 2012**), *C. elegans* (**Stothard and Pilgrim, 2003**), and vertebrates, despite in the latter case has been questioned

several times (reviewed in **Capel, 2017**).

At this point, two main objections can be moved against our approach: (1) the distribution of AASD is not appropriate for this kind of inference, as it does not represent the true gene evolutionary (or substitution) rates (which instead are those usually employed when dealing with SRGs and SDGs); (2) the three datasets are not comparable one to each other, as they take into consideration very different animal groups, with different taxonomic rankings and different divergence times (thus, the patterns of AASD are the products of other confounding factors not directly related to SD). Concerning the first objection, we are aware that the AASD does not represent the evolutionary rate itself, but rather its product. However, the two features are tightly linked, as on the long term highly-divergent proteins tend to be produced by genes with high evolutionary (or substitution) rates (**Echave et al., 2016**). By performing a GO-term enrichment, it emerged that highly-divergent genes of the mammal dataset are mainly involved in the immune response and male spermatogenesis (**Tab. 4.1** and **S10**), which are two processes notoriously connected with rapid sequence evolution (i.e., higher evolutionary rates; **Swanson and Vacquier, 2002**; **Murat et al., 2023**; **Vinkler et al., 2023**). Similarly, highly-divergent genes from the fruit fly dataset show an enrichment for GO-terms associated with meiotic-related functions (such as the formation of the synaptonemal complex by the products of *c(2)M*, *c(3)G*, *corona*, and *corolla* genes; **Tab. 4.1** and **S10**), which again are known to be rapidly evolving (**Hemmer and Blumenstiel, 2016**). In other words, the test datasets—which include well-studied and characterised model systems, allow us to directly link the high AASD (as computed in this work) with high rates of sequence evolution (as found in previous works), as they represent well-studied and characterised model systems. This consideration can thus be extended also to the bivalve dataset: highly-divergent genes in terms of AASD, which include some DSFGs and show an enrichment for GO-terms associated to macromolecule metabolism and morphological development (**Tab. 4.1** and **S10**), are also genes with accelerated substitution rates.

Concerning the second objection, we chose two test datasets with different characteristics as we wanted to check the extent of our hypothesis, that is, molecular evolution can be used to look for putative primary SDGs in taxonomic-wide analyses. The difference in divergence times and taxonomy ranks for bivalves and therians (Late Cambrian, about 498 million years ago [Mya], **Song et al., 2023**; and Early Mesozoic, 166–123 Mya, **Álvarez-Carretero et al., 2022**, respec-

tively) seems to not influence the sequence diversity of SRGs, as both *Dmrt-1L/Sox-H* for bivalves and *Sry* for mammals exhibit high AASD with respect to their own distributions, regardless of their age. *Dmrt-1L* and *Sox-H* (which are mollusc- and Bilateria-specific, respectively) are undoubtedly older than *Sry* (which, instead, emerged in the Theria common ancestor; **Foster et al., 1992**), but each of them can be considered a highly-divergent gene in bivalves and mammals, respectively (i.e., genes that are included in the 5% upper quantile of bivalve and mammal AASD distributions). Conversely, the difference in divergence times and taxonomic ranks for *Drosophila* (Paleocene/Eocene boundary, about 56 Mya; **Russo et al., 2013**) may seem to be influencing the results for the dataset, resulting in a false negative. In other words, it can be argued that: (i) the genes included in the SD cascade of *Drosophila* (such as *Sxl*, *tra*, and *dsx*; **inset** in **Fig. 4.4D**) have a high AASD, which however has not been detected by our methodological approach (for example, this may be traced back to the young diversification age of *Drosophila* species if compared to bivalves); (ii) the species included in the analysis are all congeneric, thus the sequence differentiation of SRGs may exist not at the amino acid level but at the nucleotide one. To better disentangle this issue and further discuss the fruit fly dataset, we repeated the analysis of the AASD only on species of the *Crassostrea* genus (*C. gigas*, *Crassostrea angulata*, *Crassostrea ariakensis*, and *Crassostrea virginica*), which are much younger (Middle Cretaceous, less than 100 Mya; **Qi et al., 2023**), thus comparable to *Drosophila*. Results showed that, even when analysing a smaller bivalve dataset, encompassing only 4 species of recent origin, the high AASD of *Dmrt-1L* persists, that is, *Dmrt-1L* is still grouped together with highly-divergent genes (**Fig. S15**). The same has not been recovered for *Sox-H*, which fell in genes from Group 3 (the group corresponding to the 95% interval of the AASD distribution) but still have the second highest AASD median value among DSFGs (**Fig. S15**).

Of course we should not expect that highly-divergent genes are only those involved in SD, but may participate also in other processes (as discussed earlier and shown by GO-term enrichments; **Tab. 4.1** and **S10**). Besides the genes of interest for SD (*Dmrt-1L/Sox-H* for bivalves, and *Sry* for mammals), also other components of the DSFG families have been retrieved with a high AASD, despite they have never been linked directly to SD so far: *Sox-F* in bivalves (**Fig. 4.3B**) and *Fox-D4* in mammals (**Fig. 4.4B**). This implies that our approach can't be used to unambiguously identify SDGs alone, as high AASD is exhibited also by many other genes. Instead, the analysis is meant to be used to detect highly-divergent genes and, subsequently, by

comparison with literature and a more thorough and focused functional investigation, putative SDGs among them. In this sense, the mammal dataset exemplify the importance of putting the results of our pipeline (as those of any other comparative genomics analysis) into the correct evolutionary and genomic context: among DSFGs of mammals, two genes exhibit high AASD, one of which is directly related to SD (*Sry*), while the other has a function connected with neural development (*Fox-D4*; **Klein et al., 2013**). Thus, the high AASD may arise either because of the involvement in the upper SD pathway or because of other life-history traits connected with the gene, respectively. Regarding bivalves, *Dmrt-1L* and *Sox-H* show a sharp connection with SD as a putative primary SDG, either when considering their molecular evolutionary features or when looking at their gene expression and possible function in gonad development (**Naimi et al., 2009; Teaniniuraitemoana et al., 2014; Zhang et al., 2014; Capt et al., 2018; Li, Zhang, et al., 2018; Afonso et al., 2019; Liang et al., 2019; Yue et al., 2021**). It is difficult to further speculate on the actual involvement in SD of *Dmrt-1L* and *Sox-H* without any additional information on their biology. Nonetheless, molecular evolution proves to be a valuable tool to investigate genes putatively involved in SD, and to identify major targets onto which dedicate future research effort.

4.5 Conclusions

Genes functioning in reproductive processes, and particularly SD, are often among the most variable in animal genomes, in terms of both sequence composition and regulatory interactions (**Swanson and Vacquier, 2002; Bachtrog et al., 2014**). Such high evolutionary rates may be traced back both to adaptive evolution (either as natural or sexual selection) or to non-adaptive processes (**Vicoso and Charlesworth, 2006; Meisel and Connallon, 2013; Parsch and Ellegren, 2013; Grath and Parsch, 2016**), and often results in striking differences in reproductive and sexual systems even among closely-related species. In the present work we took advantage of this characteristic to identify SDGs in bivalves among the DSFG families. By comprehensively analysing the phylogenetic history and AASD in a broad taxonomic dataset, we appointed *Dmrt-1L* and *Sox-H* as putative SDGs, thus confirming results in previous works that found them to be transcribed in a male-biased manner and/or strongly involved in male-gonad formation (**Naimi et al., 2009; Teaniniuraitemoana et al., 2014; Zhang et al., 2014; Capt et al., 2018; Li, Zhang, et al., 2018; Afonso et al., 2019; Liang et al., 2019; Yue et al., 2021**).

Future studies would now need to further investigate their evolutionary history. For example, considering that SRGs tend to accumulate in the genomic neighbourhood where primary SDGs are located (**Capel, 2017**), analysing the genomic location of DSFGs in bivalve genomes may provide enlightening results. Similarly, revealing the genetic interactions of *Dmrt-1L* and *Sox-H*, through functional and genome editing assays, would undoubtedly benefit our understanding of their role in the sexual processes of bivalves.

Chapter 5

Localisation of three sex-related genes and the germline marker Vasa/Vasa in the early developmental stages of *Mytilus* *galloprovincialis*

Filippo Nicolini^{1,2}, Sergey Nuzhdin³, Fabrizio Ghiselli¹, Andrea Luchetti¹,
Liliana Milani¹

¹*Department of Biological, Geological and Environmental Science, University of Bologna, Bologna (BO), Italy.*

²*Fano Marine Center, Fano (PU), Italy.*

³*Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, USA.*

Manuscript in preparation.

5.1 Introduction

Despite the huge socio-economic and scientific importance of bivalves, the knowledge concerning the genetic and molecular bases of their sex determination (SD) system is scarce and overlooked (**Breton et al., 2018; Nicolini, Ghiselli, et al., 2023**). Several components of the Dmrt, Sox, and Fox gene (DSFG) families have been appointed as directly involved in SD by many works, mainly thanks to differential gene expression (DGE) analyses (e.g., **Milani et al., 2013; Zhang et al., 2014; Capt et al., 2018; Shi et al., 2018**), mRNA/protein visualisation (**Li, Liu, et al., 2018; Liang et al., 2019; Wang et al., 2020; Sun et al., 2022**), RNA interference (RNAi; **Liang et al., 2019; Wang et al., 2020; Sun et al., 2022**) and quantitative real-time polymerase chain reaction (qRT-PCR; **Li, Liu, et al., 2018; Liang et al., 2019; Wang et al., 2020; Sun et al., 2022**). For example, **Li et al. (2018)** found that *Fox-L2* and *Dmrt 1-like* (*Dmrt-1L*) are predominantly transcribed in ovaries and testes, respectively, of the Yesso scallop *Patinopacten yessoensis*, and that they contribute to establish the sexual identity of immature follicles at the molecular level prior to the morphological level. **Liang et al. (2019)** showed that *Sox-2* is involved in the differentiation of male gonads and spermatogenesis of the scallop *Chlamys farreri*, and that the knocked-out phenotype results in severe loss of both germ-cell mass and spermatogonia. **Wang et al. (2020)** speculated that *Fox-L2* is involved in the sex differentiation of female gonads in the freshwater mussel *Hyriopsis cumingii*. Overall, considerable effort has been made to characterise the transcription patterns of DSFGs of interest during the adult stage of bivalves, covering various reproductive phases, while little attention has been given to the embryo and larval stages. Nonetheless, early animal development may represent a crucial moment to the establishment of the sexual identity, as the transcription of sex-determination related genes (SRGs) and SD itself begins much earlier than the onset of gonad development and differentiation (even as early as the zygote formation; **Richardson et al., 2023**). In mammals, for example, the transcription of SRGs can be detected during the embryo preimplantation stage (before 4.5 days post fertilization [dpf]; reviewed in **Richardson et al., 2023**), while *Sex-determining region of chromosome Y* (*Sry*) realises its function as the male sex-determining gene (SDG) at 10.5 days post coitum (**Beukeboom and Perrin, 2014**). In *Drosophila melanogaster*, the early female splicing variant of *Sex-lethal* (*Sxl*)—which is the top regulator of the SD cascade and is activated by a mechanism of chromosome counting (**inset** in **Fig. 4.4D**), is transcribed during the syncytial stages of the embryo (i.e., before 2 hours post fertilization [hpf]; **Salz and Erick-**

son, 2010), when it establishes the sexual identity of the embryo through a cell-autonomous mechanism. Therefore, the study of bivalve SD necessarily requires to consider also the early stages of embryonic and larval development, in order to obtain a comprehensive scenario of the process. Among bivalves, several species may constitute a model system particularly suitable to study the SD process during embryogenesis, because of the presence of the doubly uniparental inheritance (DUI) of mitochondria. This process—which involves the uniparental transmission of the maternal and paternal mitochondrial genomes through eggs and sperm, respectively, allows for an *a-priori* detection of the sexual identity of developing embryos, as early as the first cleavage division of the zygote: in female embryos, the sperm-inherited mitochondria assume a dispersed pattern between blastomeres; conversely, in males the sperm-inherited mitochondria stay assembled together, remain within one blastomere, and are eventually included in primordial germ cells (PGCs; Zouros, 2013; Ghiselli et al., 2019).

Here we sought to expand the knowledge on the process of bivalve SD, by employing the Mediterranean mussel *Mytilus galloprovincialis* as a study system, which is a species exhibiting DUI. Particularly, we aimed to investigate the transcription patterns of three DSFG (namely, *Dmrt-1L*, *Sox-H*, and *Fox-L2*; hereafter referred to as ‘SRG’) during embryonic and early larval stages. To this purpose, (i) we first performed a time-series DGE analysis by using the RNA-sequencing data published by Miglioli et al. (2024); afterwards, (ii) we investigated the temporal and spatial transcription patterns of the DSFGs of interest through mRNA *in-situ* hybridization chain reaction (HCR). To obtain a more comprehensive developmental context for the transcription patterns of DSFGs, (iii) we also traced for the first time in *M. galloprovincialis* the process of the germline specification through mRNA *in-situ* HCR and immunolocalization of Vasa/Vasa, which is a traditionally-recognised marker of PGCs and germ cells (GCs) across Metazoa (Extavour and Akam, 2003). The specification and differentiation of GCs (which are part of the gonadal tissue in adults) is in fact a critical process in sexually reproducing multicellular organisms, as it provides the groundwork for the subsequent differentiation of sexually dimorphic gametes. Therefore, understanding the developmental pathway leading to the establishment of PGCs and GCs is essential to fully characterise the sex-determining process and how the sexual fate of PGCs/GCs is directed.

5.2 Materials and methods

5.2.1 Time-series gene expression

Miglioli et al. (2024) recently produced one of the very first detailed developmental transcriptomes of the Mediterranean mussel *M. galloprovincialis*, spanning from the unfertilized oocyte to the larval stage at 72 hpf, with time points sampled every 4 hpf. A total of thirty different mRNA libraries was sequenced, consisting of fifteen developmental time points per two biological replicates each (**Tab. S11**). These data are extremely useful to thoroughly investigate the transcription patterns of genes throughout the first three days of the *M. galloprovincialis* development, to quantify the transcription level of target genes to be investigated with mRNA HCR experiments and to have an overview of the possible outcome from such analysis.

Raw reads were downloaded from the Sequence Read Archive (SRA) in NCBI (BioProject: PRJNA996031) and trimmed using Trimmomatic v0.39 (**Bolger et al., 2014**; `LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:65`). Read quality was checked using FastQC v0.12.1 (**Andrews et al., 2010**). Trimmed reads were mapped against the *M. galloprovincialis* annotated genome (GCA_900618805.1; **Gerdol et al., 2020**) using STAR v2.7.10b (**Dobin et al., 2013**) in ‘alignReads’ mode with default parameters. The resulting gene count matrix was extracted with StringTie v2.2.1 (**Pertea et al., 2015, 2016**) in expression estimation mode followed by the python script ‘prepDE.py’ (`-1 99`).

The resulting matrix was processed in R. Raw gene counts were normalised using the median of ratios method as implemented by the ‘DESeq2’ package (**Love et al., 2014**), and then transformed through the DESeq2 variance stabilising transformation (vst). Transformed gene counts were used to run a principal component analyses (PCA) and visualise sample clustering, and to plot expression values of *Vasa*, *Dmrt-1L*, *Sox-H*, and *Fox-L2* (hereafter collectively referred to as ‘target genes’). Normalised gene counts were instead used to run a time-series DGE analysis in ‘maSigPro’ (**Conesa et al., 2006**).

The entire pipeline was automated through custom python and bash scripts, which are available in a private repository on GitHub.

5.2.2 Sample collection, MitoTracker staining and fixation

Adult mussels were hand collected from various locations surrounding the AltaSea institute at the port of Los Angeles (CA, USA). Sampling took place during the spawning season of the species in California, i.e., from Oct, 2023 to Jan, 2024.

Selected mussels were thoroughly cleaned from epibionts and placed in ice for approximately 30–60 minutes, then transferred in filtered artificial sea water (FASW) at 16 °C and acclimatised for 30 minutes. All the individuals were then placed in a common tank and spawning was induced by cyclical thermal shock, that is, by exposing mussels alternatively to FASW at 24–26 °C and 14–16 °C for a time of 30–40 min each. As soon as mussels started spawning, individuals were promptly removed from the common tank, carefully washed, air dried to remove contaminant gametes from the shell, and then allowed to continue spawning in isolated containers of about 250 mL with 16 °C FASW.

Both single and multiple crosses were performed: two males (M1, M2) and two females (F1, F2) were employed for single crosses; six males and six females were employed for multiple crosses, and gametes from the same sex were mixed. One hour after the spawning started, oocytes were filtered through a 75 over a 30 µm mesh, and aged in 1 L of FASW for 40–60 min, to allow them to assume a proper circular shape. Oocyte abundance was estimated under a stereomicroscope by eye counting the number of gametes in five aliquots of 1 mL, and then calculating the mean value. Sperm mitochondria were labelled with MitoTracker Red CMXRos (Thermo Fisher Scientific) at a working concentration of 500 nM for 30 min. MitoTracker is a fluorescent, vital and fixation-resistant mitochondrial dye and was used to detect the sex of developing embryos (as early as the two-blastomere stage) according to the distribution pattern of sperm mitochondria (**Cao et al., 2004; Obata and Komaru, 2005**). From this step onward, samples were always kept in the dark.

Fertilisation was performed by mixing oocytes and sperm at a ratio of 1:10. Fertilisation success was checked after 20–30 min by the formation of polar bodies. The suspension was then carefully washed on a 30 µm mesh to remove excess sperm, and brought to a concentration of 250 zygotes/mL. The resulting suspension was transferred into cell-culture flasks of 40 mL and embryos/larvae were reared at 16 ± 1 °C in the dark. Water was changed every 24 h. After 48 hpf, larvae were fed with the unicellular microalgae *Isochrysis galbana*, at a final concentration of about 10⁵ cells/mL following **Helm et al. (2004)**.

Embryos/larvae were sampled at 1, 2, 3 and 4 hpf, and then every 12 h until 72 hpf. Proper development and vitality were checked under a stereomicroscope at every sampling time. After concentration with a mesh of proper size, embryos/larvae were fixed in 3.2% paraformaldehyde (PFA) in 1× phosphate-buffered saline (PBS; 128 mM NaCl, 2 mM KCl, 8 mM Na₂HPO₄ · 2 H₂O, 2 mM KH₂PO₄) at 4 °C overnight under constant and gentle shaking. Fixed samples were washed 3×20 min in 1× PBS with 0.1% Tween 20 (PBS-Tw) and then dehydrated 3×30 min in absolute methanol at room temperature (RT). Dehydrated samples were stored at –20 °C until usage.

5.2.3 mRNA *in-situ* hybridization chain reaction (HCR)

HCR probe design

Vasa, *Dmrt-1L*, *Sox-H*, and *Fox-L2* spliced-transcript nucleotide sequences of *M. galloprovincialis* were obtained from the previous analyses with OrthoFinder v2.5.5 (**Emms and Kelly, 2019**) on annotated bivalve genomes and transcriptomes (see **Chapter 4**). Accession numbers of spliced transcripts are 10B017427, 10B093608, 10B014180, and 10B094018, respectively. The ‘insitu_probe_generator’ script from the Ozpolat Lab (**Kuehn et al., 2022**) was used to generate pairs of probes specifically designed for third-generation HCR (**Choi et al., 2018**). The built-in BLASTN search against the annotated *M. galloprovincialis* transcriptome was employed to check for putative off-target bindings of probe pairs. B1-488, B2-647, B3-546, and B4-700 pairs of HCR amplifiers and fluorophores were chosen, as reported in **Tab. 5.1**. Resulting probes were synthesised by Integrated DNA Technologies (IDT™) in separate oligo pools.

mRNA *in-situ* HCR and microscope imaging

mRNA *in-situ* HCR in *M. galloprovincialis* embryos was performed following **Miglioli et al. (2024)**. All the steps were carried out in the dark to prevent MitoTracker from fading. Probe hybridization buffer, probe wash buffer and amplification buffer were manufactured by Molecular Instruments, Inc.

Dehydrated samples stored in methanol were washed 4×5 min and 1×10 min in PBS-Tw. Samples were then permeabilized for 30 min in a detergent solution (1% sodium dodecyl sulphate [SDS], 0.5% Tween 20, 50 mM Tris–HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 150 mM NaCl), and washed again 2×5 min in PBS-Tw. Samples were prepared for the HCR

Table 5.1 – Characteristics of fluorescent dyes used for each labelled target. HCR amplifiers and the number of probe sets (as in **Tab. S12**) are reported when applicable. Dyes for both *Vasa* and *Vasa* are reported.

Target	Dye	HCR amplifier	HCR probe pairs	Excitation (nm)	Emission (nm)
dsDNA (nuclei)	DAPI	–	–	360	460
Sperm mitochondria	MitoTracker Red CMXRos	–	–	575	600
<i>Vasa/Vasa</i>	ALEXA-488/-488	B1/–	33/–	499	520
<i>Dmrt-1L</i>	ALEXA-647	B2	18	653	670
<i>Sox-H</i>	ALEXA-546	B3	22	557	575
<i>Fox-L2</i>	ALEXA-700	B4	28	685	700

detection stage by incubation in the probe hybridization buffer for 30 min at 37 °C. Detection stage was then performed with 4 nM of each probe set in hybridization solution overnight (>12 h) at 37 °C.

Excess probes were removed by washing 4×20 min with probe wash buffer at 37 °C and 3×5 min with 5× saline-sodium citrate (150 mM NaCl, 17 mM Na₃C₆H₅O₇) with 0.1 % Tween 20 (SSC-Tw) at RT. Samples were incubated for 30 min in the amplification buffer at RT. Hairpins were heated at 95 °C for 90 s and then snap-cooled at RT for 30 min. The amplification step of HCR was performed with 6 pmol of each hairpin in the amplification buffer overnight (>12 h) at RT.

Excess hairpins were removed by washing 2×5 min, 2×10 min, and 1×5 min with SSC-Tw. If not immediately mounted on slides, samples were stored in SSC-Tw at 4 °C. Otherwise, samples were immersed first in 50 % glycerol and then in 75 % glycerol, each for 30–60 min, and then mounted with VECTASHIELD®PLUS Antifade Mounting Medium with DAPI (H-2000). Slides were imaged on a Stellaris 5 Confocal Package system with the software Las X (Leica Microsystems). Each dye was imaged sequentially in a separate channel, to enhance the yield and avoid crosstalks. **Tab. 5.1** summarises the excitation and emission peaks for each dye. Images were then manipulated and post-produced using Fiji v2.14.0.

5.2.4 Immunolocalization of Vasa

M. galloprovincialis Vasa sequence was manually inspected through multiple sequence alignment with Vasa from other bivalves (data from **Chapter 4**) and several reference species (*Danio rerio* [Ddx4: NP_571132.1]; *Homo sapiens* [Ddx4: NP_077726.1]; *Mus musculus* [Ddx4: NP_001139357.1]; *D. melanogaster* [Vasa: NP_001260458.1]; *Caenorhabditis elegans* [GLH-1: NP_001262379.1, GLH-2: NP_491876.1, GLH-3: NP_491681.1, GLH-4: NP_491207.3]), to support commercial antibody specificity in *M. galloprovincialis*. The Vasa sequence from *D. rerio* was included as the polyclonal antibody was generated using the zebrafish protein variant (manufacturer indications; ab209710 by Abcam Limited). A maximum likelihood (ML) phylogenetic tree of Vasa genes and its paralog Ddx3 (reference genes: *D. rerio* [Ddx3Xa: NP_001119895.1, Ddx3Xb: NP_571016.2]; *H. sapiens* [DDX3X: NP_001180346.1]; *M. musculus* [Pl10/Ddx3Xl: NP_149068.1]; *D. melanogaster* [Belle: NP_001262379.1]; *C. elegans* [LAF-1: NP_001254859.1, VBH-1: NP_001021793.1]) was built using IQTREE. The hidden Markov model (HMM) profile of the Asp-Glu-Ala-Asp/Asp-Glu-Ala-His box (DEAD/DEAH-box) signature domain for the amino acid guided alignment step, was built after the corresponding Pfam full database (PF00270). Methods are the same as in **Chapter 4**.

Vasa immunolocalization in *M. galloprovincialis* embryos was performed following **Milani et al. (2011)** with modifications. All the steps were carried out in the dark to prevent MitoTracker fluorescence from fading. Dehydrated samples stored in methanol were rinsed 3×10 min and 1×2 h in 1× Tris-buffered saline (TBS; 10 mM Tris–HCl, 155 mM NaCl), following an additional wash for 10 min with PBS. Samples were then digested for 6 min and 30 s with 0.01 % pronase E (Merck) in PBS, and washed again 2×5 min in PBS. Permeabilization was performed in 1× TBS with Triton X-100 (TBS-Tx) 0.1 % for 5 min at RT and in TBS-Tx 1 % overnight at 4 °C.

After an additional rinse for 5 min in TBS-Tx 0.1 %, non-specific binding sites were blocked with a TBS-Tx 0.1 % solution containing 3 % bovine serum albumin (BSA). Samples were then incubated at 4 °C for 32–48 h with primary anti-VASA/VAS antibody (polyclonal anti-VASA developed in rabbit; ab209710 by Abcam Limited), diluted 1:100.

Excess primary antibody was rinsed from samples with 4×30 min in TBS-Tx 0.1 %, followed by an incubation of 1 h in TBS-Tx 0.1 % containing 3 % BSA. Samples were then incubated at 4 °C for 24–32 h with secondary antibody HRP anti-rabbit in goat (Santa Cruz Biotechnology Inc.), diluted 1:400.

Excess secondary antibody was rinsed with 4×30 min in TBS-Tx 0.1% and 1×1 h in 1%. Samples were immersed first in 50% glycerol and then in 75% glycerol, each for 30–60 min, and then mounted with VECTASHIELD®PLUS Antifade Mounting Medium with DAPI (H-2000). Slides were imaged on a Nikon A1R+ HD25 confocal microscope. Each dye was imaged sequentially in a separate channel, to enhance the yield and avoid any crosstalks. **Tab. 5.1** summarises the excitation and emission peaks for each dye. Images were then manipulated and post-produced using Fiji v2.14.0.

5.3 Results

5.3.1 Differential gene expression analysis of *Vasa* and SRGs in embryo time-series

Over 24 M reads were mapped for each RNA-sequencing library (86.58% of the total input reads), with an average of 26.8 M (**Tab. S11**). Of these, an average of 22 M reads were uniquely mapped (71.86% of the total input reads), while an average of 4.5 M were multi-mapped (14.72% of the total input reads). The average of unmapped reads was 4.1 M (13.42% of the total input reads; **Tab. S11**). The PCA on normalised read counts returned well-clustered experimental groups for time points between 8 and 36 hpf, while for stages before 8 hpf and after 36 hpf, experimental groups are more homogeneous among each other (**Fig. 5.1A**). This situation may reflect major developmental dynamics during embryogenesis and larval development. As a matter of fact, before 8 hpf, the embryo undergoes segmentation and no big morphogenetic movements are usually detected. Between 8 and 13 hpf, instead, gastrulation begins, the embryo experiences strong morphogenetic rearrangements (such as the formation of embryonic layers) and the trochophore larva develops, all processes which are expected to be detected also at the molecular level. After 36 hpf, instead, the larva does not show any dramatic morphogenetic event, as the D-veliger is almost formed and the gross advanced larval morphology is established. The hierarchical clustering of differentially expressed genes computed by **Miglioli et al. (2024)** is concordant with this view.

Transcription levels of *Vasa*, SRGs, *Fox-B2*, and *Wnt-8a* were plotted individually (**Fig. 5.1B**) to obtain a proxy of the expected outcome of HCR. *Fox-B2* and *Wnt-8a* were employed as control genes to get support for handling of data and of the pipeline, as they were also analysed

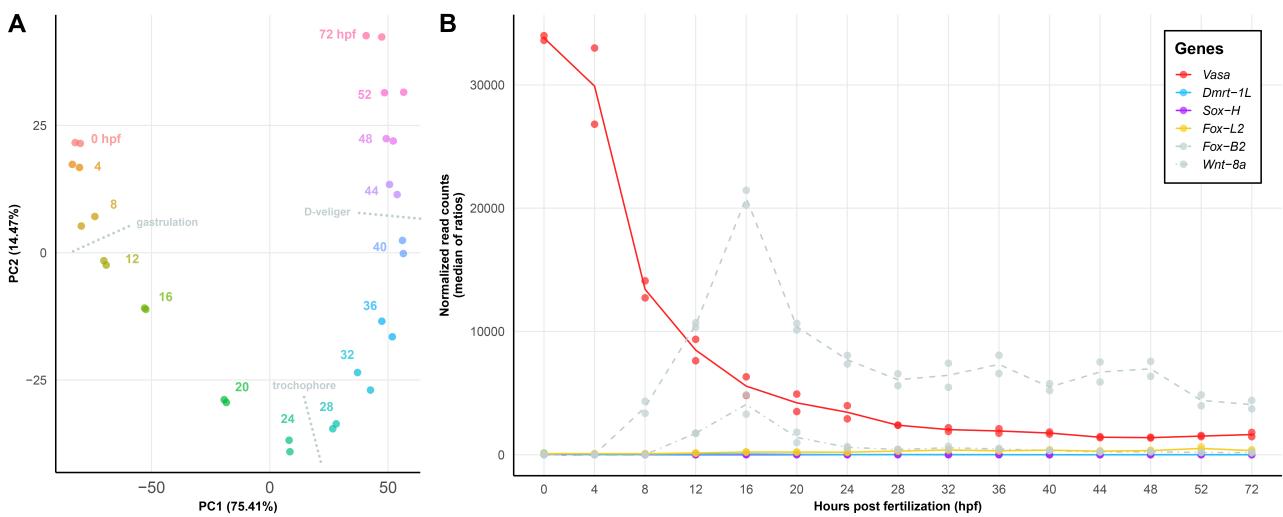


Figure 5.1 – PCA of DESeq2-normalised read counts (A) and transcription levels of target and reference genes (B). (A) Principal components (PCs) 1 and 2 are plotted in the x and y axes, respectively; the proportion of variance explained by each PC is shown in parentheses. Sampled time-points are shown in different colours and are indicated by the hours hours post fertilization (hpf). Major developmental transitions are marked with dotted lines. PCA has been performed on vst-transformed, normalised read counts (DESeq2 median of ratios). (B) Transcription levels of target (*Vasa*, *Dmrt-1L*, *Sox-H*, and *Fox-L2*) and reference genes (*Fox-B2* and *Wnt-8a*) as expressed by normalised read counts (DESeq2 median of ratios).

by Miglioli et al. (2024). The transcription of both genes starts at 4 and 8 hpf, respectively, reaches a peak at 16 hpf, and then constantly decreases (Fig. 5.1B). *Vasa* transcripts are highly abundant in unfertilized oocytes and in embryos 4 hpf, then constantly decrease throughout time; conversely, *Fox-L2* transcripts increase from 12 hpf onward (Fig. 5.1B). Both *Dmrt-1L* and *Sox-H*, instead, show low or null levels of transcriptions throughout the entire time series (Fig. 5.1B).

The maSigPro DGE analysis of the *M. galloprovincialis* developmental time series found 13,067 differentially expressed genes (about 17% of the analysed genes) and clustered them into 9 different groups, according to their specific transcriptional profiles (Fig. 5.2). Among the genes of interest, only *Vasa* and *Fox-L2* showed a significantly different transcriptional profile throughout the time series, and were included in clusters 3 and 1, respectively. As already discussed, *Vasa* and *Fox-L2* transcription levels show an opposite tendency, with the former decreasing and the latter increasing throughout time. Both *Dmrt-1L* and *Sox-H* were not found to be differentially transcribed by maSigPro and, thus, were not included in any cluster. The same holds true for *Wnt-8a* and *Fox-B2*.

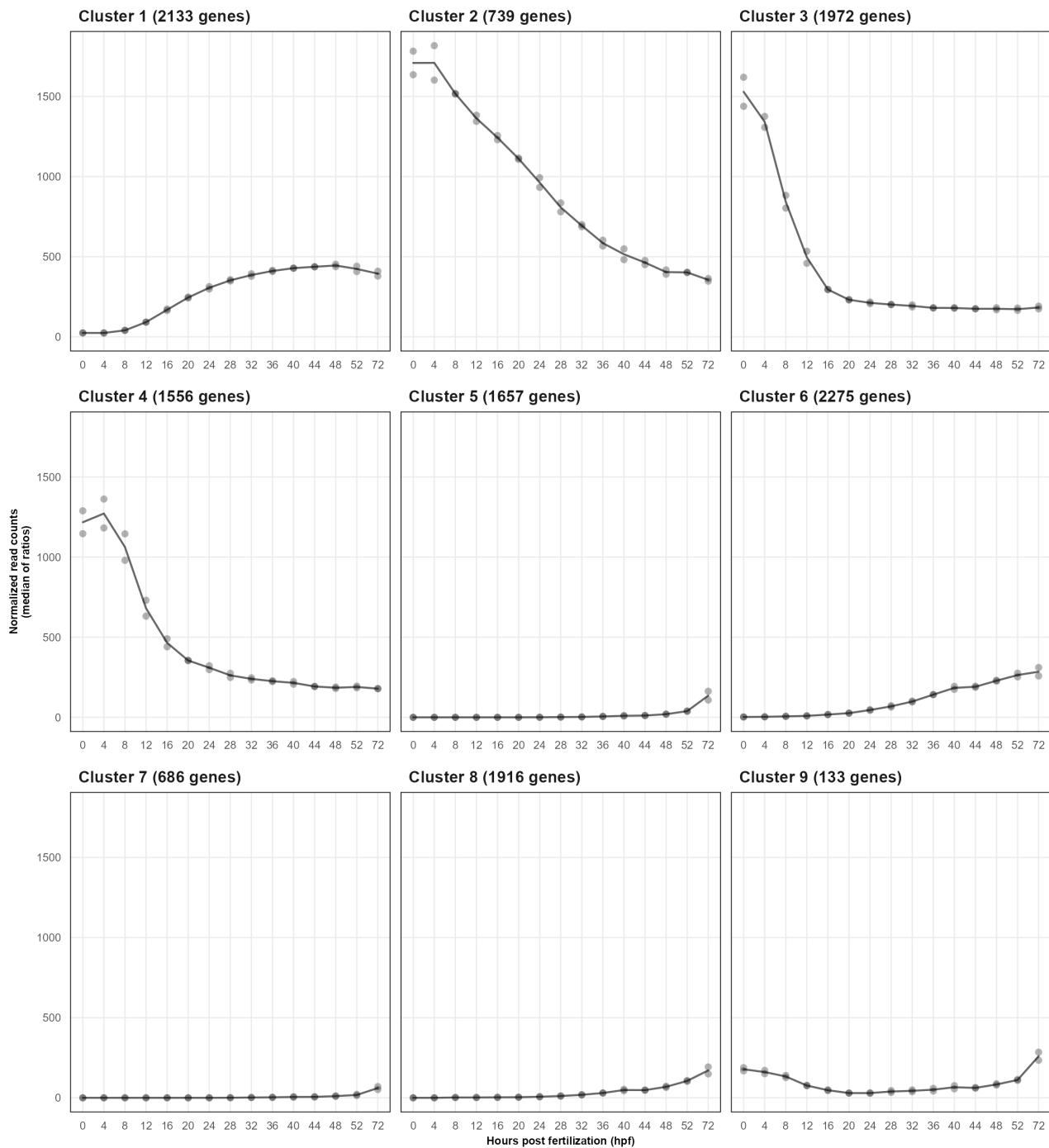


Figure 5.2 – Transcription patterns of differentially-expressed genes as inferred by maSigPro. Genes are divided into 9 different clusters according to their transcription patterns throughout 15 sampled time points. Median values of the two biological replicates are shown for each time point and represented by points. Mean values are shown for each time point and represented by solid lines. *Vasa* is included in Cluster 3, while *Fox-L2* in Cluster 1.

Table 5.2 – Number of imaged samples, divided by developmental stage, experiment, and sex.

Stage	Experiment	Females	Males	Undetermined	Total
Oocytes	HCR	–	–	–	11
2-cell embryos	HCR	8	9	1	18
4-cell embryos	HCR	9	3	0	12
8-cell embryos	HCR	11	3	0	14
12-hpf embryos	HCR	7	6	0	13
Total embryos	HCR	35	21	1	57
24-hpf larvae	HCR	0	1	11	12
48-hpf larvae	HCR	0	0	11	11
72-hpf larvae	HCR	1	1	8	10
Total larvae	HCR	1	2	30	33
Oocytes	Negative control	–	–	–	5
2-cell embryos	Negative control	7	2	0	9
4-cell embryos	Negative control	7	2	0	9
8-cell embryos	Negative control	5	1	0	6
12-hpf embryos	Negative control	0	0	3	3
Total embryos	Negative control	19	5	3	27
Total larvae	Negative control	0	0	4	4
Total imaged samples	All	55	28	38	137

5.3.2 mRNA *in-situ* HCR of Vasa and SRGs

Overall, a total of 80 adult *M. galloprovincialis* individuals were sampled and staged for thermal-shock induced spawning. Of these, 8 males and 8 females were eventually selected as parents for single (2 of each sex) and multiple (6 of each sex) crosses, on the basis of their gamete quality (i.e., presence sperm motility, and oocyte transparency and rounded shape). MitoTracker labelling was successfully retained in developing embryos of *M. galloprovincialis* until 12 hpf. After that stage, the stained sperm mitochondria were difficult to detect, and so was the dispersal pattern to establish the sexual identity.

After embryo rearing, fixation, and mRNA *in-situ* HCR of target genes, a total of 16 oocytes, 81 embryos and 33 mussel larvae were imaged (**Tab. 5.2**). Of these, on the basis of sperm mitochondria dispersal patterns, 55 were females (dispersed pattern), 28 were males (aggregated pattern) and 38 were of indeterminable sex (ambiguous pattern or unlabelled sperm mitochondria). For each stage, negative controls were also imaged (final count of 36), by staining just sperm mitochondria with MitoTracker and nuclei with DAPI, and going through the HCR protocol without adding probes in the hybridization step. A total of 137 samples were imaged.

The ‘*insitu_probe_generator*’ script (**Kuehn et al., 2022**) generated: (i) 33 probe pairs conjugated with hairpin B1 and ALEXA-488 for *Vasa*; (ii) 32 probe pairs conjugated with hairpin B2 and ALEXA-647 for *Dmrt-1L*; (iii) 27 probe pairs conjugated with hairpin B3 and ALEXA-546 for *Sox-H*; and (iv) 28 probe pairs conjugated with hairpin B4 and ALEXA-700 for *Fox-L2* (**Tab. 5.1** and **S12**). HCR labelling of genes of interest proved to be concordant with results obtained from RNA-seq analysis (see **Section 5.3.1**; **Fig. 5.1B**). Concerning *Vasa*, it has been detected throughout every sampled stage (**Fig. 5.3A**): transcripts were identified homogeneously in the cytoplasm of unfertilized oocytes, 2-, 4-, and 8-cell embryos; in gastrulae, *Vasa* is located mainly in the ingressed cells; in trophophores, it forms a cup-like structure in the region opposite to the shell-field; in D-larvae, it is mainly retained in two central areas adjacent to the valves (right and left sides of the larvae) in a sort of a comma-shaped region. Concerning *Dmrt-1L* (**Fig. 5.3B**), final images were quite noisy and showed putative non-specific staining at the level of embryo external surface and larvae shell, which may have interfered with the true signal of HCR for this gene; in any case, no clear labelling distribution pattern was found in embryos of both sexes. *Sox-H* mRNAs (**Fig. 5.3C**) were not detected during the imaged developmental stages. Conversely, *Fox-L2* transcripts have been detected starting from the 8-cell stage—where they are homogeneously present, to the D-veliger larvae—where they appear to be mostly co-localized with *Vasa* (**Fig. 5.3D** and **E**). Imaging of control samples (i.e., without mRNA *in-situ* staining) can be found in **Fig. S16**.

5.3.3 Immunolocalization of Vasa

To determine whether the commercial polyclonal antibody (ab209710 by Abcam Limited) could successfully bind *M. galloprovincialis* Vasa, we conducted a phylogenetic analysis (**Fig. 5.4A**) and a multiple sequence alignment inspection (**Fig. 5.4B**) of Vasa/Ddx4 proteins, along with its paralog Ddx3, starting from the bivalve curated genome and transcriptome dataset analysed in **Chapter 4**. We retrieved three different Vasa sequences in the *M. galloprovincialis* genome (**Tab. S13**). Two of them (VDI03911.1 and VDI03912.1) are splicing variants of the same mRNA (acc. no. 10B017427) investigated through DGE and mRNA *in-situ* HCR in previous sections. Both variants are constituted by 17 exons and differ from each other for only eight leading amino acids at the protein N-terminus (**Fig. 5.4C**). Their DEAD/DEAH-box and C-terminal domains show high levels of sequence conservation with respect to *D. rerio* Vasa (**Fig. 5.4C**). Concerning the other Vasa *M. galloprovincialis* sequence (VDI58335.1), it originates from a

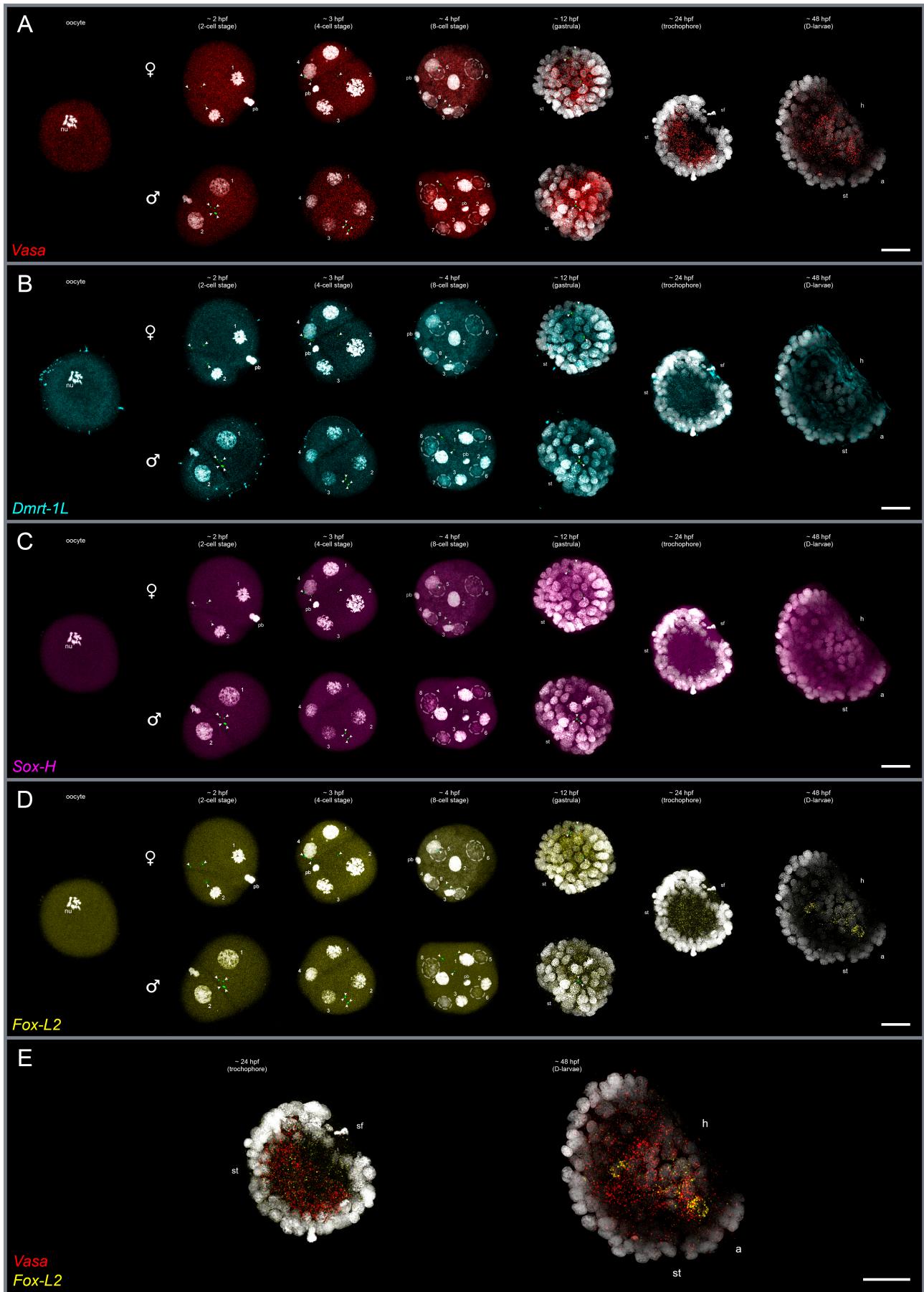


Figure 5.3 – Caption on next page.

Figure 5.3 – mRNA *in-situ* HCR of *Vasa* (A), *Dmrt-1L* (B), *Sox-H* (C), *Fox-L2* (D), and *Vasa+Fox-L2* (merged; E) in several developmental stages of *M. galloprovincialis*. Nuclei are shown in white; in the 2-, 4-, and 8-cell stages, nuclei are also marked with numbers; in the 8-cell stage, nuclei of blastomeres in the background are highlighted with dashed circles. Sperm mitochondria, when stained (shown in green), are marked with arrowheads. For 2-cell, 4-cell, 8-cell, and gastrula stages, embryos of both sexes are reported (top rows: females; bottom rows: males). a: anus; h: hinge; nu: oocyte nucleus; pb: polar body; sf: shell field; st: stomodeum. Scale bar: 20 µm. (Figure on previous page.)

separate genomic locus and appears very short (105 amino acid positions) if compared to other Vasa from bivalves (>800; data not shown). Additionally, it does not exhibit any complete DEAD/DEAH-box domain (as per CDD domain annotation). This considered, the additional Vasa gene (and its relative protein) may be an artefact due to genome mis-assembly and/or mis-annotation, or a non-functional gene. Thus, we argue that it should not affect the correct immunolocalization of the *M. galloprovincialis* Vasa proteins, and will not be further discussed.

Unfortunately, the amount of available samples and antibodies for the experiment was limited. Therefore, we managed to acquire just two oocytes, two 2-cell embryos, and one gastrula. Nonetheless, from the obtained images, we found that Vasa proteins are apparently missing from the oocytes, but can be detected at increasing levels on embryos after 4 hpf (**Fig. 5.5**) and during gastrulation, with a localization matching that of *Vasa* mRNAs (**Fig. 5.4A**). Imaging of control samples (i.e., without primary antibody reaction) can be found in **Fig. S17**.

5.4 Discussion

5.4.1 Sperm mitochondria are not detected after 12 hpf because of Mito-Tracker misincorporation or fading

Because of the presence of the unique DUI of mitochondria, *M. galloprovincialis* offers a compelling system to investigate SD during the early stages of embryogenesis. As a matter of fact, the sexual fate of an embryo appears to be established as soon as the first cleavage division, according to the dispersal pattern of sperm mitochondria (**Saavedra et al., 1997; Cao et al., 2004**): (i) if the embryo is going to develop into a female, sperm mitochondria can be found scattered across different blastomeres; (ii) if the embryo is going to develop into a male, sperm mitochondria are found aggregated all in the same blastomere (usually the macromere), being

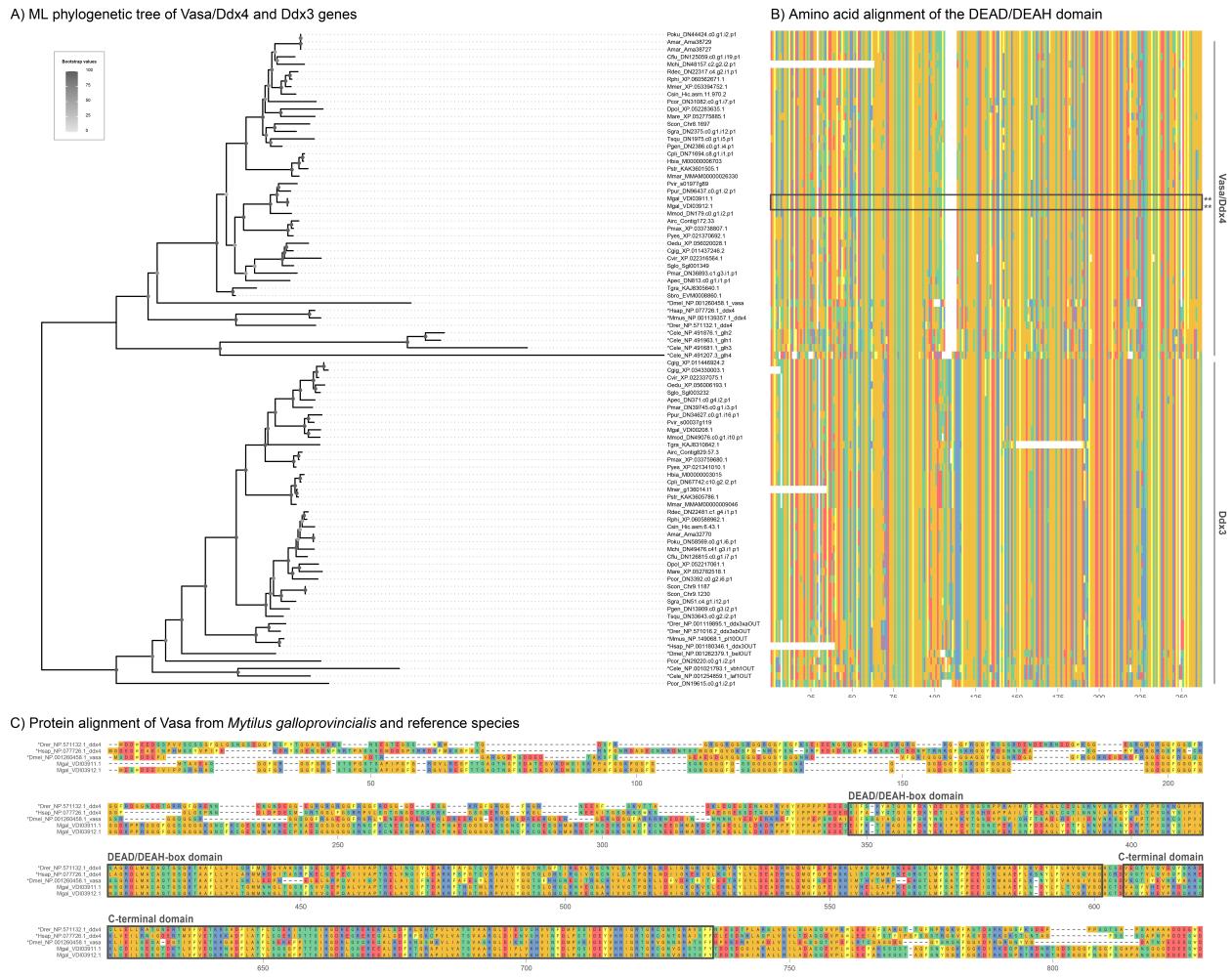


Figure 5.4 – ML phylogenetic tree of Vasa/Ddx3 and Ddx4 proteins from bivalves and reference species (A), along with the amino acid alignment of the relative DEAD/DEAH-box (B) and of Vasa proteins from *M. galloprovincialis* and reference species (C). (A) The tree has been rooted considering the Ddx4 clade as the outgroup. Reference genes from *D. rerio*, *H. sapiens*, *M. musculus*, *D. melanogaster*, and *C. elegans* are marked with an asterisk (*) at the beginning of the tip. Bootstrap values are shown for each node. (B) The alignment of the DEAD/DEAH-box is shown for each tip. The signature DEAD (Asp-Glu-Ala-Asp) motif can be found at positions 198–201. Vasa sequences from *M. galloprovincialis* are highlighted with a solid rectangle and two asterisks (**) on the right. (C) The alignment of complete Vasa sequences from *M. galloprovincialis*, *D. rerio*, *H. sapiens*, and *D. melanogaster* is shown. *C. elegans* has not been included for clarity purpose, because the species has multiple Vasa orthologs. The signature DEAD/DEAH-box and C-terminal-associated domains are highlighted with solid rectangles. Note that position coordinates are not the same between (B) and (C). Colours of amino acid residues in both (B) and (C) correspond to the ‘Chemistry_AA’ scheme from the R package ‘ggmsa’, which highlights the amino acid side-chain chemistry (Zhou et al., 2022). Bivalve species IDs as in Tab. S1. Cele: *Caenorhabditis elegans*; Drer: *Danio rerio*; Dmel: *Drosophila melanogaster*; Hsap: *Homo sapiens*; Mmus: *Mus musculus*. Full descriptions of gene names, accession numbers, and species can be found in Tab. S13.

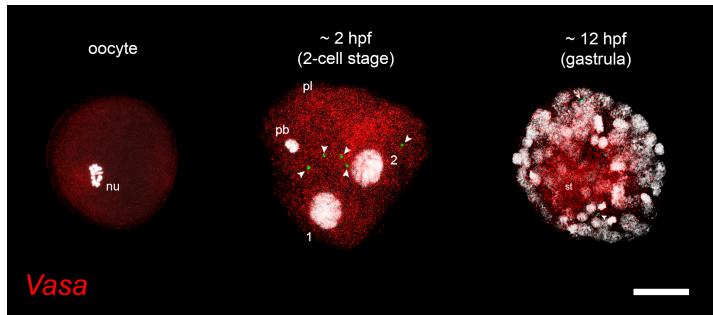


Figure 5.5 – Immunolocalization of Vasa in *M. galloprovincialis* oocyte and embryos. Nuclei are shown in white. Sperm mitochondria (in green) are marked with arrowheads. nu: oocyte nucleus; pl: polar lobe; pb: polar body; st: stomodeum. Scale bar: 20 μ m.

subsequently transferred to PGCs as part of the germ plasm. Therefore, in order to be able to establish the sexual identity of embryos and link it to any differential transcriptions of DSFGs, we labelled sperm mitochondria with MitoTracker (prior to oocyte fertilisation), and check for their dispersal patterns throughout the various sampled stages. Note that the processes of SD and DUI are not necessarily causally linked (**Kenchington et al., 2009**). Despite being successfully retained in developing embryos up until 12 hpf, the MitoTracker fluorescence was difficult to detect in later stages, and so was the mitochondrial dispersal pattern. This phenomenon may have been caused by: (i) misincorporation of MitoTracker during pre-fertilization sperm incubation; (ii) MitoTracker fading after intense manipulation of samples; (iii) MitoTracker dye being incompatible with proper embryo development, i.e., labelled embryos not surviving after 12 hpf. Based on previous studies showing that MitoTracker labelling of sperm mitochondria (including the relative dispersed and aggregated patterns) can be observed up until the late D-larva stage in *Mytilus edulis* (72 hpf; **Cao et al., 2004**), we argue that (i) and (ii) are the most likely explanation for the dye not being detected in samples after 12 hpf. Therefore, MitoTracker has not interfered with the correct development of embryos. However, it must be considered that we employed a rosamine-based MitoTracker dye (MitoTracker Red; which better resists aldehydic fixation), while **Cao et al. (2004)** used a carbocyanine-based Mito-Tracker dye (MitoTracker Green). This may have determined different effects on cell vitality, thus making results not comparable to each other. As a matter of fact, despite MitoTracker dyes are life-compatible as per manufacturer's indications, **Minamikawa et al. (1999)** showed that rosamine-based MitoTracker dyes have photosensitising effects on cells. This means that cells labelled with MitoTracker Red may be committed to apoptosis if exposed to intense light, which induce the loss of the mitochondrial membrane potential and consequent mitochondria swelling. However, based on our experimental conditions, we argue that MitoTracker Red photosensitisation has had a minimal effect, if any, on embryo development. As a matter of fact, after sperm MitoTracker staining, samples were kept in the dark throughout the entire

sampling period (MitoTracker Red cytotoxicity is not evident in the dark; **Minamikawa et al., 1999**), with limited exposure to environmental light just in correspondence with water changes; furthermore, the photosensitisation has been shown to significantly increase with light doses exceeding 0.4 J/cm^2 (cell-colony mortality rate of 60–80 % compared to control colonies; **Minamikawa et al., 1999**), which is higher than typical environmental light exposure (the solar constant is measured at around 1.362 kW/m^2 , equivalent to $0.1367\text{ J/cm}^2\text{ s}$); (iii) only a little proportion of mitochondria (the 5 sperm-derived mitochondria) have been labelled with MitoTracker, while the oocyte-derived ones remained unlabelled. Altogether, we think that MitoTracker Red staining did not determine a cytotoxic effect on *M. galloprovincialis* embryos and the consequent survival of only unlabelled embryos. Thus, we conclude that MitoTracker was not properly detected on samples older than 12 hpf because of misincorporation since the beginning or the dye fading. However, we acknowledge that a formal survival and vitality test should be performed on *M. galloprovincialis* embryos marked with MitoTracker Red, in order to exclude any possible cytotoxic effect.

5.4.2 Exploring the processes of SD in the *M. galloprovincialis* early development

To date, the molecular basis of bivalve SD has been investigated mainly in adult tissues (e.g., **Li, Zhang, et al., 2018; Liang et al., 2019; Wang et al., 2020; Sun et al., 2022; Wang et al., 2022**). As a matter of fact, considering that in many bivalve species gonads form anew at the beginning of every reproductive season from several populations of PGCs (**Filanti et al., 2021**), it can be speculated that the sexual identity may be established in correspondence with each new gonad formation. This observation would also explain the process by which many bivalve species are capable of sex changes and sex reversal from one reproductive season to the other (**Breton et al., 2018**). Nonetheless, animal SD is a key developmental process often triggered soon after fertilisation and occurring throughout the early development, as can be observed for example in mammals and fruit flies (**Salz and Erickson, 2010; Beukeboom and Perrin, 2014; Richardson et al., 2023**). Consequently, a full understanding of SD in bivalves needs to account also for the events taking place during embryo and larval life stages. To our best knowledge, the only investigation of bivalve SRGs during non-adult stages comes from the Pacific oyster *Crassostrea gigas* (**Naimi et al., 2009**), where the transcription levels of *Vasa*,

Dmrt-1L, and *Fox-L2* have been investigated through qRT-PCR. In this work, however, only stages between 7 dpf larvae and 4-month-old spats have been tested, and a direct association of the *Dmrt-1L/Fox-L2* transcription levels with SD could not be established. As a matter of fact, sexes cannot be differentiated in oysters before the onset of gametogenesis, and thus the sex of developing embryos/larvae/spats cannot be properly established (**Naimi et al., 2009**).

In this work, we aimed to expand the knowledge of bivalve SD by investigating for the first time the transcription patterns of three bivalve SDG candidates—belonging to the DSFG families, during the embryogenesis and early larval development of the Mediterranean mussel *M. galloprovincialis*. This species allows to infer sex of developing embryos by tracing the sperm mitochondria distribution patterns (see **Section 5.1; Section 5.4.1**). To this purpose, we employed an explorative investigation through a DGE analysis, and mRNA *in-situ* HCR. Our experimental setting, which included the sperm mitochondria labelling, allowed us to *a-priori* establish the sex of developing embryos and larvae, and thus to link any differential transcription pattern of DSFGs to the sexual identity.

The DGE analysis showed that the inferred transcription levels of control genes, *Fox-B2* and *Wnt-8a* (**Fig. 5.1B**), are coherent with the ones reported by **Miglioli et al. (2024)**, indicating that the results obtained from other genes can be considered reliable. The low or null transcription levels of both *Dmrt-1L* and *Sox-H* (**Fig. 5.1B**) may derive from the absence of transcription itself. However, it must be taken into account that *M. galloprovincialis* shows a mather-dependent sex ratio (**Saavedra et al., 1997**), that is, the percentage of females and males in the progeny is tightly linked to the mother’s nuclear genome, while being independent from the father’s. Thus, considering that **Miglioli et al. (2024)** do not specify the sex-ratio of the sequenced embryo pool, the possibility that the low expression levels of *Dmrt-1L* and *Sox-H* may be caused by some sex-biased related effect cannot be ruled out. Nonetheless, mRNA *in-situ* HCR supports also for our samples the scenario depicted by the DGE analysis, that is, the two genes are likely not transcribed, as no unambiguous signal was detected (**Fig. 5.3B–C**). Concerning *Dmrt-1L*, we think that an additional and more thorough investigation is needed, as the confocal imaging step seemed to have been affected by autofluorescent signals coming from the embryo surface and the larval shell, and/or by a-specific binding of probes (**Fig. 5.3B**). Thus, to obtain more reliable results, a new mRNA *in-situ* HCR experiment on *Dmrt-1L* should be designed, possibly using a set of amplifiers and fluorophores which is different from the ones

employed here (B2-647; **Tab. 5.1**).

The transcription levels of *Fox-L2* are opposed to those of *Vasa* (see **Section 5.3; Section 5.4.3; Fig. 5.1A**): the gene is not transcribed up until about 12 hpf, i.e., corresponding to gastrulation; from this stage onward, the gene transcription is detected homogeneously all over the embryo, and then becomes restricted to two regions located at both sides of the D-veliger larvae (**Fig. 5.3D**). In particular, in this stage of development, *Fox-L2* appears to be co-localised with *Vasa* (**Fig. 5.3E**), suggesting a role in PGC specification and/or differentiation. No sex-biased transcription has been detected for *Fox-L2*, though it must be considered that after 12 hpf we were not able to confidently establish the sexual identity of embryos/larvae through the localization of sperm-derived mitochondria (see **Section 5.4.1**). Thus, although it is tempting to speculate a possible role of *Fox-L2* in the specification of female gonads, as proposed by **Zhang et al. (2014)** in *C. gigas*, no definitive conclusions can be drawn at this time.

All considered, this work suggests two scenarios: (1) *Dmrt-1L*, *Sox-H*, and/or *Fox-L2* are truly SDGs, or in any case are top regulators in the bivalve SD process (as proposed by previous authors [**Zhang et al., 2014; Li, Zhang, et al., 2018**] and by the comparative genomics analysis of **Chapter 4**), though in *M. galloprovincialis* their activation (hence, SD) does not occur in early development, but in later stages; (2) *Dmrt-1L*, *Sox-H*, and *Fox-L2* are not SDGs, but are involved in gonad differentiation and maintenance in adult individuals (as found in previous works). That said, the two possibilities should not be viewed as mutually exclusive. As a matter of fact, *Dmrt-1L*, *Sox-H*, and/or *Fox-L2* may be required during *M. galloprovincialis* development for early SD, which however occur at advanced larval/spat stages, as observed in *C. gigas* (SD occurs at 40–60 dpf; **Naimi et al., 2009; Santerre et al., 2013**), but they are also required in adults to allow PGCs to initiate the sex-specific gonad development and differentiation at every reproductive season. As a matter of fact, a similar expression pattern has been seen for *M. galloprovincialis* *Vasa*, whose transcription is detectable in PGCs at a low level during the non-reproductive season, and at a high level both in immature mussels and in the reproductive season (**Obata et al., 2010**). Thus, it can be speculated that the genes triggering the SD cascade may be following similar paths.

5.4.3 Primordial germ cells are specified by both preformation and epigenesis in *M. galloprovincialis*

The process of gonad specification (including PGCs) in bivalves have been studied in several species, both in adults (e.g., **Fabioux, Pouvreau, et al., 2004; Obata et al., 2010; Filanti et al., 2021**) and during early development (**Woods, 1932; Fabioux, Huvet, et al., 2004; Kakoi et al., 2008**). In the pea clam *Sphaerium striatum* (**Woods, 1932**), the specification of the germline is traced back to the unfertilized oocyte, where the germline determinants (in the form of electron-dense granules, mainly containing mitochondria) are included in an asymmetric region of the cytoplasm (the germ plasm). The zygote segmentation then segregates the germ plasm in single blastomeres, until in the gastrula it is found only in two quiescent PGCs, which derives from the 4d blastomere (nomenclature as per **Lyons et al., 2012**). In the Pacific oyster *C. gigas* (**Fabioux, Huvet, et al., 2004**), a similar process of PGC formation has been described. The germline marker *Vasa* is found to be maternally transmitted to the embryo and deposited in a region (the germ plasm) at the vegetal pole of the oocyte. With the onset of segmentation, *Vasa* progressively segregates in single blastomeres, until it is found only in two separated cell clumps at both sides of the pericardic region in the D-larva (**Fig. 5.6A**), where they will form PGCs. The two cell clumps derive again from the 4d blastomere and, in adult oysters, they will periodically proliferate and migrate to the adjacent connective tissue to build gonad acini during the reproductive season (**Fabioux, Pouvreau, et al., 2004; Milani et al., 2017**). A different mechanism has instead been proposed for the Japanese spiny oyster *Saccostrea kegaki* (**Kakoi et al., 2008**). Here, *Vasa* is said to be transcribed all over the embryo until the 8-cell stage (data are not available on the original publication), becoming progressively more restricted to certain blastomeres only after the 50-cell stage. With gastrulation, *Vasa* is detected only at the posterior mesoderm, which derives from the 4d blastomere. In the present work, in the attempt to investigate the transcription patterns of several SDGs, we also characterised the emergence of PGCs in the early development of *M. galloprovincialis* by mean of *Vasa/Vasa* localization, thus providing an additional description of PGC development in bivalves.

According to the DGE analysis, *Vasa* shows a transcription pattern typical of maternal factors (**Xu et al., 2018**), which are stored as transcripts in the oocytes during oogenesis and then constantly decrease throughout embryo segmentation, gastrulation and early larval development, down to undetectable levels (**Fig. 5.1B**). mRNA *in-situ* HCR confirmed these results,

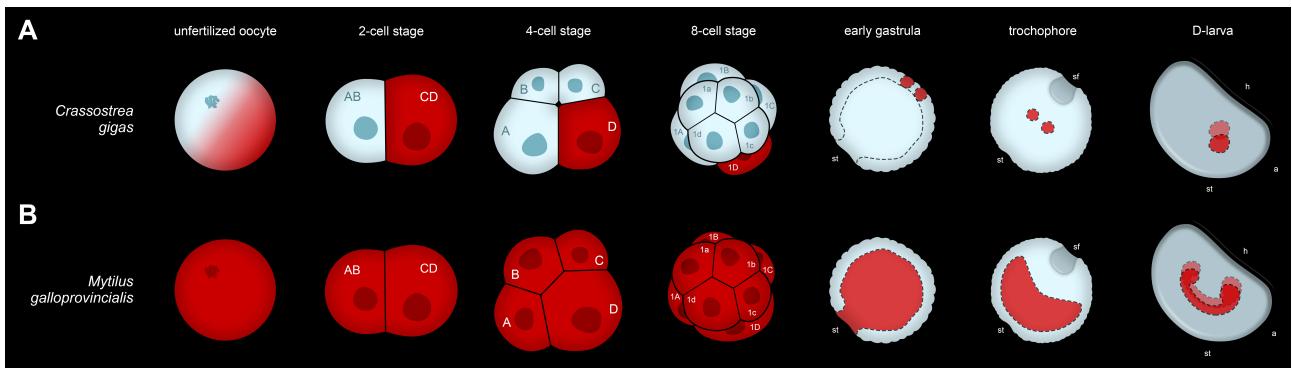


Figure 5.6 – Comparison of Vasa localization (in red) during the Pacific oyster *C. gigas* (A) and the Mediterranean mussel *M. galloprovincialis* (B) early development. Drawings not in scale. Data of *C. gigas* from **Fabioux et al. (2004)**. Blastomere nomenclature as per **Lyons et al. (2012)**. a: anus; h: hinge; st: stomodeum; sf: shell field.

showing that *Vasa* mRNA is located all over the cytoplasm of the oocyte and in all the blastomeres up until the gastrulation stage, when cells positive to *Vasa* move to constitute internal cell layers; following additional morphogenetic movements, *Vasa*-positive cells are eventually present only in two limited regions at both the lateral sides of the D-veliger (**Fig. 5.3A**). On the contrary, immunolocalization showed a different temporal distribution pattern compared to its mRNA, revealing that *Vasa* does not occur in the oocyte, but that its translation begins at low level only with segmentation of the zygote and then increases with gastrulation (**Fig. 5.5**). This different localisation is likely the result of a delay in *Vasa* mRNA translation, which is activated only during the embryo segmentation and grows with the increase of cell number. Nonetheless, this different localisation pattern of mRNA and protein may also have been determined by the differential transcription/translation of the two *Vasa* splicing variants annotated in the *M. galloprovincialis* genome, or by a non-specific binding of either the HCR DNA probes or the primary polyclonal antibody. However, considering that the two *Vasa*/Vasa variants are mostly identical, except for eight leading amino acids at the protein N-terminus (**Fig. 5.4C**), in our experiments we should have been able to target both. As a matter of fact, identifying any differential expression between the two variants would be almost impossible, either through mRNA *in-situ* HCR or immunolocalization. Accordingly, the DGE analyses failed in retrieving any dissimilarity in the transcription levels of the two splicing variants (data not shown), as the experiment was based on short-read sequencing (**Miglioli et al., 2024**). Regarding the HCR probes, they have been specifically designed on the complete *Vasa* mRNA spliced sequence, which has already been proven to specifically label PGCs in a previous analysis on *M. gallo-*

provincialis sub-adults and adults individuals (**Obata et al., 2010**). Regarding the commercial antibody, given the high sequence similarity between the Vasa proteins from *M. galloprovincialis* and *D. rerio* (the protein used to produce the antibody; see **Section 5.2**), at least in their core regions (i.e., the DEAD/DEAH-box and the C-terminal domains; **Fig. 5.4C**), we are confident that the immunolocalization procedure correctly labelled Vasa proteins; furthermore, the same set of polyclonal antibodies has been shown to successfully target PGCs/GCs in another bivalve species, *Ruditapes philippinarum*, and its specificity has been supported by Western blotting (**Filanti et al., 2021**). We thus consider our results to be strongly reliable in the correct localization of *Vasa*/Vasa.

Altogether, the present study shows a process of germline specification in the Mediterranean mussel which resembles that of *S. kegaki*, but differs from the one described in *C. gigas* (**Fig. 5.6A**) and *S. striatum*. As a matter of fact, contrary to these latter two species, in *M. galloprovincialis* *Vasa* transcripts do not form any evident gradient in the oocyte and early stages of embryogenesis (until 8-cell stage), while becoming progressively more restricted to specific cell populations only later in development (**Fig. 5.3** and **5.6B**). Particularly, with the onset of gastrulation, *Vasa*-positive cells are internalised in the developing gastrula, and *Vasa* is thus retained only by the inner cell layers. Once the embryo metamorphoses into a trochophore larva, *Vasa* transcripts arrange in a cup-like structure in the region opposite to the shell field (the ventral side), while in the early D-veliger *Vasa* is present only in two lateral regions next to the valves. Here, PGCs are going to form, eventually constituting two symmetrical linear clumps at the base of the dorsal mantle (**Obata et al., 2010**), which is going to represent the primary source of stem cells for gonad acini formation at every reproductive season (**Obata et al., 2010**). This mechanism is also reflected in the localization of Vasa proteins, which do not show any clear gradient in the oocyte (from which is absent) and at least up until gastrulation. Altogether, these findings suggest that in the Mediterranean mussel, *Vasa*/Vasa may segregate in the PGCs not only because of their inheritance as maternal factors (through preformation), but also in response to some external (and unknown) zygotic signal (through epigenesis). Therefore, *Vasa* alone do not allow to identify the presumptive primordial germ cells (pPGCs) or the PGCs during the earliest stages of development, as instead it has been shown in adult individuals (i.e., upon PGC formation; **Obata et al., 2010**). Given that *Vasa*/Vasa mark a population of cells instead of few blastomeres (those constituting the pPGCs), *Vasa* may consequently play a role also in the broader field of stem cell specification during embryo-

genesis, as shown in the marine polychaete *Platynereis dumerilii* (**Rebscher et al., 2007**) and the sea urchin *Strongylocentrotus purpuratus* (**Voronina et al., 2008**). In these two species, the germline is specified via an intermediate process relying on both preformation and epigenesis, which can be considered a ‘two-step process’ (**Rebscher et al., 2007; Kumano, 2015**). In this model, a lineage of pluripotent stem cells (PSCs) first segregates during early embryogenesis, and then produces PGCs and other mesodermal somatic structures by unequal cell division. The germline markers, including *Vasa*, are thus localised in both the PSCs and in the descendant PGCs. A similar pattern of *Vasa* localisation (i.e., ubiquitously present in oocytes and during early cleavage of the embryo, then progressively restricted to specific blastomeres) has also been shown in the snail *Ilyanassa obsoleta* (**Swartz et al., 2008**) and the abalone *Haliotis asinina* (**Kranz et al., 2010**), despite not being directly linked to PSC specification.

5.5 Conclusion

In the present work, we hypothesise that the PGC specification in *M. galloprovincialis* follows a two-step process (i.e., a combination of both preformation and epigenesis), which involves the PGCs to be formed only after embryogenesis (**Kumano, 2015**). A similar process may also be hypothesised for *S. kegaki* (**Kakoi et al., 2008**). This mechanism may explain why *Dmrt-1L* and *Sox-H*, if confirmed as SDGs, are not transcribed during the investigated developmental stages. SD would in fact happen only upon PGC commitment, thus during advanced larval development. The present work represents the first attempt to characterise the spatial localisation of three DSFGs in the Mediterranean mussel embryonic and larval development, along with *Vasa/Vasa*, and proves the importance of considering also the developmental stages when investigating new species in a comparative framework. Adopting such an evolutionary developmental perspective may in fact reveal new processes and patterns in animal biology, even when considering closely-related species. As a matter of fact, on the basis of available studies, it has been previously proposed that PGC specification is generally based on epigenesis in gastropods and preformation in bivalves (**Obata and Komaru, 2012**), even though the underlying mechanisms may be species-specific (**Obata and Komaru, 2012**). However, the model represented by the Mediterranean mussel showed that the process of PGC specification may be more diverse in bivalves than expected: preformation happens in *C. gigas* and *S. striatum*, and the two-step process happens instead in *S. kegaki*, and, as we propose, in *M. galloprovincialis*. Therefore,

results provided by the present work support the idea that the traditional preformation and epigenesis should not be accounted as mutually-exclusive phenomena nor as the only mechanisms of PGC formation (**Extavour, 2007; Kumano, 2015**). Clearly, given the unavailability of any PGC marker in bivalve embryos and larvae (and in mollusc in general), at the moment it is not possible to unambiguously establish the emergence and commitment of PGCs during embryogenesis (**Rebscher, 2014**), especially if based only on few germline genes. As a matter of fact, PGCs may share certain genetic markers (e.g., *Vasa*, *Nanos*, *Piwi*, and *Pl-10*) also with some stem cell lineages (**Extavour and Akam, 2003; Extavour, 2007; Rebscher et al., 2007; Voronina et al., 2008; Rebscher, 2014; Piccinini and Milani, 2023**). Thus, more comprehensive investigations are needed to fully and unambiguously characterise the emergence of the germline in bivalve embryos, for example through the examination of the histological and cytological morphology and of genetic regulations (**Extavour and Akam, 2003**). A similar scenario holds true also for SD and SDGs. In fact, we should not expect that the sex-determining process, together with its underlying gene regulatory networks and the timing of its expression, is the same across the entire bivalve diversity. SD is indeed one of the most variable developmental processes, despite its importance in the morphological development of an organism (**Capel, 2017**). However, it can be expected that the main actors, being them genetics or environmental or of multiple origin, are conserved, at least in having a role along the whole SD process (**Capel, 2017**). Future studies would thus need to further address the functions of the main DSFG candidates, as well as the modes of SD and germline development, through cutting-edge techniques (such as single-cell RNA-sequencing) and possibly also encompassing various life stages. In this sense, it is tempting to consolidate the role of the Mediterranean mussel as a model system for SD and germline studies by taking advantage not only of the DUI of mitochondria as a proxy for the sexual identity, but also of the ability of the species to produce a sex-biased offspring with the sole maternal influence (**Saavedra et al., 1997**). This would allow a more thorough and straightforward investigation of the determinants influencing the sex and germline specification of developing mussels, by means of targeted RNA-sequencing and transcript/protein localisation.

Chapter 6

Conclusions

The main objective of this PhD thesis was to investigate bivalve sex determination (SD) through the lens of evolutionary and integrative biology. Bivalves is a group of animals characterised by highly heterogeneous sexual and reproductive modes, with strictly gonochoristic species, obligate and facultative hermaphrodites (either protandrous, protogynous and bidirectional), as well as androgenetic systems. Both genetic and environmental factors seem to influence the sexual identity, at various degrees according to the species, and heteromorphic sex chromosomes (HeSCs) seem to have not been selected throughout the bivalve evolutionary history. Therefore, a rigorous comparative approach is essential to unravel the extreme complexity that regulates bivalve SD. Particularly, by combining bioinformatics with *wet-lab* techniques, including genomics, phylogenetics, molecular evolution analyses, differential gene expression (DGE), mRNA *in-situ* hybridization chain reaction (HCR), and immunolocalization, this work lays the foundation to understanding how sex-determination related genes (SRGs), with a special focus on the Dmrt, Sox, and Fox gene (DSFG) families, have evolved and may function in sex-determining processes across the bivalve taxonomic diversity.

In **Chapter 3**, the emerging role of bivalves as model organisms for SD studies has been emphasised through a critical examination of the current knowledge. The complexity of bivalve reproduction and sexual systems underscored the need to view SD not as a binary and stationary process, but rather as a highly dynamic continuum influenced by multiple genetic and environmental factors. Adopting this broader perspective will allow for a more effective investigation of the biology of SD.

In **Chapter 4**, the molecular evolution of SRGs across a range of bivalve species was analysed. The findings revealed patterns of accelerated amino acid sequence divergence (AASD)

in key SRGs, namely *Dmrt 1-like* (*Dmrt-1L*) and *Sox-H*, supporting the hypothesis that these genes are deeply involved in SD mechanisms, possibly even as primary sex-determining genes (SDGs). Thanks to a comparative study which encompassed the analysis of additional control datasets—mammals and *Drosophila*, the validity of the results has been confirmed and discussed in the light of a broader framework. This comparative approach allowed for the identification of evolutionary convergences and divergences, advancing our understanding of the patterns of molecular evolution in animal SRGs.

Chapter 5 focused on gene expression studies in the Mediterranean mussel *Mytilus gallo-provincialis*, offering insights into the SD process in early development. Particularly, *Dmrt-1L* and *Sox-H* appear to not be expressed during these stages, while *Fox-L2* transcription starts only with the onset of gastrulation. This suggests that either these genes are not top regulators of SD, or that SD occurs only later in development, thus their expression is not found during the analysed stages. The latter interpretation would be in line with the pattern of primordial germ cell (PGC) specification in *M. galloprovincialis*, which begins only in correspondence with the onset of gastrulation, thus not following a strict preformation model as in other studied bivalves.

Overall, this thesis further demonstrates that bivalves, with their vast reproductive and sexual diversity, serve as ideal models for investigating the complexity of SD. By integrating genomic analyses with developmental biology, this work provides a new framework for understanding how SRG evolve and function in diverse species. Future studies could build on these insights by exploring the functional roles of SRGs through other advanced techniques (such as CRISPR-Cas9), thus expanding our understanding of the genetic underpinnings of SD and differentiation. This integrative approach has the potential to unlock new knowledge not only in bivalves but across a wide array of species, deepening our understanding of the evolutionary forces shaping reproductive biology in animals.

References

- Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(suppl_2), W7–W13.
- Abbott, J. K. (2011). Intra-locus sexual conflict and sexually antagonistic genetic variation in hermaphroditic animals. *Proceedings of the Royal Society B: Biological Sciences*, 278(1703), 161–169.
- Afonso, L. F., Americo, J. A., Soares-Souza, G. B., Torres, A. L. Q., Wajsenzon, I. J. R., & Rebelo, M. d. F. (2019). Gonad transcriptome of golden mussel *Limnoperna fortunei* reveals potential sex differentiation genes. *bioRxiv*, 818757.
- Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C. W., & Rokhsar, D. S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, 524(7564), 220–224.
- Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv*, 27, 1–26.
- Allen, S. L., Delaney, E. K., Kopp, A., & Chenoweth, S. F. (2017). Single-molecule sequencing of the *Drosophila serrata* genome. *G3: Genes, Genomes, Genetics*, 7(3), 781–788.
- Altenhoff, A. M., Warwick Vesztrocy, A., Bernard, C., Train, C.-M., Nicheperovich, A., Prieto Baños, S., Julca, I., Moi, D., Nevers, Y., Majidian, S., et al. (2024). OMA orthology in 2024: Improved prokaryote coverage, ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA Ecosystem. *Nucleic Acids Research*, 52(D1), D513–D521.

- Álvarez-Carretero, S., Tamuri, A. U., Battini, M., Nascimento, F. F., Carlisle, E., Asher, R. J., Yang, Z., Donoghue, P. C., & Dos Reis, M. (2022). A species-level timeline of mammal evolution integrating phylogenomic data. *Nature*, 602(7896), 263–267.
- Andrews, S., et al. (2010). FastQC: A quality control tool for high throughput sequence data.
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, 12(7), e1001899.
- Bai, C.-M., Xin, L.-S., Rosani, U., Wu, B., Wang, Q.-C., Duan, X.-K., Liu, Z.-H., & Wang, C.-M. (2019). Chromosomal-level assembly of the blood clam, *Scapharca (Anadara) broughtonii*, using long sequence reads and Hi-C. *GigaScience*, 8(7), giz067.
- Baldwin-Brown, J. G., Weeks, S. C., & Long, A. D. (2018). A new standard for crustacean genomes: The highly contiguous, annotated genome assembly of the clam shrimp *Eulimnadia texana* reveals HOX gene order and identifies the sex chromosome. *Genome Biology and Evolution*, 10(1), 143–156.
- Baral, S., Arumugam, G., Deshmukh, R., & Kunte, K. (2019). Genetic architecture and sex-specific selection govern modular, male-biased evolution of *doublesex*. *Science Advances*, 5(5), eaau3753.
- Bear, A., & Monteiro, A. (2013). Both cell-autonomous mechanisms and hormones contribute to sexual development in vertebrates and insects. *BioEssays*, 35(8), 725–732.
- Benayoun, B. A., Caburet, S., & Veitia, R. A. (2011). Forkhead transcription factors: Key players in health and disease. *Trends in Genetics*, 27(6), 224–232.
- Beukeboom, L. W., & Perrin, N. (2014). *The evolution of sex determination*. Oxford University Press.
- Bewick, A. J., Anderson, D. W., & Evans, B. J. (2011). Evolution of the closely related, sex-related genes *DM-W* and *DMRT1* in African clawed frogs (*Xenopus*). *Evolution*, 65(3), 698–712.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Bopp, D., Saccone, G., & Beye, M. (2014). Sex determination in insects: Variations on a common theme. *Sexual Development*, 8(1-3), 20–28.
- Bredemeyer, K. R., Hillier, L., Harris, A. J., Hughes, G. M., Foley, N. M., Lawless, C., Carroll, R. A., Storer, J. M., Batzer, M. A., Rice, E. S., et al. (2023). Single-haplotype com-

- parative genomics provides insights into lineage-specific structural variation during cat evolution. *Nature Genetics*, 55(11), 1953–1963.
- Breton, S., Capt, C., Guerra, D., & Stewart, D. (2018). Sex-Determining mechanisms in bivalves. In J. L. Leonard (Ed.), *Transitions Between Sexual Systems: Understanding the Mechanisms of, and Pathways Between, Dioecy, Hermaphroditism and Other Sexual Systems* (pp. 165–192). Springer International Publishing.
- Breton, S., Stewart, D. T., Brémaud, J., Havird, J. C., Smith, C. H., & Hoeh, W. R. (2022). Did doubly uniparental inheritance (DUI) of mtDNA originate as a cytoplasmic male sterility (CMS) system? *BioEssays*, 44(4), 2100283.
- Briones, C., Nuñez, J. J., Pérez, M., Espinoza-Rojas, D., Molina-Quiroz, C., & Guiñez, R. (2018). De novo male gonad transcriptome draft for the marine mussel *Perumytilus purpuratus* with a focus on its reproductive-related proteins. *Journal of Genomics*, 6, 127.
- Calcino, A. D., de Oliveira, A. L., Simakov, O., Schwaha, T., Zieger, E., Wollesen, T., & Wanninger, A. (2019). The quagga mussel genome and the evolution of freshwater tolerance. *DNA Research*, 26(5), 411–422.
- Cao, L., Kenchington, E., & Zouros, E. (2004). Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). *Genetics*, 166(2), 883–894.
- Capel, B. (2017). Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*, 18(11), 675–689.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973.
- Capt, C., Bouvet, K., Guerra, D., Robicheau, B. M., Stewart, D. T., Pante, E., & Breton, S. (2020). Unorthodox features in two venerid bivalves with doubly uniparental inheritance of mitochondria. *Scientific Reports*, 10(1), 1087.
- Capt, C., Renaut, S., Ghiselli, F., Milani, L., Johnson, N. A., Sietman, B. E., Stewart, D. T., & Breton, S. (2018). Deciphering the link between doubly uniparental inheritance of mtDNA and sex determination in bivalves: Clues from comparative transcriptomics. *Genome Biology and Evolution*, 10(2), 577–590.
- Chakraborty, M., Chang, C.-H., Khost, D. E., Vedanayagam, J., Adrión, J. R., Liao, Y., Mon-tooth, K. L., Meiklejohn, C. D., Larracuente, A. M., & Emerson, J. (2021). Evolution of

- genome structure in the *Drosophila simulans* species complex. *Genome Research*, 31(3), 380–396.
- Chávez-Villalba, J., Hernández-Ibarra, A., López-Tapia, M. R., & Mazón-Suástegui, J. M. (2008). Prospective culture of the Cortez oyster *Crassostrea corteziensis* from north-western Mexico: Growth, gametogenic activity, and condition index. *Journal of Shellfish Research*, 27(4), 711–720.
- Chen, H., Xiao, G., Chai, X., Lin, X., Fang, J., & Teng, S. (2017). Transcriptome analysis of sex-related genes in the blood clam *Tegillarca granosa*. *PLoS One*, 12(9), e0184584.
- Choi, H. M., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation *in situ* hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development*, 145(12), dev165753.
- Civetta, A., & Singh, R. S. (1998). Sex-related genes, directional sexual selection, and speciation. *Molecular Biology and Evolution*, 15(7), 901–909.
- Collin, R. (2013). Phylogenetic patterns and phenotypic plasticity of molluscan sexual systems. *Integrative and Comparative Biology*, 53(4), 723–735.
- Conesa, A., Nueda, M. J., Ferrer, A., & Talón, M. (2006). MaSigPro: A method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics*, 22(9), 1096–1102.
- Corrochano-Fraile, A., Davie, A., Carboni, S., & Bekaert, M. (2022). Evidence of multiple genome duplication events in *Mytilus* evolution. *BMC Genomics*, 23(1), 340.
- Dainat, J., Hereñú, D., Davis, E., Crouch, K., Sol, L., & Agostinho, N. (2022). Another Gff analysis toolkit to handle annotations in any GTF/GFF format (Version v1. 0). *Zenodo*. doi, 10.
- Dalpé, A., Lalonde-Larue, A., Choquette, T., Cimmino, J., Leroux, É., Leung, C., Stewart, D. T., Angers, B., & Breton, S. (2022). The influence of environmental conditions on sex determination in the blue mussel *Mytilus edulis*. *ICES Journal of Marine Science*, 79(2), 394–402.
- de Bono, M., & Hodgkin, J. (1996). Evolution of sex determination in *Caenorhabditis*: Unusually high divergence of *tra-1* and its functional consequences. *Genetics*, 144(2), 587–595.
- Deng, T., Pang, C., Lu, X., Zhu, P., Duan, A., Tan, Z., Huang, J., Li, H., Chen, M., & Liang, X. (2016). *De novo* transcriptome assembly of the Chinese swamp buffalo by RNA sequencing and SSR marker discovery. *PLoS One*, 11(1), e0147132.

- Dermauw, W., Van Leeuwen, T., & Feyereisen, R. (2020). Diversity and evolution of the P450 family in arthropods. *Insect Biochemistry and Molecular Biology*, *127*, 103490.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.
- Dong, C., Zhang, L., Xia, S., Sosa, D., Arsala, D., & Long, M. (2022). New gene evolution with subcellular expression patterns detected in PacBio-sequenced genomes of *Drosophila* genus. *bioRxiv*, 2022–11.
- Eads, B. D., Colbourne, J. K., Bohuski, E., & Andrews, J. (2007). Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genomics*, *8*, 1–14.
- Echave, J., Spielman, S. J., & Wilke, C. O. (2016). Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics*, *17*(2), 109–121.
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10), e1002195.
- Elbers, J. P., Rogers, M. F., Perelman, P. L., Proskuryakova, A. A., Serdyukova, N. A., Johnson, W. E., Horin, P., Corander, J., Murphy, D., & Burger, P. A. (2019). Improving Illumina assemblies with Hi-C and long reads: An example with the North African dromedary. *Molecular Ecology Resources*, *19*(4), 1015–1026.
- Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*(9), 689–698.
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 1–14.
- Evensen, K. G., Robinson, W. E., Krick, K., Murray, H. M., & Poynton, H. C. (2022). Comparative phylotranscriptomics reveals putative sex differentiating genes across eight diverse bivalve species. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *41*, 100952.
- Extavour, C. G., & Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: Epigenesis and preformation. *Development*, *130*(24), 5869–5884.
- Extavour, C. G. (2007). Evolution of the bilaterian germ line: Lineage origin and modulation of specification mechanisms. *Integrative and Comparative Biology*, *47*(5), 770–785.

- Fabioux, C., Huvet, A., Lelong, C., Robert, R., Pouvreau, S., Daniel, J.-Y., Minguant, C., & Le Pennec, M. (2004). Oyster *vasa*-like gene as a marker of the germline cell development in *Crassostrea gigas*. *Biochemical and Biophysical Research Communications*, 320(2), 592–598.
- Fabioux, C., Pouvreau, S., Le Roux, F., & Huvet, A. (2004). The oyster *vasa*-like gene: A specific marker of the germline in *Crassostrea gigas*. *Biochemical and Biophysical Research Communications*, 315(4), 897–904.
- Fan, H., Wu, Q., Wei, F., Yang, F., Ng, B. L., & Hu, Y. (2019). Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biology*, 20, 1–12.
- Farhat, S., Bonnivard, E., Pales Espinosa, E., Tanguy, A., Boutet, I., Guiglielmoni, N., Flot, J.-F., & Allam, B. (2022). Comparative analysis of the *Mercenaria mercenaria* genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks. *BMC Genomics*, 23(1), 192.
- Filanti, B., Piccinini, G., Bettini, S., Lazzari, M., Franceschini, V., Maurizii, M. G., & Milani, L. (2021). Early germline differentiation in bivalves: TDRD7 as a candidate investigational unit for *Ruditapes philippinarum* germ granule assembly. *Histochemistry and Cell Biology*, 156, 19–34.
- Foote, A. D., Liu, Y., Thomas, G. W., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C. E., Hunter, M. E., Joshi, V., et al. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics*, 47(3), 272–275.
- Foster, J. W., Brennan, F. E., Hampikian, G. K., Goodfellow, P. N., Sinclair, A. H., Lovell-Badge, R., Selwood, L., Renfree, M. B., Cooper, D. W., & Marshall Graves, J. A. (1992). Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. *Nature*, 359(6395), 531–533.
- Fouret, J., Brunet, F. G., Binet, M., Aurine, N., Enchéry, F., Croze, S., Guinier, M., Goumaidi, A., Preininger, D., Volff, J.-N., et al. (2020). Sequencing the genome of Indian flying fox, natural reservoir of nipah virus, using hybrid assembly and conservative secondary scaffolding. *Frontiers in Microbiology*, 11, 1807.
- Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M. A., Murgarella, M., Greco, S., et al. (2020). Massive gene presence-

- absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology*, 21, 1–21.
- Ghiselli, F., Iannello, M., Piccinini, G., & Milani, L. (2021). Bivalve molluscs as model systems for studying mitochondrial biology. *Integrative and Comparative Biology*, 61(5), 1699–1714.
- Ghiselli, F., Iannello, M., Puccio, G., Chang, P. L., Plazzi, F., Nuzhdin, S. V., & Passamonti, M. (2018). Comparative transcriptomics in two bivalve species offers different perspectives on the evolution of sex-biased genes. *Genome Biology and Evolution*, 10(6), 1389–1402.
- Ghiselli, F., Maurizii, M. G., Reunov, A., Ariño-Bassols, H., Cifaldi, C., Pecci, A., Alexandrova, Y., Bettini, S., Passamonti, M., Franceschini, V., et al. (2019). Natural heteroplasmy and mitochondrial inheritance in bivalve molluscs. *Integrative and Comparative Biology*, 59(4), 1016–1032.
- Ghiselli, F., Milani, L., Guerra, D., Chang, P. L., Breton, S., Nuzhdin, S. V., & Passamonti, M. (2013). Structure, transcription, and variability of metazoan mitochondrial genome: Perspectives from an unusual mitochondrial inheritance system. *Genome Biology and Evolution*, 5(8), 1535–1554.
- Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C., & Froufe, E. (2020). Molluscan genomics: The road so far and the way forward. *Hydrobiologia*, 847(7), 1705–1726.
- Gomes-dos-Santos, A., Lopes-Lima, M., Machado, A. M., Marcos Ramos, A., Usié, A., Bolotov, I. N., Vikhrev, I. V., Breton, S., Castro, L. F. C., da Fonseca, R. R., et al. (2021). The Crown Pearl: A draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). *DNA Research*, 28(2), dsab002.
- Gomes-dos-Santos, A., Lopes-Lima, M., Machado, M. A., Teixeira, A., C. Castro, L. F., & Froufe, E. (2023). PacBio Hi-Fi genome assembly of the Iberian dolphin freshwater mussel *Unio delphinus* Spengler, 1793. *Scientific Data*, 10(1), 340.
- Gómez-Chiarri, M., Warren, W. C., Guo, X., & Proestou, D. (2015). Developing tools for the study of molluscan immunity: The sequencing of the genome of the eastern oyster, *Crassostrea virginica*. *Fish & Shellfish Immunology*, 46(1), 2–4.
- González, V. L., Andrade, S. C., Bieler, R., Collins, T. M., Dunn, C. W., Mikkelsen, P. M., Taylor, J. D., & Giribet, G. (2015). A phylogenetic backbone for Bivalvia: An RNA-seq approach. *Proceedings of the Royal Society B: Biological Sciences*, 282(1801), 20142332.

- Grath, S., & Parsch, J. (2016). Sex-biased gene expression. *Annual Review of Genetics*, 50, 29–44.
- Grau-Bové, X., & Sebé-Pedrós, A. (2021). Orthology clusters from gene trees with Possvm. *Molecular Biology and Evolution*, 38(11), 5204–5208.
- Gusman, A., Lecomte, S., Stewart, D. T., Passamonti, M., & Breton, S. (2016). Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. *PeerJ*, 4, e2760.
- Haag, E. S., & Doty, A. V. (2005). Sex determination across evolution: Connecting the dots. *PLoS Biology*, 3(1), e21.
- Habtewold, T., Wagah, M., Tambwe, M. M., Moore, S., Windbichler, N., Christophides, G., Johnson, H., Heaton, H., Collins, J., Krasheninnikova, K., et al. (2023). A chromosomal reference genome sequence for the malaria mosquito, *Anopheles gambiae*, Giles, 1902, Ifakara strain. *Wellcome Open Research*, 8.
- Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A., et al. (2007). Evolution in the fast lane: Rapidly evolving sex-related genes in *Drosophila*. *Genetics*, 177(3), 1321–1335.
- Han, F., Wang, Z., Wu, F., Liu, Z., Huang, B., & Wang, D. (2010). Characterization, phylogeny, alternative splicing and expression of *Sox30* gene. *BMC Molecular Biology*, 11, 1–11.
- Han, W., Liu, L., Wang, J., Wei, H., Li, Y., Zhang, L., Guo, Z., Li, Y., Liu, T., Zeng, Q., et al. (2022). Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. *Nature Ecology & Evolution*, 1–16.
- Hart, S. F., Yonemitsu, M. A., Giersch, R. M., Garrett, F. E., Beal, B. F., Arriagada, G., Davis, B. W., Ostrander, E. A., Goff, S. P., & Metzger, M. J. (2023). Centuries of genome instability and evolution in soft-shell clam, *Mya arenaria*, bivalve transmissible neoplasia. *Nature Cancer*, 4(11), 1561–1574.
- Haszprunar, G., & Wanninger, A. (2012). Molluscs. *Current Biology*, 22(13), R510–R514.
- Heenan, P., Zondag, L., & Wilson, M. J. (2016). Evolution of the Sox gene family within the chordate phylum. *Gene*, 575(2), 385–392.
- Helm, M. M., Bourne, N., & Lovatelli, A. (2004). *Hatchery culture of bivalves: A practical manual*.

- Hemmer, L. W., & Blumenstiel, J. P. (2016). Holding it together: Rapid evolution and positive selection in the synaptonemal complex of *Drosophila*. *BMC Evolutionary Biology*, 16, 1–17.
- Holleley, C. E., O'Meally, D., Sarre, S. D., Marshall Graves, J. A., Ezaz, T., Matsubara, K., Azad, B., Zhang, X., & Georges, A. (2015). Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. *Nature*, 523(7558), 79–82.
- Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., Pfeiffer, B. D., George, R. A., Svirskas, R., et al. (2015). The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research*, 25(3), 445–458.
- Iannello, M., Forni, G., Piccinini, G., Xu, R., Martelossi, J., Ghiselli, F., & Milani, L. (2023). Signatures of extreme longevity: A perspective from bivalve molecular evolution. *Genome Biology and Evolution*, 15(11), evad159.
- Inoue, K., Yoshioka, Y., Tanaka, H., Kinjo, A., Sassa, M., Ueda, I., Shinzato, C., Toyoda, A., & Itoh, T. (2021). Genomics and transcriptomics of the green mussel explain the durability of its byssus. *Scientific Reports*, 11(1), 5992.
- Ip, J. C.-H., Xu, T., Sun, J., Li, R., Chen, C., Lan, Y., Han, Z., Zhang, H., Wei, J., Wang, H., et al. (2021). Host–endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Molecular Biology and Evolution*, 38(2), 502–518.
- Jackson, B. C., Carpenter, C., Nebert, D. W., & Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. *Human Genomics*, 4, 1–8.
- Jombart, T., & Dray, S. (2010). Adephylo: Exploratory analyses for the phylogenetic comparative method. *Bioinformatics*, 26(15), 1–21.
- Kakoi, S., Kin, K., Miyazaki, K., & Wada, H. (2008). Early development of the Japanese spiny oyster (*Saccostrea kegaki*): Characterization of some genetic markers. *Zoological Science*, 25(5), 455–464.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589.
- Kenchington, E. L., Hamilton, L., Cogswell, A., & Zouros, E. (2009). Paternal mtDNA and maleness are co-inherited but not causally linked in mytilid mussels. *PLoS One*, 4(9), e6976.

- Kenny, N. J., McCarthy, S. A., Dudchenko, O., James, K., Betteridge, E., Corton, C., Dolucan, J., Mead, D., Oliver, K., Omer, A. D., et al. (2020). The gene-rich genome of the scallop *Pecten maximus*. *GigaScience*, 9(5),giaa037.
- Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., Comeault, A. A., Peede, D., D'Agostino, E. R., Pelaez, J., et al. (2021). Highly contiguous assemblies of 101 drosophilid genomes. *eLife*, 10, e66405.
- Klein, S. L., Neilson, K. M., Orban, J., Yaklichkin, S., Hoffbauer, J., Mood, K., Daar, I. O., & Moody, S. A. (2013). Conserved structural domains in FoxD4L1, a neural forkhead box transcription factor, are required to repress or activate target genes. *PLoS One*, 8(4), e61845.
- Knudsen, B., Kohn, A. B., Nahir, B., McFadden, C. S., & Moroz, L. L. (2006). Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: Conservation of the gene order in Euthyneura. *Molecular Phylogenetics and Evolution*, 38(2), 459–469.
- Kopp, A. (2012). Dmrt genes in the development and evolution of sexual dimorphism. *Trends in Genetics*, 28(4), 175–184.
- Korpelainen, H. (1990). Sex ratios and conditions required for environmental sex determination in animals. *Biological Reviews*, 65(2), 147–184.
- Kranz, A. M., Tollenaere, A., Norris, B. J., Degnan, B. M., & Degnan, S. M. (2010). Identifying the germline in an equally cleaving mollusc: *Vasa* and *Nanos* expression during embryonic and larval development of the vetigastropod *Haliotis asinina*. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 314(4), 267–279.
- Kuehn, E., Clausen, D. S., Null, R. W., Metzger, B. M., Willis, A. D., & Özpolat, B. D. (2022). Segment number threshold determines juvenile onset of germline cluster expansion in *Platynereis dumerilii*. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 338(4), 225–240.
- Kumano, G. (2015). Evolution of germline segregation processes in animal development. *Development, Growth & Differentiation*, 57(4), 324–332.
- Lan, Y., Sun, J., Chen, C., Sun, Y., Zhou, Y., Yang, Y., Zhang, W., Li, R., Zhou, K., Wong, W. C., et al. (2021). Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nature Communications*, 12(1), 1165.

- Lan, Y., Sun, J., Zhang, W., Xu, T., Zhang, Y., Chen, C., Feng, D., Wang, H., Tao, J., Qiu, J.-W., et al. (2019). Host–symbiont interactions in deep-sea chemosymbiotic vesicomyid clams: Insights from transcriptome sequencing. *Frontiers in Marine Science*, 6, 680.
- Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D. S., Shimeld, S. M., & Degnan, B. M. (2008). Genesis and expansion of metazoan transcription factor gene classes. *Molecular Biology and Evolution*, 25(5), 980–996.
- Lemoine, F., & Gascuel, O. (2021). Gotree/Goalign: Toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genomics and Bioinformatics*, 3(3), lqab075.
- Li, A., Dai, H., Guo, X., Zhang, Z., Zhang, K., Wang, C., Wang, X., Wang, W., Chen, H., Li, X., et al. (2021). Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity. *Communications Biology*, 4(1), 1287.
- Li, C., Liu, X., Liu, B., Ma, B., Liu, F., Liu, G., Shi, Q., & Wang, C. (2018). Draft genome of the Peruvian scallop *Argopecten purpuratus*. *GigaScience*, 7(4), giy031.
- Li, F., Bian, L., Ge, J., Han, F., Liu, Z., Li, X., Liu, Y., Lin, Z., Shi, H., Liu, C., et al. (2020). Chromosome-level genome assembly of the East Asian common octopus (*Octopus sinensis*) using PacBio sequencing and Hi-C technology. *Molecular Ecology Resources*, 20(6), 1572–1582.
- Li, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics*, 39(1), btad014.
- Li, J., Zhou, Y., Zhou, Z., Lin, C., Wei, J., Qin, Y., Xiang, Z., Ma, H., Zhang, Y., Zhang, Y., et al. (2020). Comparative transcriptome analysis of three gonadal development stages reveals potential genes involved in gametogenesis of the fluted giant clam (*Tridacna squamosa*). *BMC Genomics*, 21, 1–16.
- Li, R., Zhang, L., Li, W., Zhang, Y., Li, Y., Zhang, M., Zhao, L., Hu, X., Wang, S., & Bao, Z. (2018). *FOXL2* and *DMRT1L* are yin and yang genes for determining timing of sex differentiation in the bivalve mollusk *Patinopecten yessoensis*. *Frontiers in Physiology*, 9, 1166.
- Li, X.-Y., Mei, J., Ge, C.-T., Liu, X.-L., & Gui, J.-F. (2022). Sex determination mechanisms and sex control approaches in aquaculture animals. *Science China Life Sciences*, 65(6), 1091–1122.
- Liang, S., Liu, D., Li, X., Wei, M., Yu, X., Li, Q., Ma, H., Zhang, Z., & Qin, Z. (2019). *SOX2* participates in spermatogenesis of Zhikong scallop *Chlamys farreri*. *Scientific Reports*, 9(1), 76.

- Liao, Y., Zhang, X., Chakraborty, M., & Emerson, J. (2021). Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Research*, 31(3), 397–410.
- Liu, C., Zhang, Y., Ren, Y., Wang, H., Li, S., Jiang, F., Yin, L., Qiao, X., Zhang, G., Qian, W., et al. (2018). The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *GigaScience*, 7(9), giy101.
- Liu, T., Zhang, Y., Nie, H., Sun, J., & Yan, X. (2024). Characterization and expression patterns of the Fox gene family under heat and cold stress in Manila clam *Ruditapes philippinarum* based on genome-wide identification. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 52, 101313.
- Liu, X., Li, C., Chen, M., Liu, B., Yan, X., Ning, J., Ma, B., Liu, G., Zhong, Z., Jia, Y., et al. (2020). Draft genomes of two Atlantic bay scallop subspecies *Argopecten irradians irradians* and *A. i. concentricus*. *Scientific Data*, 7(1), 99.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21.
- Lyons, D. C., Perry, K. J., Lesoway, M. P., & Henry, J. Q. (2012). Cleavage pattern and fate map of the mesentoblast, 4d, in the gastropod *Crepidula*: A hallmark of spiralian development. *EvoDevo*, 3, 1–26.
- Mahajan, S., Wei, K. H.-C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young *Drosophila* Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biology*, 16(7), e2006348.
- Mai, D., Nalley, M. J., & Bachtrog, D. (2020). Patterns of genomic differentiation in the *Drosophila nasuta* species complex. *Molecular Biology and Evolution*, 37(1), 208–220.
- Mank, J. E., Axelsson, E., & Ellegren, H. (2007). Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Research*, 17(5), 618–624.
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654.
- Marshall Graves, J. A., & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology*, 11, 1–12.

- Matson, C. K., & Zarkower, D. (2012). Sex and the singular DM domain: Insights into sexual regulation, evolution and plasticity. *Nature Reviews Genetics*, 13(3), 163–174.
- Mawaribuchi, S., Ito, Y., & Ito, M. (2019). Independent evolution for sex determination and differentiation in the *DMRT* family in animals. *Biology Open*, 8(8), bio041962.
- Mawaribuchi, S., Yoshimoto, S., Ohashi, S., Takamatsu, N., & Ito, M. (2012). Molecular evolution of vertebrate sex-determining genes. *Chromosome Research*, 20, 139–151.
- Mazet, F., Yu, J.-K., Liberles, D. A., Holland, L. Z., & Shimeld, S. M. (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene*, 316, 79–89.
- McCartney, M. A., Auch, B., Kono, T., Mallez, S., Zhang, Y., Obille, A., Becker, A., Abrahante, J. E., Garbe, J., Badalamenti, J. P., et al. (2022). The genome of the zebra mussel, *Dreissena polymorpha*: A resource for comparative genomics, invasion genetics, and biocontrol. *G3: Genes, Genomes, Genetics*, 12(2), jkab423.
- Mead, D., Fingland, K., Cripps, R., Miguez, R. P., Smith, M., Corton, C., Oliver, K., Skelton, J., Betteridge, E., Doulcan, J., et al. (2020). The genome sequence of the eastern grey squirrel, *Sciurus carolinensis* Gmelin, 1788. *Wellcome Open Research*, 5.
- Meisel, R. P. (2011). Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Molecular Biology and Evolution*, 28(6), 1893–1900.
- Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, 29(9), 537–544.
- Meng, J., Yang, M., Xu, F., Li, X., & Li, L. (2018). Transcriptome assembly of Modiolus modiolus and comparative analysis with *Bathymodiolus platifrons*. *Acta Oceanologica Sinica*, 37, 38–45.
- Miglioli, A., Tredez, M., Boosten, M., Sant, C., Carvalho, J. E., Dru, P., Canesi, L., Schubert, M., & Dumollard, R. (2024). The Mediterranean mussel *Mytilus galloprovincialis*: A novel model for developmental studies in mollusks. *Development*, 151(4), dev202256.
- Milani, L., & Ghiselli, F. (2020). Faraway, so close. The comparative method and the potential of non-model animals in mitochondrial research. *Philosophical Transactions of the Royal Society B*, 375(1790), 20190186.
- Milani, L., Ghiselli, F., Maurizii, M. G., Nuzhdin, S. V., & Passamonti, M. (2014). Paternally transmitted mitochondria express a new gene of potential viral origin. *Genome Biology and Evolution*, 6(2), 391–405.

- Milani, L., Ghiselli, F., Maurizii, M. G., & Passamonti, M. (2011). Doubly uniparental inheritance of mitochondria as a model system for studying germ line formation. *PLoS One*, 6(11), e28194.
- Milani, L., Ghiselli, F., Nuzhdin, S. V., & Passamonti, M. (2013). Nuclear genes with sex bias in *Ruditapes philippinarum* (Bivalvia, Veneridae): Mitochondrial inheritance and sex determination in DUI species. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 320(7), 442–454.
- Milani, L., Pecci, A., Ghiselli, F., Passamonti, M., Bettini, S., Franceschini, V., & Maurizii, M. G. (2017). VASA expression suggests shared germ line dynamics in bivalve molluscs. *Histochemistry and Cell Biology*, 148, 157–171.
- Minamikawa, T., Sriratana, A., Williams, D. A., Bowser, D. N., Hill, J. S., & Nagley, P. (1999). Chloromethyl-X-rosamine (MitoTracker Red) photosensitises mitochondria and induces apoptosis in intact human cells. *Journal of Cell Science*, 112(14), 2419–2430.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534.
- Moreno, J. A., Dudchenko, O., Feigin, C. Y., Mereby, S. A., Chen, Z., Ramos, R., Almet, A. A., Sen, H., Brack, B. J., Johnson, M. R., et al. (2024). *Emx2* underlies the development and evolution of marsupial gliding membranes. *Nature*, 1–9.
- Mork, L., Czerwinski, M., & Capel, B. (2014). Predetermination of sexual fate in a turtle with temperature-dependent sex determination. *Developmental Biology*, 386(1), 264–271.
- Mullon, C., Pomiankowski, A., & Reuter, M. (2012). Molecular evolution of *Drosophila Sex-lethal* and related sex determining genes. *BMC Evolutionary Biology*, 12, 1–11.
- Murat, F., Mbengue, N., Winge, S. B., Trefzer, T., Leushkin, E., Sepp, M., Cardoso-Moreira, M., Schmidt, J., Schneider, C., Mößinger, K., et al. (2023). The molecular evolution of spermatogenesis across mammals. *Nature*, 613(7943), 308–316.
- Murata, Y., Nikaido, M., Sasaki, T., Cao, Y., Fukumoto, Y., Hasegawa, M., & Okada, N. (2003). Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 28(2), 253–260.
- Naimi, A., Martinez, A.-S., Specq, M.-L., Diss, B., Mathieu, M., & Sourdaine, P. (2009). Molecular cloning and gene expression of *Cg-Foxl2* during the development and the adult game-

- togenetic cycle in the oyster *Crassostrea gigas*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 154(1), 134–142.
- Nakagawa, S., Gisselbrecht, S. S., Rogers, J. M., Hartl, D. L., & Bulyk, M. L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proceedings of the National Academy of Sciences*, 110(30), 12349–12354.
- Nicolini, F., Ghiselli, F., Luchetti, A., & Milani, L. (2023). Bivalves as emerging model systems to study the mechanisms and evolution of sex determination: A genomic point of view. *Genome Biology and Evolution*, 15(10), evad181.
- Nicolini, F., Martelossi, J., Forni, G., Savojardo, C., Mantovani, B., & Luchetti, A. (2023). Comparative genomics of Hox and ParaHox genes among major lineages of Branchiopoda with emphasis on tadpole shrimps. *Frontiers in Ecology and Evolution*, 11, 1046960.
- Nie, H., Dong, S., Li, D., Zheng, M., Jiang, L., Li, X., & Yan, X. (2018). RNA-Seq analysis of differentially expressed genes in the grand jackknife clam *Solen grandis* under aerial exposure. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 28, 54–62.
- Obata, M., & Komaru, A. (2012). The mechanisms of primordial germ cell determination during embryogenesis in molluscan species. *Invertebrate Survival Journal*, 9(2), 223–229.
- Obata, M., & Komaru, A. (2005). Specific location of sperm mitochondria in mussel *Mytilus galloprovincialis* zygotes stained by MitoTracker. *Development, Growth & Differentiation*, 47(4), 255–263.
- Obata, M., Sano, N., Kimata, S., Nagasawa, K., Yoshizaki, G., & Komaru, A. (2010). The proliferation and migration of immature germ cells in the mussel, *Mytilus galloprovincialis*: Observation of the expression pattern in the *M. galloprovincialis* vasa-like gene (*Myvlg*) by in situ hybridization. *Development Genes and Evolution*, 220, 139–149.
- of Sex Consortium, T. (2014). Tree of sex: A database of sexual systems. *Scientific Data*, 1.
- O'Neil, M. T., & Belote, J. M. (1992). Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics*, 131(1), 113–128.
- Orkin, J. D., Montague, M. J., Tejada-Martinez, D., De Manuel, M., Del Campo, J., Cheves Hernandez, S., Di Fiore, A., Fontserè, C., Hodgson, J. A., Janiak, M. C., et al. (2021). The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys

- revealed with fecalFACS. *Proceedings of the National Academy of Sciences*, 118(7), e2010632118.
- Paggeot, L. X., DeBiase, M. B., Escalona, M., Fairbairn, C., Marimuthu, M. P., Nguyen, O., Sahasrabudhe, R., & Dawson, M. N. (2022). Reference genome for the California ribbed mussel, *Mytilus californianus*, an ecosystem engineer. *Journal of Heredity*, 113(6), 681–688.
- Pamilo, P., & O'Neill, R. (1997). Evolution of the *Sry* genes. *Molecular Biology and Evolution*, 14(1), 49–55.
- Panara, V., Budd, G. E., & Janssen, R. (2019). Phylogenetic analysis and embryonic expression of panarthropod Dmrt genes. *Frontiers in Zoology*, 16, 1–18.
- Papa, F., Windbichler, N., Waterhouse, R. M., Cagnetti, A., D'Amato, R., Persampieri, T., Lawniczak, M. K., Nolan, T., & Papathanos, P. A. (2017). Rapid evolution of female-biased genes among four species of *Anopheles* malaria mosquitoes. *Genome Research*, 27(9), 1536–1548.
- Paris, M., Boyer, R., Jaenichen, R., Wolf, J., Karageorgi, M., Green, J., Cagnon, M., Parinello, H., Estoup, A., Gautier, M., et al. (2020). Near-chromosome level genome assembly of the fruit pest *Drosophila suzukii* using long-read sequencing. *Scientific Reports*, 10(1), 11227.
- Parsch, J., & Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. *Nature Reviews Genetics*, 14(2), 83–87.
- Patnaik, B. B., Wang, T. H., Kang, S. W., Hwang, H.-J., Park, S. Y., Park, E. B., Chung, J. M., Song, D. K., Kim, C., Kim, S., et al. (2016). Sequencing, de novo assembly, and annotation of the transcriptome of the endangered freshwater pearl bivalve, *Cristaria plicata*, provides novel insights into functional genes and marker discovery. *PLoS One*, 11(2), e0148622.
- Penaloza, C., Gutierrez, A. P., Eöry, L., Wang, S., Guo, X., Archibald, A. L., Bean, T. P., & Houston, R. D. (2021). A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *GigaScience*, 10(3), giab020.
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9), 1650–1667.

- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295.
- Phochanukul, N., & Russell, S. (2010). No backbone but lots of Sox: Invertebrate Sox genes. *The International Journal of Biochemistry & Cell Biology*, 42(3), 453–464.
- Piccinini, G., Iannello, M., Puccio, G., Plazzi, F., Havird, J. C., & Ghiselli, F. (2021). Mitonuclear coevolution, but not nuclear compensation, drives evolution of OXPHOS complexes in bivalves. *Molecular Biology and Evolution*, 38(6), 2597–2614.
- Piccinini, G., & Milani, L. (2023). Germline-related molecular phenotype in Metazoa: Conservation and innovation highlighted by comparative transcriptomics. *EvoDevo*, 14(1), 2.
- Powell, D., Subramanian, S., Suwansa-Ard, S., Zhao, M., O'Connor, W., Raftos, D., & Elizur, A. (2018). The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. *DNA Research*, 25(6), 655–665.
- Purandare, S. R., Bickel, R. D., Jaquière, J., Rispe, C., & Brisson, J. A. (2014). Accelerated evolution of morph-biased genes in pea aphids. *Molecular Biology and Evolution*, 31(8), 2073–2083.
- Putnam, H. M., Trigg, S. A., White, S. J., Spencer, L. H., Vadopalas, B., Natarajan, A., Hetzel, J., Jaeger, E., Soohoo, J., Gallardo-Escárate, C., et al. (2022). Dynamic DNA methylation contributes to carryover effects and beneficial acclimatization in geoduck clams. *bioRxiv*, 2022–06.
- Qi, H., Cong, R., Wang, Y., Li, L., & Zhang, G. (2023). Construction and analysis of the chromosome-level haplotype-resolved genomes of two *Crassostrea* oyster congeners: *Crassostrea angulata* and *Crassostrea gigas*. *GigaScience*, 12, giad077.
- Ran, Z., Li, Z., Yan, X., Liao, K., Kong, F., Zhang, L., Cao, J., Zhou, C., Zhu, P., He, S., et al. (2019). Chromosome-level genome assembly of the razor clam *Sinonovacula constricta* (Lamarck, 1818). *Molecular Ecology Resources*, 19(6), 1647–1658.
- Ranz, J. M., Go, A. C., González, P. M., Clifton, B. D., Gomes, S., Jaberyzadeh, A., Woodbury, A., Chan, C., Gandasetiawan, K. A., Jayasekera, S., et al. (2023). Gene expression differentiation in the reproductive tissues of *Drosophila willistoni* subspecies and their hybrids. *Molecular Ecology*, 32(13), 3605–3623.

- Rebscher, N. (2014). Establishing the germline in spiralian embryos. *International Journal of Developmental Biology*, 58.
- Rebscher, N., Zelada-González, F., Banisch, T. U., Raible, F., & Arendt, D. (2007). Vasa unveils a common origin of germ cells and of somatic stem cells from the posterior growth zone in the polychaete *Platynereis dumerilii*. *Developmental Biology*, 306(2), 599–611.
- Renfree, M. B., O, W.-S., Short, R., & Shaw, G. (1996). Sexual differentiation of the urogenital system of the fetal and neonatal tammar wallaby, *Macropus eugenii*. *Anatomy and Embryology*, 194, 111–134.
- Renschler, G., Richard, G., Valsecchi, C. I. K., Toscano, S., Arrigoni, L., Ramírez, F., & Akhtar, A. (2019). Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes & Development*, 33(21-22), 1591–1612.
- Richardson, V., Engel, N., & Kulathinal, R. J. (2023). Comparative developmental genomics of sex-biased gene expression in early embryogenesis across mammals. *Biology of Sex Differences*, 14(1), 30.
- Rogers, R. L., Grizzard, S. L., Titus-McQuillan, J. E., Bockrath, K., Patel, S., Wares, J. P., Garner, J. T., & Moore, C. C. (2021). Gene family amplification facilitates adaptation in freshwater unionid bivalve *Megalonaia nervosa*. *Molecular Ecology*, 30(5), 1155–1173.
- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, 47(W1), W5–W10.
- Russo, C. A., Mello, B., Frazão, A., & Voloch, C. M. (2013). Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). *Zoological Journal of the Linnean Society*, 169(4), 765–775.
- Saavedra, C., Reyero, M.-I., & Zouros, E. (1997). Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus galloprovincialis*. *Genetics*, 145(4), 1073–1082.
- Salz, H., & Erickson, J. W. (2010). Sex determination in *Drosophila*: The view from the top. *Fly*, 4(1), 60–70.
- Sanchez-Flores, A., Peñaloza, F., Carpinetyro-Ponce, J., Nazario-Yepiz, N., Abreu-Goodger, C., Machado, C. A., & Markow, T. A. (2016). Genome evolution in three species of cactophilic *Drosophila*. *G3: Genes, Genomes, Genetics*, 6(10), 3097–3105.

- Santerre, C., Sourdaine, P., Marc, N., Mingant, C., Robert, R., & Martinez, A.-S. (2013). Oyster sex determination is influenced by temperature—first clues in spat during first gonadic differentiation and gametogenesis. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 165(1), 61–69.
- Sarkar, A., & Hochedlinger, K. (2013). The sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell*, 12(1), 15–30.
- Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593.
- Schomburg, C., Janssen, R., & Prpic, N.-M. (2022). Phylogenetic analysis of forkhead transcription factors in the Panarthropoda. *Development Genes and Evolution*, 232(1), 39–48.
- Seudre, O., Martín-Zamora, F. M., Rapisarda, V., Luqman, I., Carrillo-Baltodano, A. M., & Martín-Durán, J. M. (2022). The *Fox* gene repertoire in the annelid *Owenia fusiformis* reveals multiple expansions of the *foxQ2* class in Spiralia. *Genome Biology and Evolution*, 14(10), evac139.
- Shi, J., Hong, Y., Sheng, J., Peng, K., & Wang, J. (2015). De novo transcriptome sequencing to identify the sex-determination genes in *Hyriopsis schlegelii*. *Bioscience, Biotechnology, and Biochemistry*, 79(8), 1257–1265.
- Shi, Y., Liu, W., & He, M. (2018). Proteome and transcriptome analysis of ovary, intersex gonads, and testis reveals potential key sex reversal/differentiation genes and mechanism in scallop *Chlamys nobilis*. *Marine Biotechnology*, 20, 220–245.
- Shimeld, S. M., Boyle, M. J., Brunet, T., Luke, G. N., & Seaver, E. C. (2010). Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. *Developmental Biology*, 340(2), 234–248.
- Shimizu, K., Negishi, L., Ito, T., Touma, S., Matsumoto, T., Awaji, M., Kurumizaka, H., Yoshitake, K., Kinoshita, S., Asakawa, S., et al. (2022). Evolution of nacre-and prisms-related shell matrix proteins in the pen shell, *Atrina pectinata*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 44, 101025.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539.

- Smith, C. H. (2021). A high-quality reference genome for a parasitic bivalve with doubly uniparental inheritance (Bivalvia: Unionida). *Genome Biology and Evolution*, 13(3), evab029.
- Song, H., Guo, X., Sun, L., Wang, Q., Han, F., Wang, H., Wray, G. A., Davidson, P., Wang, Q., Hu, Z., et al. (2021). The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biology*, 19, 1–20.
- Song, H., Wang, Y., Shao, H., Li, Z., Hu, P., Yap-Chiongco, M. K., Shi, P., Zhang, T., Li, C., Wang, Y., et al. (2023). Scaphopoda is the sister taxon to Bivalvia: Evidence of ancient incomplete lineage sorting. *Proceedings of the National Academy of Sciences*, 120(40), e2302361120.
- Stammnitz, M. R., Gori, K., Kwon, Y. M., Harry, E., Martin, F. J., Billis, K., Cheng, Y., Baez-Ortega, A., Chow, W., Comte, S., et al. (2023). The evolution of two transmissible cancers in Tasmanian devils. *Science*, 380(6642), 283–293.
- Stothard, P., & Pilgrim, D. (2003). Sex-determination gene and pathway evolution in nematodes. *BioEssays*, 25(3), 221–231.
- Sun, D., Yu, H., & Li, Q. (2022). Examination of the roles of *Foxl2* and *Dmrt1* in sex differentiation and gonadal development of oysters by using RNA interference. *Aquaculture*, 548, 737732.
- Sun, J., Zhang, Y., Xu, T., Zhang, Y., Mu, H., Zhang, Y., Lan, Y., Fields, C. J., Hui, J. H. L., Zhang, W., et al. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution*, 1(5), 0121.
- Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, 3(2), 137–144.
- Swartz, S. Z., Chan, X. Y., & Lambert, J. D. (2008). Localization of *Vasa* mRNA during early cleavage of the snail *Ilyanassa*. *Development Genes and Evolution*, 218, 107–113.
- Teaniniuraitemoana, V., Huvet, A., Levy, P., Klopp, C., Lhuillier, E., Gaertner-Mazouni, N., Gueguen, Y., & Le Moullac, G. (2014). Gonad transcriptome analysis of pearl oyster *Pinctada margaritifera*: Identification of potential sex differentiation and sex determining genes. *BMC Genomics*, 15, 1–20.
- Teaniniuraitemoana, V., Leprêtre, M., Levy, P., Vanaa, V., Parrad, S., Gaertner-Mazouni, N., Gueguen, Y., Huvet, A., & Le Moullac, G. (2016). Effect of temperature, food availability, and estradiol injection on gametogenesis and gender in the pearl oyster *Pinctada*

- margaritifera*. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 325(1), 13–24.
- Teng, W., Fu, H., Li, Z., Zhang, Q., Xu, C., Yu, H., Kong, L., Liu, S., & Li, Q. (2023). Parallel evolution in *Crassostrea* oysters along the latitudinal gradient is associated with variation in multiple genes involved in adipogenesis. *Molecular Ecology*, 32(19), 5276–5287.
- Todd, E. V., Liu, H., Muncaster, S., & Gemmell, N. J. (2016). Bending genders: The biology of natural sex change in fish. *Sexual Development*, 10(5-6), 223–241.
- Tu, Q., Brown, C. T., Davidson, E. H., & Oliveri, P. (2006). Sea urchin Forkhead gene family: Phylogeny and embryonic expression. *Developmental Biology*, 300(1), 49–62.
- Tvedte, E. S., Gasser, M., Sparklin, B. C., Michalski, J., Hjelmen, C. E., Johnston, J. S., Zhao, X., Bromley, R., Tallon, L. J., Sadzewicz, L., et al. (2021). Comparison of long-read sequencing technologies in interrogating bacteria and fly genomes. *G3: Genes, Genomes, Genetics*, 11(6), jkab083.
- Uller, T., & Helanterä, H. (2011). From the origin of sex-determining factors to the evolution of sex-determining systems. *The Quarterly Review of Biology*, 86(3), 163–180.
- Verhulst, E. C., van de Zande, L., & Beukeboom, L. W. (2010). Insect sex determination: It all evolves around *transformer*. *Current Opinion in Genetics & Development*, 20(4), 376–383.
- Vicoso, B., & Charlesworth, B. (2006). Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews Genetics*, 7(8), 645–653.
- Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C., Weinstock, J., Froese, D., Vasiliev, S. K., Ovodov, N. D., et al. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PLoS One*, 8(2), e55950.
- Vinkler, M., Fiddaman, S. R., Těšický, M., O'Connor, E. A., Savage, A. E., Lenz, T. L., Smith, A. L., Kaufman, J., Bolnick, D. I., Davies, C. S., et al. (2023). Understanding the evolution of immune genes in jawed vertebrates. *Journal of Evolutionary Biology*, 36(6), 847–873.
- Vizueta Moraga, J., Sánchez-Gracia, A., & Rozas Liras, J. A. (2020). Bitacora: A comprehensive tool for the identification and annotation of gene families in genome assemblies. *Molecular Ecology Resources*, 20(5), 1445–1452.
- Voronina, E., Lopez, M., Juliano, C. E., Gustafson, E., Song, J. L., Extavour, C., George, S., Oliveri, P., McClay, D., & Wessel, G. (2008). Vasa protein expression is restricted

- to the small micromeres of the sea urchin, but is inducible in other lineages early in development. *Developmental Biology*, 314(2), 276–286.
- Wang, C., Wallerman, O., Arendt, M.-L., Sundström, E., Karlsson, Å., Nordin, J., Mäkeläinen, S., Pielberg, G. R., Hanson, J., Ohlsson, Å., et al. (2021). A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Communications Biology*, 4(1), 185.
- Wang, G., Dong, S., Guo, P., Cui, X., Duan, S., & Li, J. (2020). Identification of *Foxl2* in freshwater mussel *Hyriopsis cumingii* and its involvement in sex differentiation. *Gene*, 754, 144853.
- Wang, J., & Nie, H. (2024). Genome-wide identification and expression analysis of Sox gene family in the Manila clam (*Ruditapes philippinarum*). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 50, 101244.
- Wang, M., Xia, J., Jawad, M., Wei, W., Gui, L., Liang, X., Yang, J.-L., & Li, M. (2022). Transcriptome sequencing analysis of sex-related genes and miRNAs in the gonads of *Mytilus coruscus*. *Frontiers in Marine Science*, 9, 1013857.
- Wang, Q., Cao, T., & Wang, C. (2023). Genome-wide identification and expression analysis of Dmrt genes in bivalves. *BMC Genomics*, 24(1), 457.
- Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., et al. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution*, 1(5), 0120.
- Wang, X., Werren, J. H., & Clark, A. G. (2015). Genetic and epigenetic architecture of sex-biased expression in the jewel wasps *Nasonia vitripennis* and *giraulti*. *Proceedings of the National Academy of Sciences*, 112(27), E3545–E3554.
- Wang, X., Liu, Z., & Wu, W. (2017). Transcriptome analysis of the freshwater pearl mussel (*Cristaria plicata*) mantle unravels genes involved in the formation of shell and pearl. *Molecular Genetics and Genomics*, 292, 343–352.
- Wei, M., Ge, H., Shao, C., Yan, X., Nie, H., Duan, H., Liao, X., Zhang, M., Chen, Y., Zhang, D., et al. (2020). Chromosome-level clam genome helps elucidate the molecular basis of adaptation to a buried lifestyle. *iScience*, 23(6).
- Wexler, J. R., Plachetzki, D. C., & Kopp, A. (2014). Pan-metazoan phylogeny of the DMRT gene family: A framework for functional studies. *Development Genes and Evolution*, 224, 175–181.

- Whitfield, L. S., Lovell-Badge, R., & Goodfellow, P. N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. *Nature*, 364(6439), 713–715.
- Wilkins, A. S. (1995). Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *BioEssays*, 17(1), 71–77.
- Willson, J., Roddur, M. S., Liu, B., Zaharias, P., & Warnow, T. (2022). DISCO: Species tree inference using multicopy gene family tree decomposition. *Systematic Biology*, 71(3), 610–629.
- Woods, F. H. (1932). Keimbahn determinants and continuity of the germ cells in *Sphaerium striatinum* (Lam.) *Journal of Morphology*, 53(2), 345–365.
- Wu, S., Zhang, Y., Li, Y., Wei, H., Guo, Z., Wang, S., Zhang, L., & Bao, Z. (2020). Identification and expression profiles of Fox transcription factors in the Yesso scallop (*Patinopecten yessoensis*). *Gene*, 733, 144387.
- Xiong, Y., Brandley, M. C., Xu, S., Zhou, K., & Yang, G. (2009). Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. *BMC Evolutionary Biology*, 9, 1–13.
- Xu, R., Iannello, M., Havird, J. C., Milani, L., & Ghiselli, F. (2022). Lack of transcriptional coordination between mitochondrial and nuclear oxidative phosphorylation genes in the presence of two divergent mitochondrial genomes. *Zoological Research*, 43(1), 111.
- Xu, R., Martelossi, J., Smits, M., Iannello, M., Peruzza, L., Babbucci, M., Milan, M., Dunham, J. P., Breton, S., Milani, L., et al. (2022). Multi-tissue RNA-Seq analysis and long-read-based genome assembly reveal complex sex-specific gene regulation and molecular evolution in the Manila clam. *Genome Biology and Evolution*, 14(12), evac171.
- Xu, R., Li, Q., Yu, H., & Kong, L. (2018). Oocyte maturation and origin of the germline as revealed by the expression of *Nanos*-like in the Pacific oyster *Crassostrea gigas*. *Gene*, 663, 41–50.
- Yang, J.-L., Feng, D.-D., Liu, J., Xu, J.-K., Chen, K., Li, Y.-F., Zhu, Y.-T., Liang, X., & Lu, Y. (2021). Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *GigaScience*, 10(4), giab024.
- Yang, M., Xu, F., Liu, J., Que, H., Li, L., & Zhang, G. (2014). Phylogeny of forkhead genes in three spiralians and their expression in Pacific oyster *Crassostrea gigas*. *Chinese Journal of Oceanology and Limnology*, 32(6), 1207–1223.

- Yu, J., Zhang, L., Li, Y., Li, R., Zhang, M., Li, W., Xie, X., Wang, S., Hu, X., & Bao, Z. (2017). Genome-wide identification and expression profiling of the SOX gene family in a bivalve mollusc *Patinopecten yessoensis*. *Gene*, *627*, 530–537.
- Yu, J.-K., Mazet, F., Chen, Y.-T., Huang, S.-W., Jung, K.-C., & Shimeld, S. M. (2008). The Fox genes of *Branchiostoma floridae*. *Development Genes and Evolution*, *218*, 629–638.
- Yue, C., Li, Q., & Yu, H. (2021). Variance in expression and localization of sex-related genes *CgDsx*, *CgBHMG1* and *CgFoxl2* during diploid and triploid Pacific oyster *Crassostrea gigas* gonad differentiation. *Gene*, *790*, 145692.
- Yuen, B., Polzin, J., & Petersen, J. M. (2019). Organ transcriptomes of the lucinid clam *Loripes orbiculatus* (Poli, 1791) provide insights into their specialised roles in the biology of a chemosymbiotic bivalve. *BMC Genomics*, *20*, 1–14.
- Zeng, Y., Zheng, H., He, C., Zhang, C., Zhang, H., & Zheng, H. (2024). Genome-wide identification and expression analysis of Dmrt gene family and their role in gonad development of Pacific oyster (*Crassostrea gigas*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *269*, 110904.
- Zhang, N., Xu, F., & Guo, X. (2014). Genomic analysis of the Pacific oyster (*Crassostrea gigas*) reveals possible conservation of vertebrate sex determination in a mollusc. *G3: Genes, Genomes, Genetics*, *4*(11), 2207–2217.
- Zhao, D., McBride, D., Nandi, S., McQueen, H., McGrew, M., Hocking, P., Lewis, P., Sang, H., & Clinton, M. (2010). Somatic sex identity is cell autonomous in the chicken. *Nature*, *464*(7286), 237–242.
- Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L., et al. (2022). Ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Briefings in Bioinformatics*, *23*(4), bbac222.
- Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J. C., Peel, E., Cheng, Y., Pajpach, F., et al. (2021). Platypus and echidna genomes reveal mammalian biology and evolution. *Nature*, *592*(7856), 756–762.
- Zhu, C., Zhang, L., Ding, H., & Pan, Z. (2019). Transcriptome-wide identification and characterization of the *Sox* gene family and microsatellites for *Corbicula fluminea*. *PeerJ*, *7*, e7770.
- Zouros, E. (2013). Biparental inheritance through uniparental transmission: The doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evolutionary Biology*, *40*, 1–31.

Appendix

The appendix includes the titles and abstracts of the papers published during my PhD that are not part of this thesis.

Taxonomic revision of the Australian stick insect genus *Candovia* (Phasmida: Necrosciinae): insight from molecular systematics and species-delimitation approaches.

Giobbe Forni^{1,2}, Alex Cussigh^{1,2}, Paul D. Brock³, Braxton R. Jones⁴, Filippo Nicolini¹, Jacopo Martelossi¹, Andrea Luchetti¹, Barbara Mantovani¹

¹Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.

²Department of Agricultural and Environmental Sciences, University of Milan, Milano, Italy.

³The Natural History Museum, Cromwell Road, London, UK.

⁴School of Life and Environmental Sciences, The University of Sydney, Sydney NSW 2006, Australia.

Published in: 2023, *Zoological Journal of the Linnean Society*, 197:189–210. doi: [10.1093/zoolinnean/zlac074](https://doi.org/10.1093/zoolinnean/zlac074)

Abstract. The Phasmida genus *Candovia* comprises nine traditionally recognized species, all endemic to Australia. In this study, *Candovia* diversity is explored through molecular species-delimitation analyses using the *COI_{Fol}* gene fragment and phylogenetic inferences leveraging seven additional mitochondrial and nuclear loci. Molecular results were integrated with morphological observations, leading us to confirm the already described species and to the delineation of several new taxa and of the new genus *Paracandovia*. New *Candovia* species from various parts of Queensland and New South Wales are described and illustrated (*C. alata* sp. nov., *C. byfieldensis* sp. nov., *C. dalgleishae* sp. nov., *C. eungellensis* sp. nov., *C. karasi* sp. nov., *C. koensi* sp. nov. and *C. wollumbinensis* sp. nov.). New combinations are proposed and species removed from synonymy with the erection of the new genus *Paracandovia* (*P. cercata* stat. rev., comb. nov., *P. longipes* stat. rev., comb. nov., *P. pallida* comb. nov., *P. peridromes* comb. nov., *P. tenera* stat. rev., comb. nov.). Phylogenetic analyses suggest that the egg capitulum may have independently evolved multiple times throughout the evolutionary history of these insects. Furthermore, two newly described species represent the first taxa with fully developed wings in this previously considered apterous clade.

Comparative genomics of *Hox* and *ParaHox* genes among major lineages of Branchiopoda with emphasis on tadpole shrimps.

Filippo Nicolini^{1,2}, Jacopo Martelossi¹, Giobbe Forni³, Castrense Savojardo⁴, Barbara Mantovani¹, Andrea Luchetti¹

¹Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.

²Fano Marine Center, Fano (PU), Italy.

³Department of Agricultural and Environmental Sciences, University of Milan, Milan, Italy.

⁴Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy.

Published in: 2023, *Frontiers in Ecology and Evolution*, 11:1046960.

doi: [10.3389/fevo.2023.1046960](https://doi.org/10.3389/fevo.2023.1046960)

Abstract. *Hox* and *ParaHox* genes (HPHGs) are key developmental genes that pattern regional identity along the anterior-posterior body axis of most animals. Here, we identified HPHGs in tadpole shrimps (Pancrustacea, Branchiopoda, Notostraca), an iconic example of the so-called “living fossils” and performed a comparative genomics analysis of HPHGs and the *Hox* cluster among major branchiopod lineages. Notostraca possess the entire *Hox* complement, and the *Hox* cluster seems to be split into two different subclusters, although we were not able to support this finding with chromosome-level assemblies. However, the genomic structure of *Hox* genes in Notostraca appears more derived than that of *Daphnia* spp., which instead retains the plesiomorphic condition of a single compact cluster. Spinicaudata and *Artemia franciscana* show instead a *Hox* cluster subdivided across two or more genomic scaffolds with some orthologs either duplicated or missing. Yet, branchiopod HPHGs are similar among the various clades in terms of both intron length and number, as well as in their pattern of molecular evolution. Sequence substitution rates are in fact generally similar for most of the branchiopod *Hox* genes and the few differences we found cannot be traced back to natural selection, as they are not associated with any signals of diversifying selection or substantial switches in selective modes. Altogether, these findings do not support a significant stasis in the Notostraca *Hox* cluster and further confirm how morphological evolution is not tightly associated with genome dynamics.

Multiple and diversified transposon lineages contribute to early and recent bivalve genome evolution.

Jacopo Martelossi¹, Filippo Nicolini^{1,2}, Simone Subacchi¹, Daniela Pasquale¹, Fabrizio Ghiselli¹, Andrea Luchetti¹

¹*Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.*

²*Fano Marine Center, Fano (PU), Italy.*

Published in: 2023, *BMC Biology*, 21:145. doi: [10.1186/s12915-023-01632-z](https://doi.org/10.1186/s12915-023-01632-z)

Abstract. Background. Transposable elements (TEs) can represent one of the major sources of genomic variation across eukaryotes, providing novel raw materials for species diversification and innovation. While considerable effort has been made to study their evolutionary dynamics across multiple animal clades, molluscs represent a substantially understudied phylum. Here, we take advantage of the recent increase in mollusc genomic resources and adopt an automated TE annotation pipeline combined with a phylogenetic tree-based classification, as well as extensive manual curation efforts, to characterize TE repertoires across 27 bivalve genomes with a particular emphasis on DDE/D class II elements, long interspersed nuclear elements (LINEs), and their evolutionary dynamics. **Results.** We found class I elements as highly dominant in bivalve genomes, with LINE elements, despite less represented in terms of copy number per genome, being the most common retroposon group covering up to 10 % of their genome. We mined 86,488 reverse transcriptases (RVT) containing LINE coming from 12 clades distributed across all known superfamilies and 14,275 class II DDE/D-containing transposons coming from 16 distinct superfamilies. We uncovered a previously underestimated rich and diverse bivalve ancestral transposon complement that could be traced back to their most recent common ancestor that lived about 500 million years ago [Mya]. Moreover, we identified multiple instances of lineage-specific emergence and loss of different LINEs and DDE/D lineages with the interesting cases of CR1-Zenon, Proto2, RTE-X, and Academ elements that underwent a bivalve-specific amplification likely associated with their diversification. Finally, we found that this LINE diversity is maintained in extant species by an equally diverse set of long-living and potentially active elements, as suggested by their evolutionary history and transcription pro-

files in both male and female gonads. **Conclusions.** We found that bivalves host an exceptional diversity of transposons compared to other molluscs. Their LINE complement could mainly follow a “stealth drivers” model of evolution where multiple and diversified families are able to survive and co-exist for a long period of time in the host genome, potentially shaping both recent and early phases of bivalve genome evolution and diversification. Overall, we provide not only the first comparative study of TE evolutionary dynamics in a large but understudied phylum such as Mollusca, but also a reference library for ORF-containing class II DDE/D and LINE elements, which represents an important genomic resource for their identification and characterization in novel genomes.

Towards a time-tree solution for Branchiopoda diversification: a jackknife assessment of fossil age priors.

Niccolò Righetti^{1*}, Filippo Nicolini^{2*}, Giobbe Forni², Andrea Luchetti²

¹*Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Sorbonne Université, CNRS, IBPS, UMR7238, Paris, France.*

²*Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.*

* the authors equally contributed to this work.

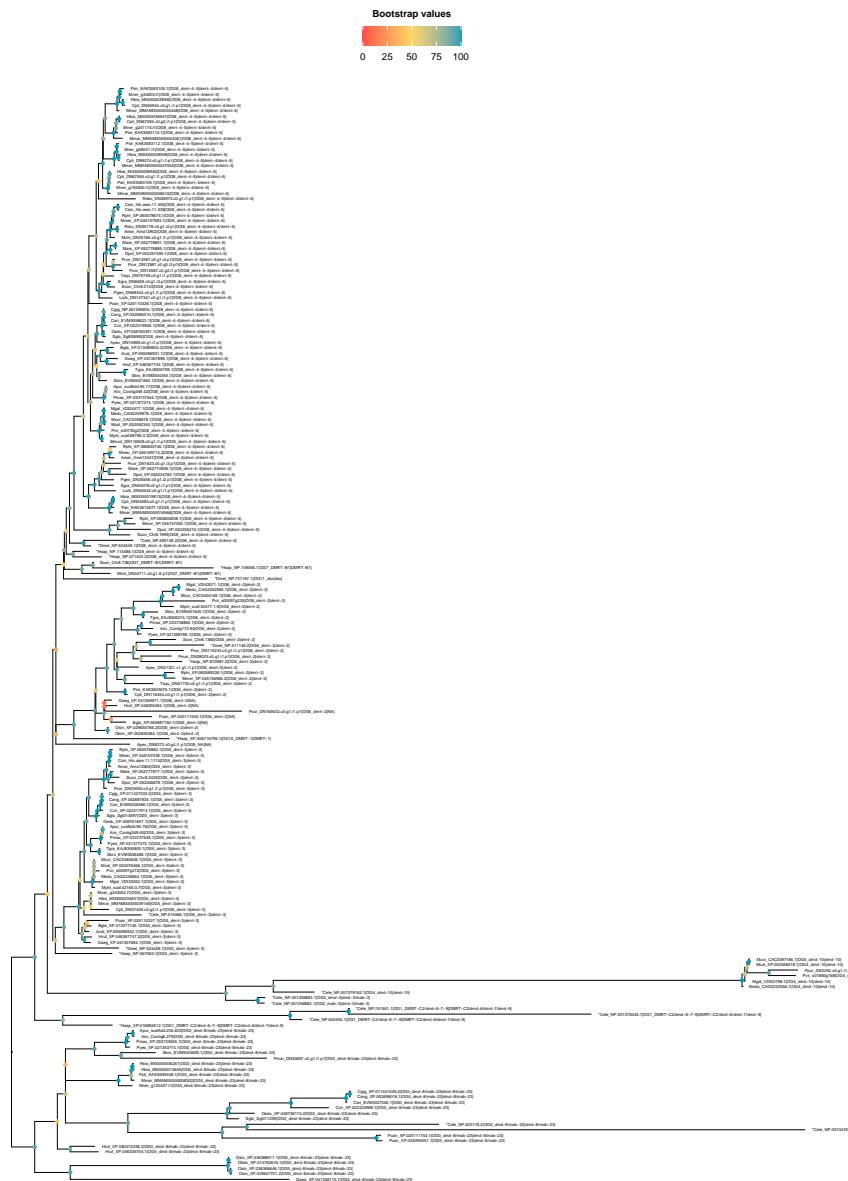
Submitted for peer-review.

Abstract. An understanding of Branchiopoda's evolutionary history is crucial for a comprehensive knowledge of the Pancrustacea tree of life, given their close evolutionary relationship with Hexapoda. Despite significant advances in molecular and morphological phylogenetics that have resolved much of the branchiopod backbone topology, a reliable temporal framework remains elusive. Key challenges include a sparse fossil record, long-term morphological stasis, and past topological inconsistencies. Leveraging a Bayesian Inference approach and the most extensive phylogenomic dataset for branchiopod to date, encompassing 46 species and over 130 genes, we inferred a time-calibrated phylogenetic tree. Furthermore, to strengthen the confidence in our divergence times estimation, we assessed the impact of age priors, topological uncertainties, and gene trees which are discordant from the species trees. Our results are largely consistent with the fossil record and with previous studies, indicating that Branchiopoda originated between 400 and 500 Mya, and the orders of large branchiopods diversified during the Mesozoic. Concerning Cladocera, results remain problematic, with a sharper uncertainty in the diversification time with respect to the fossil record. Though, the jackknife resampling of fossils and the other sensitivity analyses proved our calibration method to be robust, suggesting that the difficulties in obtaining a paleontological-consistent time tree may be hindered by the variability in branchiopod substitution rates and topological instability within certain clades.

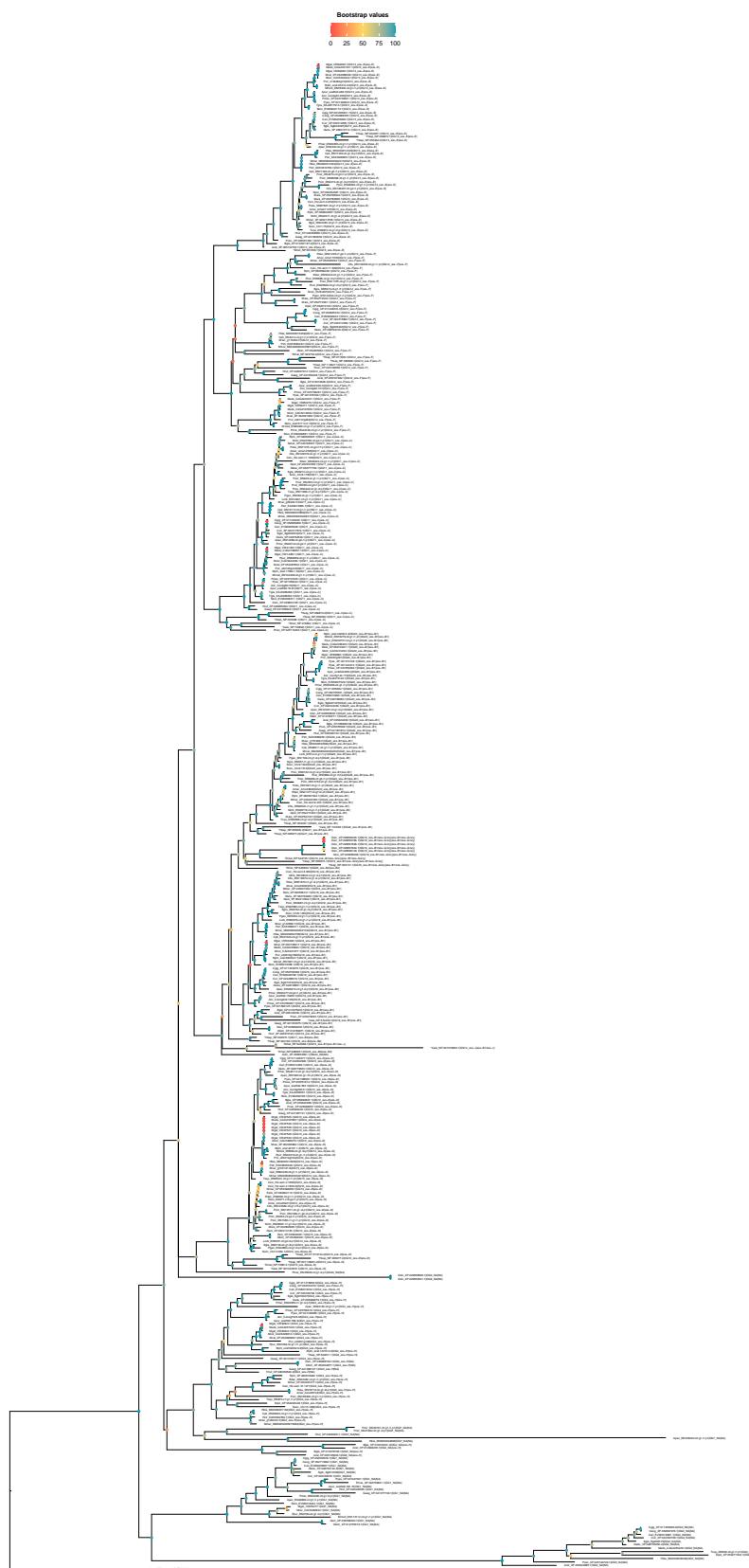
Supplementary figures

High-quality supplementary figures are available at the following GitHub repository:

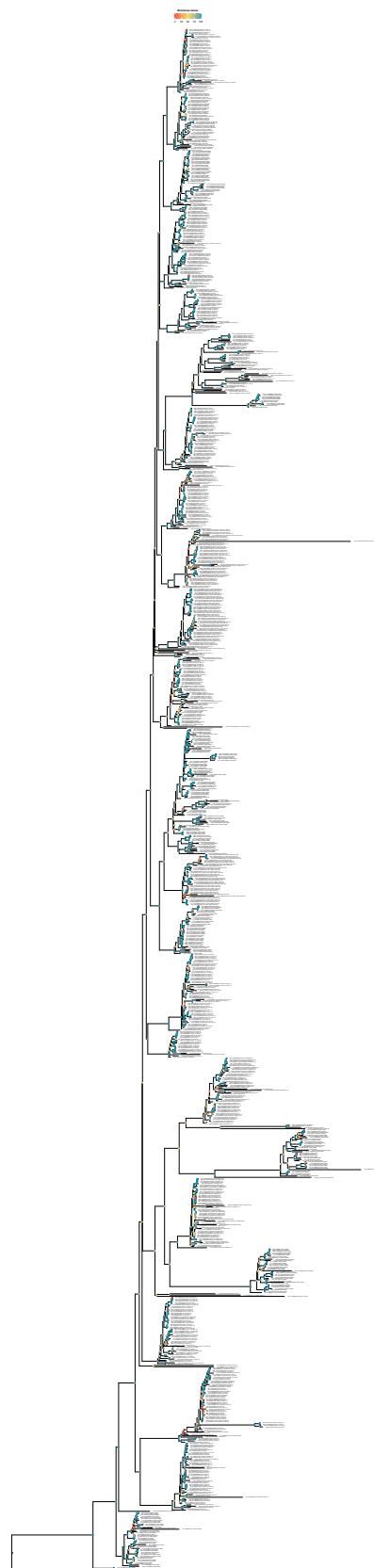
https://github.com/filonico/phd_thesis_tex



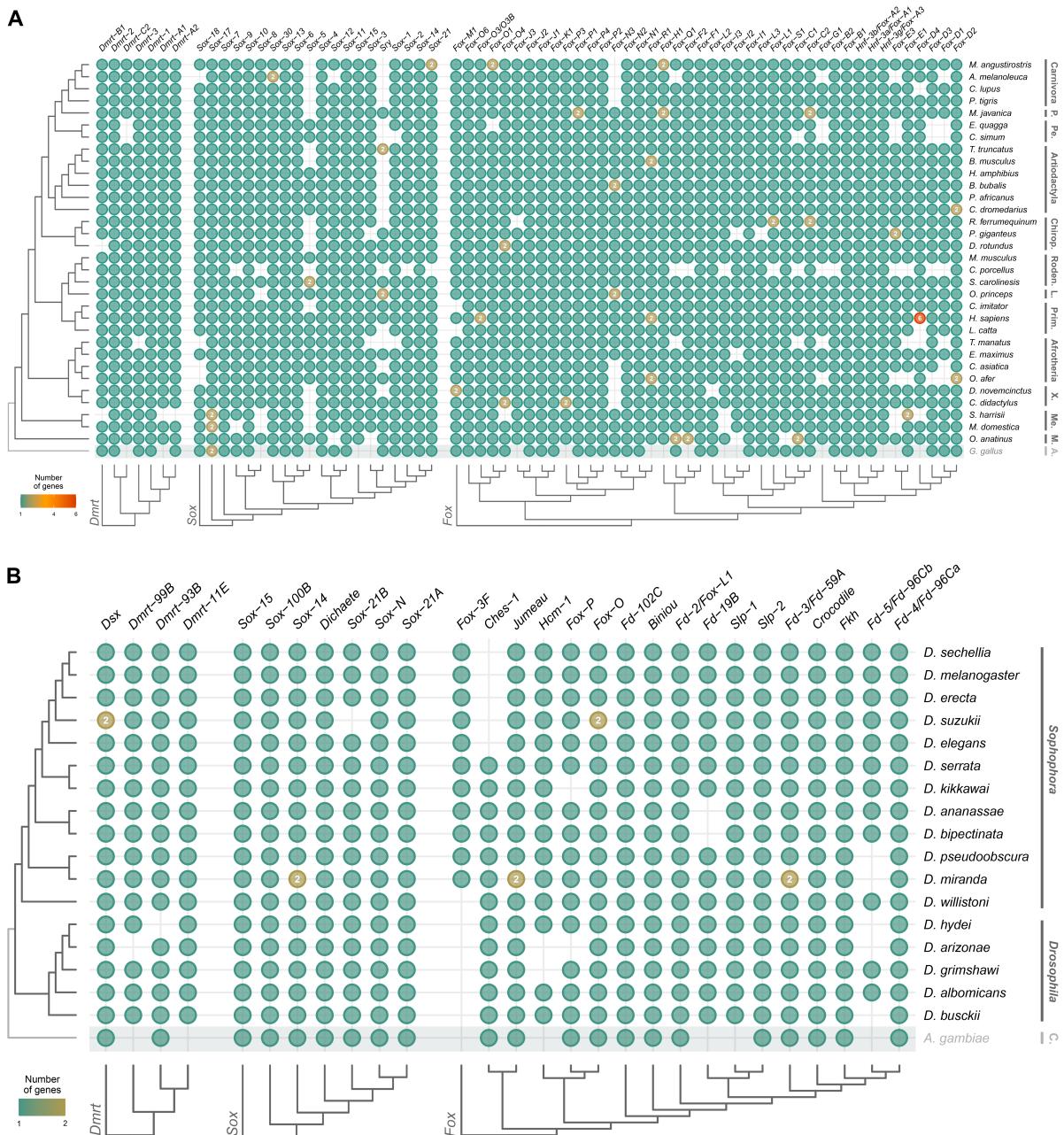
Supplementary Figure S1 – maximum likelihood (ML) phylogenetic tree of the Dmrt gene family in molluscs, including the posvm orthology inference. Reference genes from *Homo sapiens*, *Caenorhabditis elegans*, and *Drosophila melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



Supplementary Figure S2 – ML phylogenetic tree of the Sox gene family in molluscs, including the possvm orthology inference. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. Bootstrap values are shown for each node.



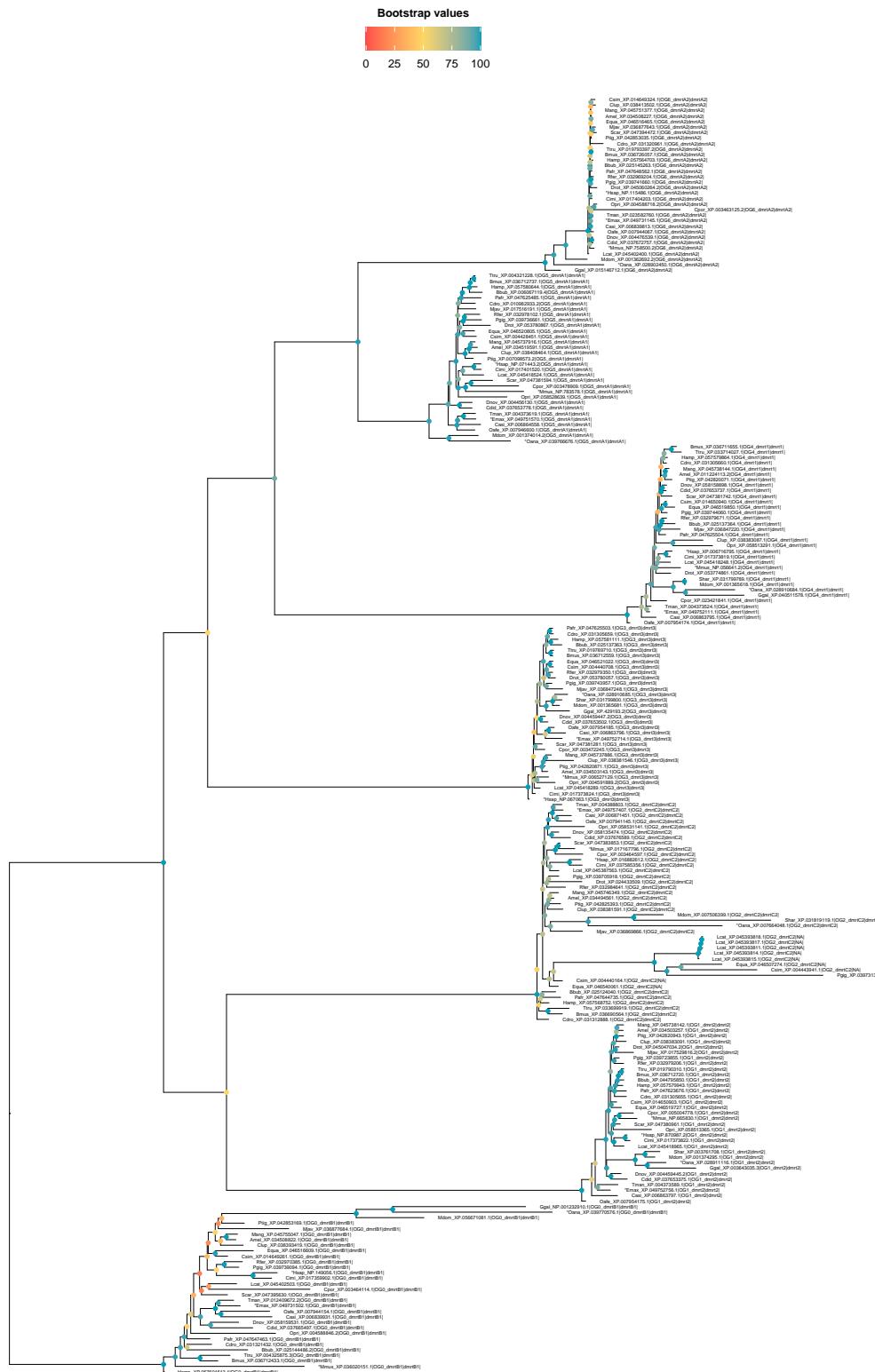
Supplementary Figure S3 – ML phylogenetic tree of the Fox gene family in molluscs, including the possvm orthology inference. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. Bootstrap values are shown for each node.



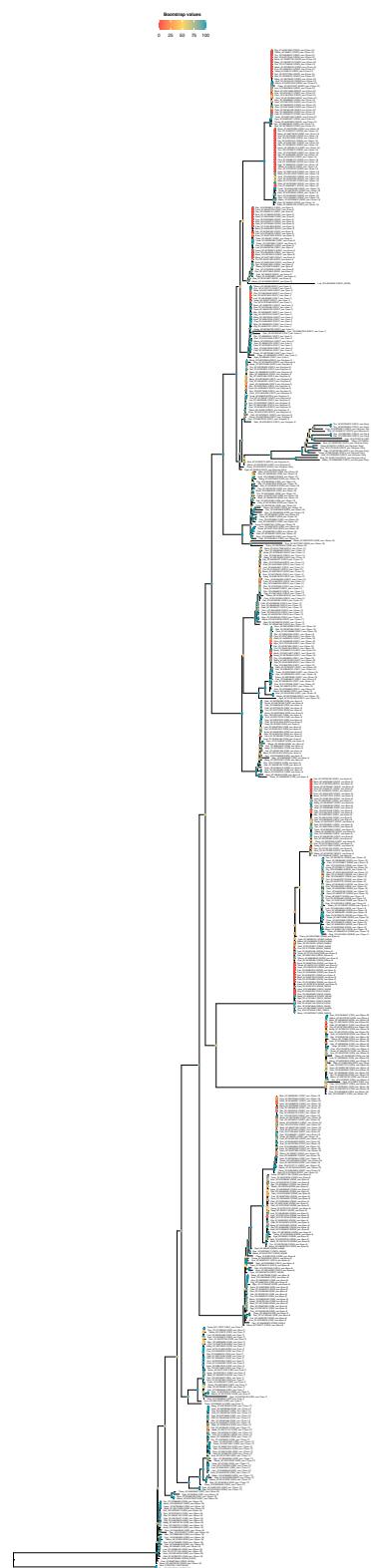
Supplementary Figure S4 – The DSFG complement in Mammalia (A) and *Drosophila* spp. (B).

Presence/absence of genes in various species are indicated by filled circles. Numbers inside each circle specify genes with 2 or more copies. The shaded area highlights outgroup species, *Gallus gallus* (Aves) for mammals and *Anopheles gambiae* (Culicidae) for fruit flies. The phylogenetic tree of analysed species, as inferred from literature, is shown on the left, while major taxonomic groups are reported on the right. All species are represented by genomic data. Dmrt, Sox, and Fox gene (DSFG) trees are shown on the bottom (full trees can be found in Fig. S5 and S7). Full species names for both mammals and fruit flies, along with all assembly and taxonomic information, can be found in Tab. S4 and S5, respectively.

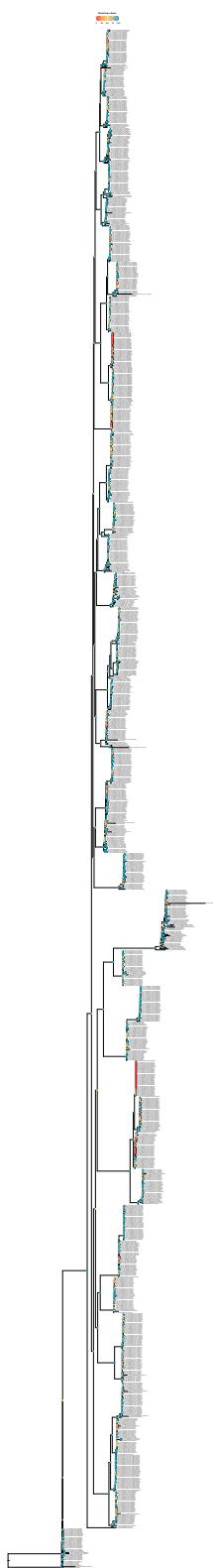
A.: Aves; Chirop.: Chiroptera; L.: Lagomorpha; M.: Monotremata; Me.: Metatheria; P.: Pholidota; Pe.: Perissodactyla; Prim.: Primates; Roden.: Rodentia; X.: Xenarthra; C.: Culicidae.



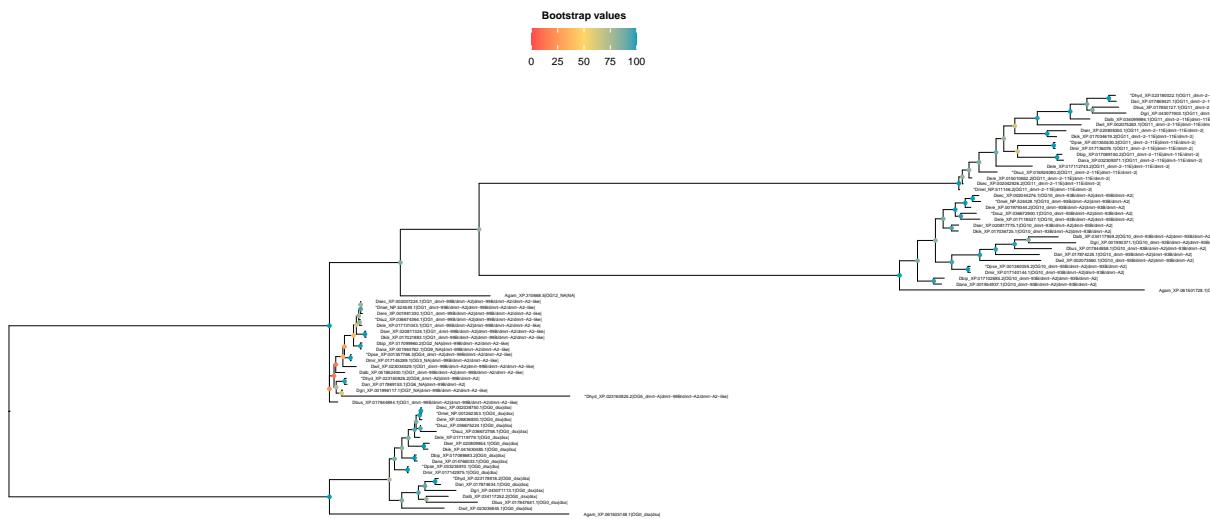
Supplementary Figure S5 – ML phylogenetic tree of the *dsx* and *mab-3* related transcription factor (Dmrt) gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *Mus musculus*, *Elephas maximus indicus*, and *Ornithorhynchus anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S4**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



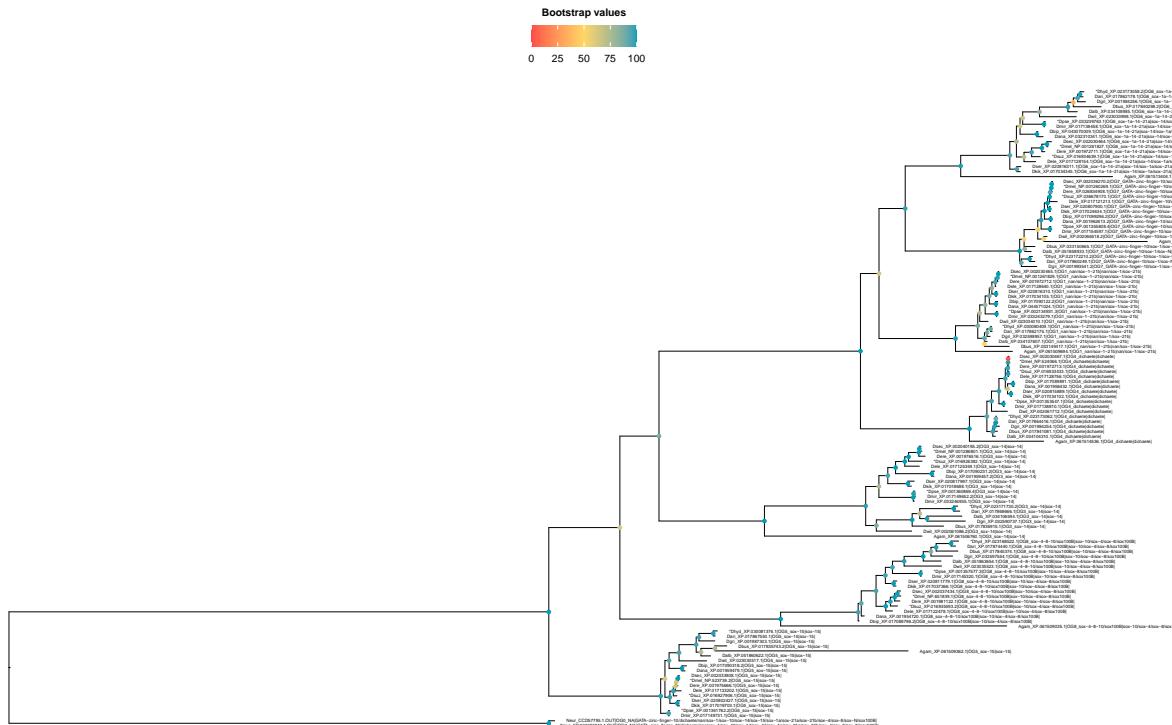
Supplementary Figure S6 – ML phylogenetic tree of the *Sry*-related HMG-box (Sox) gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *M. musculus*, *E. maximus indicus*, and *O. anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S4**. Bootstrap values are shown for each node.



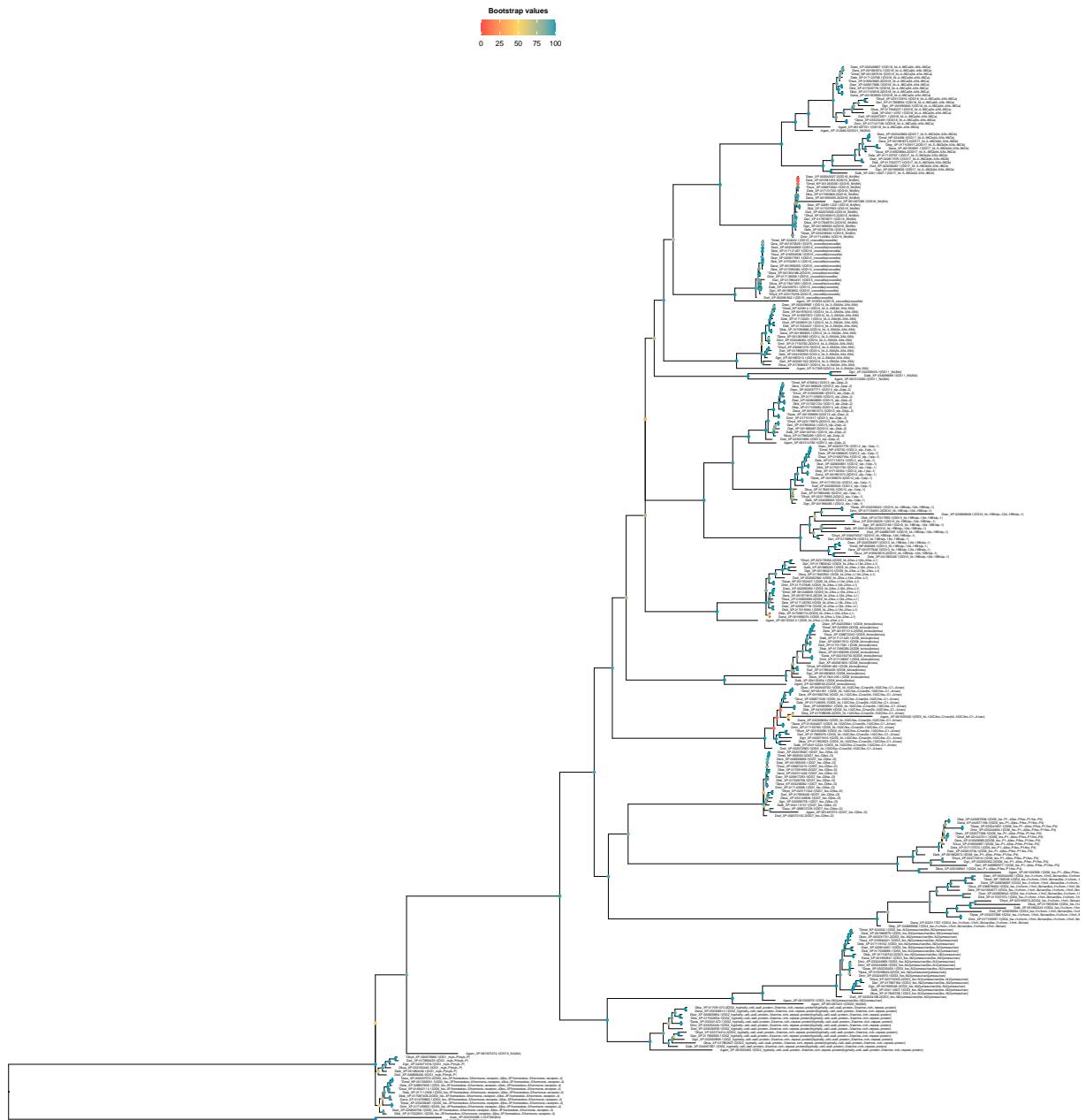
Supplementary Figure S7 – ML phylogenetic tree of the forkhead box (Fox) gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *M. musculus*, *E. maximus indicus*, and *O. anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S4**. Bootstrap values are shown for each node.



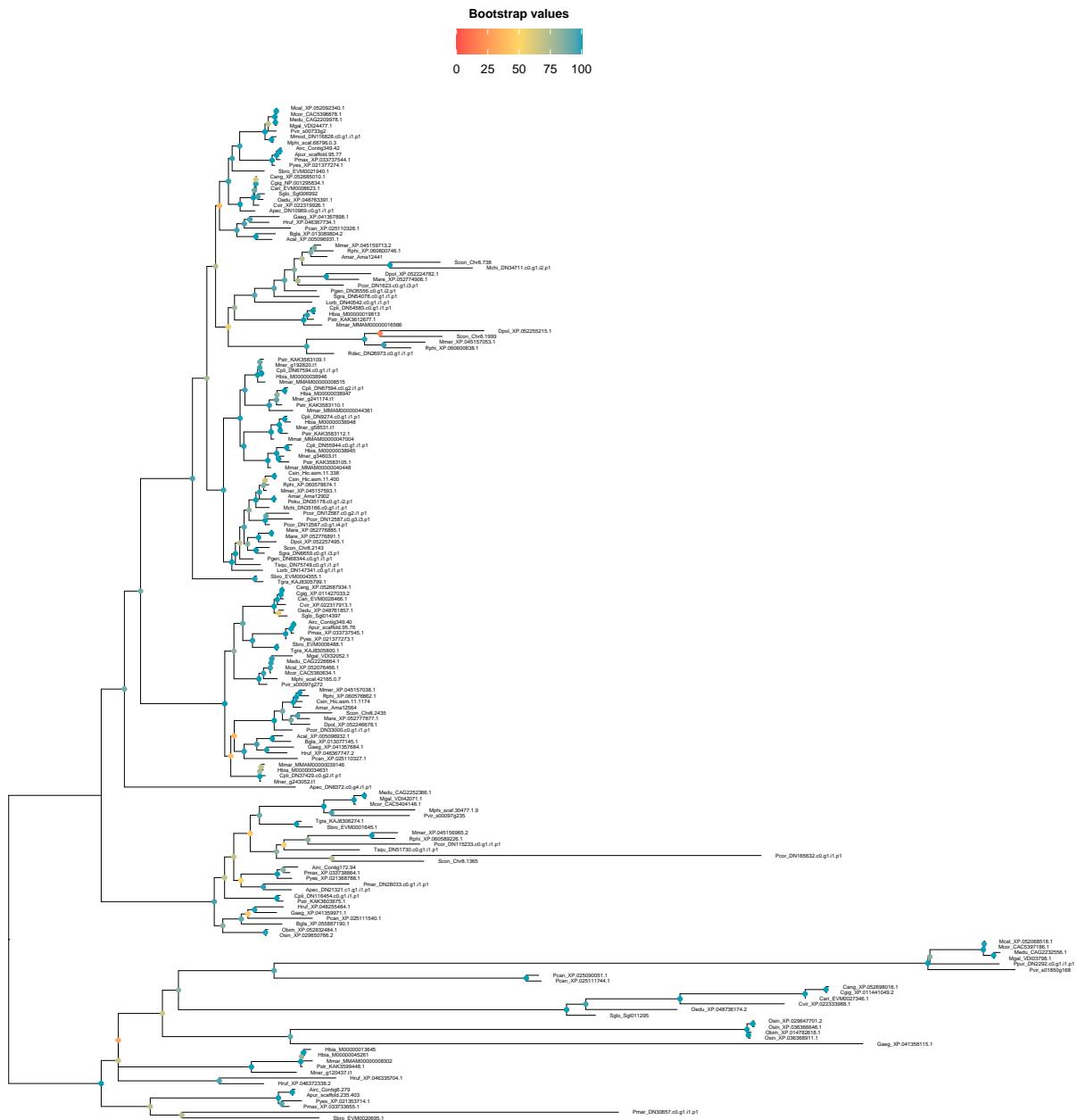
Supplementary Figure S8 – ML phylogenetic tree of the Dmrt gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *Drosophila hydei*, *Drosophila pseudoobscura*, and *Drosophila suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S5**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



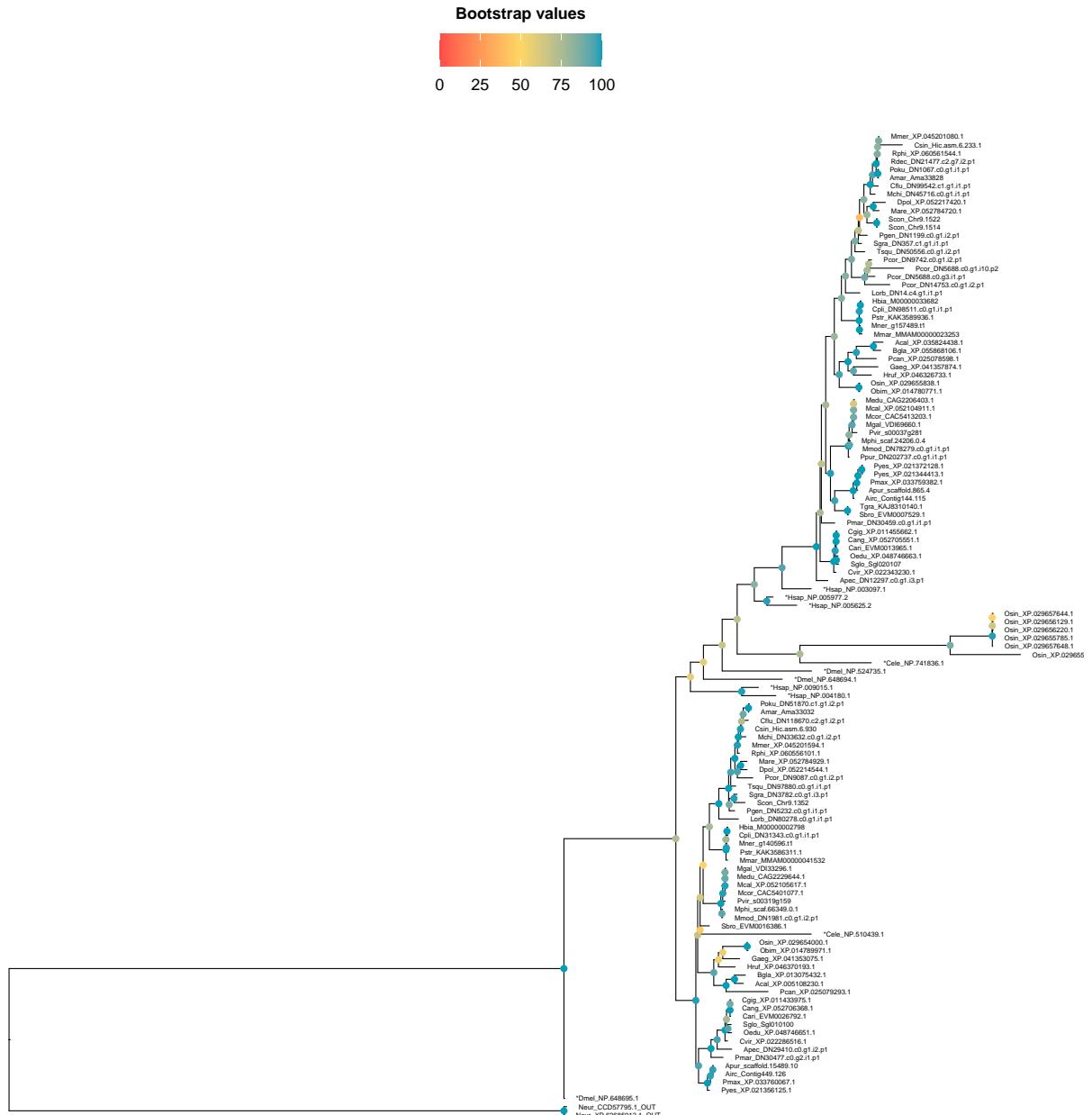
Supplementary Figure S9 – ML phylogenetic tree of the Sox gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *D. hydei*, *D. pseudoobscura*, and *D. suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S5**. Bootstrap values are shown for each node.



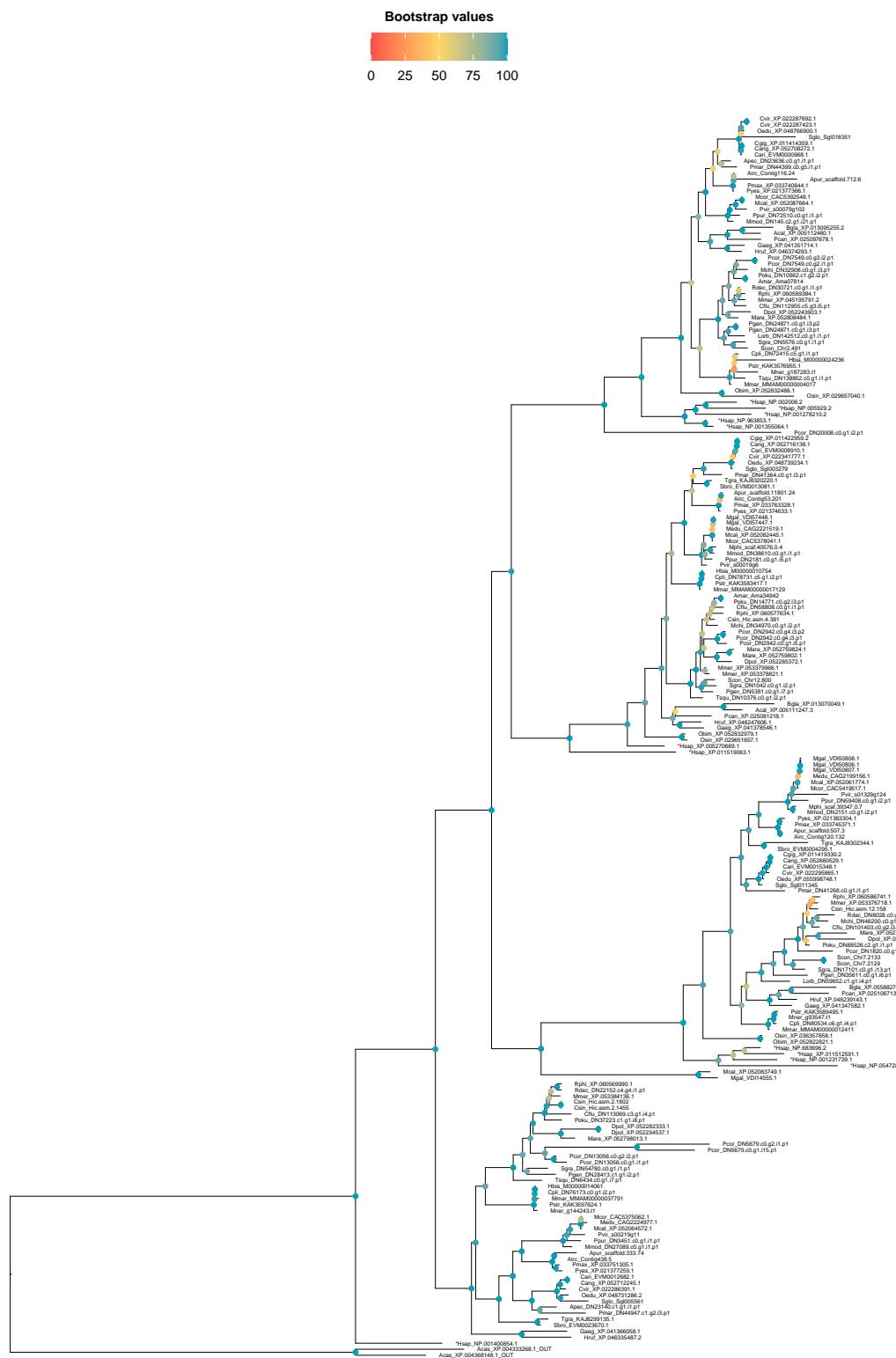
Supplementary Figure S10 – ML phylogenetic tree of the Fox gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *D. hydei*, *D. pseudoobscura*, and *D. suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S5**. Bootstrap values are shown for each node.



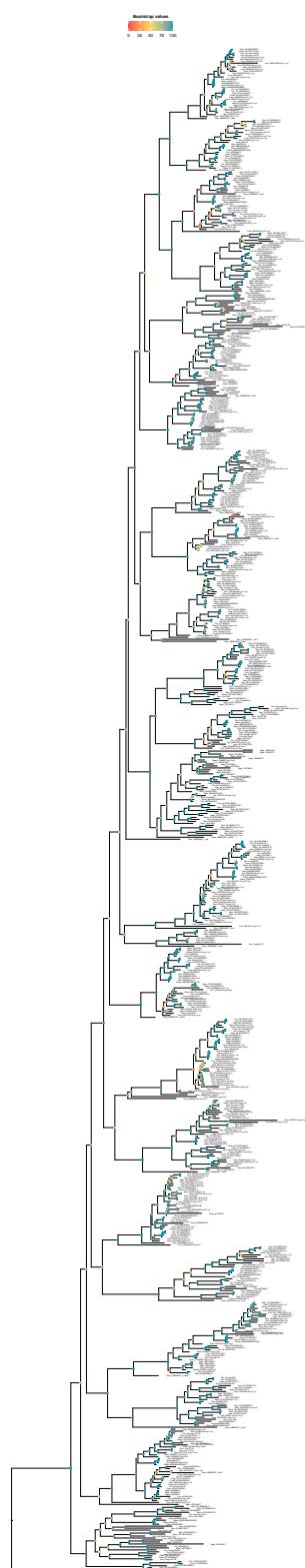
Supplementary Figure S11 – ML phylogenetic tree of the Dmrt gene family in mollusc species.
 Species ID can be found in **Tab. S1**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



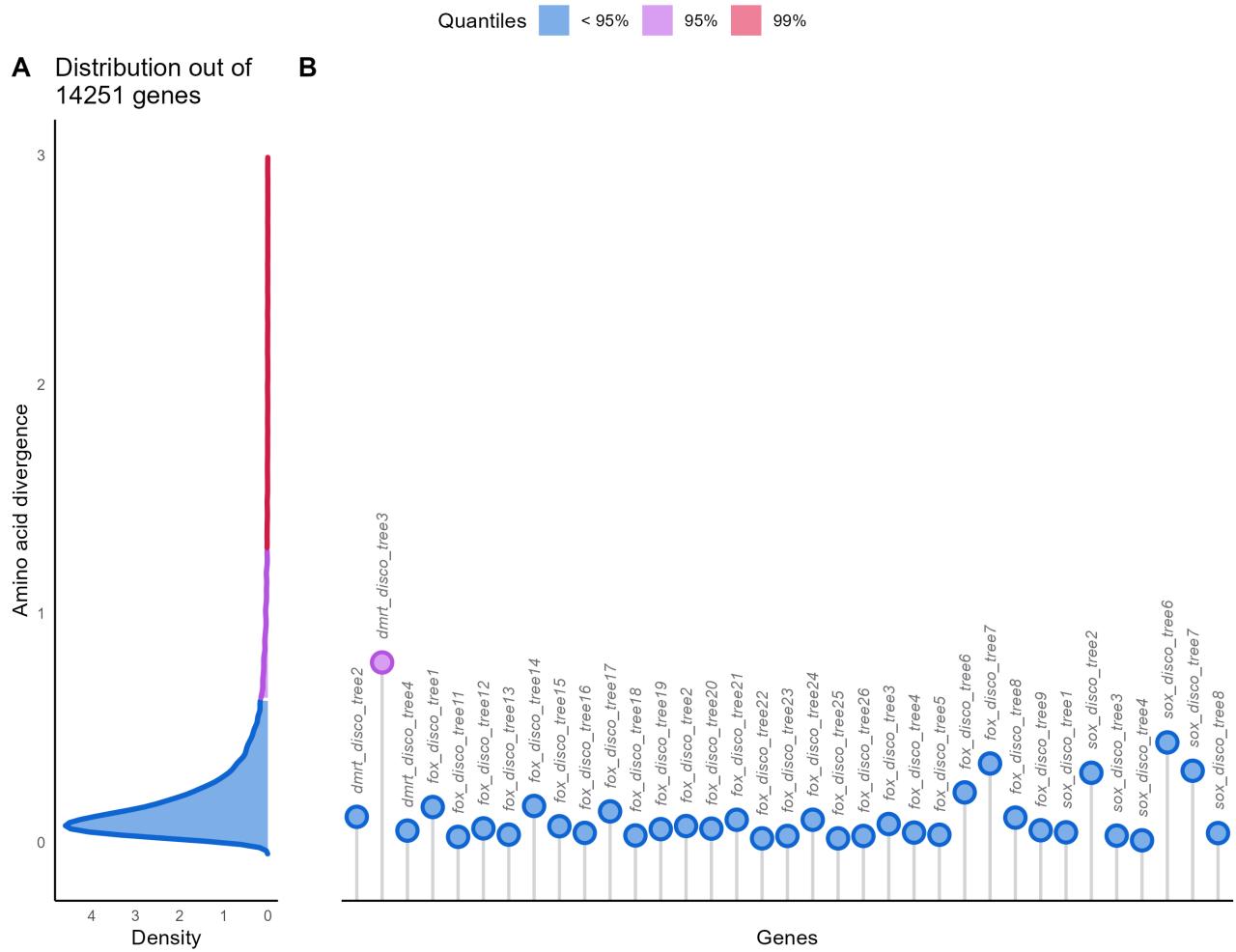
Supplementary Figure S12 – ML phylogenetic tree of *Sox-B1* and *Sox-B2* genes in mollusc and reference species. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. Bootstrap values are shown for each node.



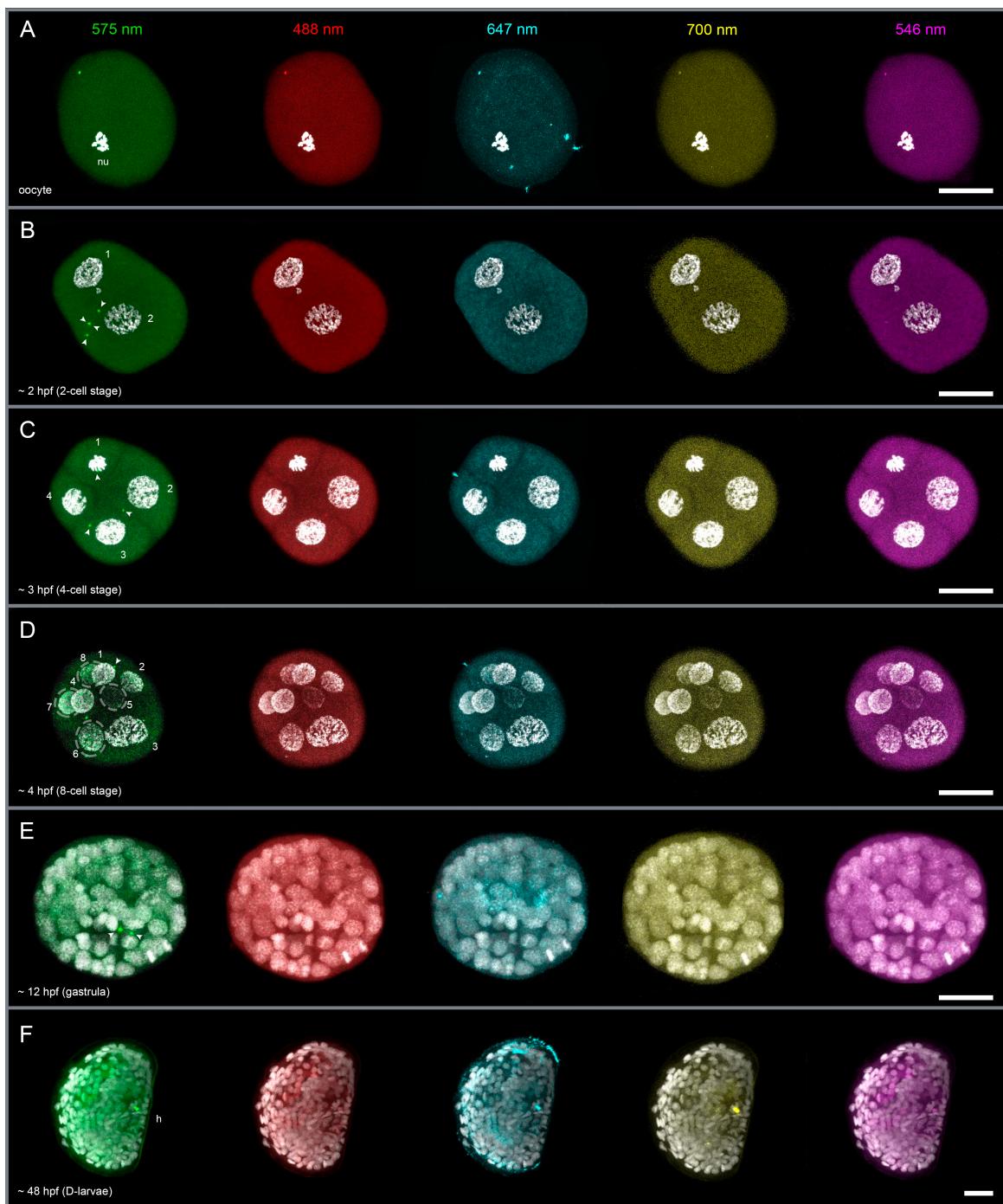
Supplementary Figure S13 – ML phylogenetic tree of *Fox-J2*, *Fox-M*, *Fox-O*, and *Fox-P* genes in mollusc and reference species. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. Bootstrap values are shown for each node.



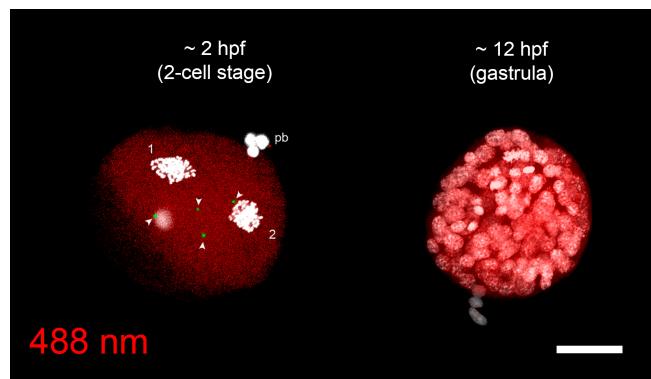
Supplementary Figure S14 – ML phylogenetic tree of the Fox gene family in bivalves and the sea urchin *Strongylocentrotus purpuratus* (Spur). Reference genes from *S. purpuratus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Tab. S1**. *S. purpuratus* genes are those given by **Tu et al., 2006**. Bootstrap values are shown for each node.



Supplementary Figure S15 – Distribution of amino acid sequence divergence (AASD) of single-copy orthogroups in *Crassostrea gigas*, *Crassostrea angulata*, *Crassostrea ariakensis*, and *Crassostrea virginica* (A), including DSFG (B). The distribution of AASD in *Crassostrea* has been computed on the median values of pairwise distances of over 14 k single-copy orthogroups (SCOs). Circle heights of DSFGs show the median value of their AASD. *Dmrt-1L* genes are indicated as 'dmrt₋disco₋tree3'.



Supplementary Figure S16 – MitoTracker staining and negative controls of mRNA *in-situ* hybridization chain reaction (HCR) in *Mytilus galloprovincialis* (A) oocyte, (B) 2-cell male embryo, (C) 4-cell female embryo, (D) 8-cell female embryo, (E) 12 hours post fertilization [hpf] embryo (gastrula), and (F) 48 hpf larvae (D-veliger). Nuclei are shown in white; in the 2-, 4-, and 8-cell stages, nuclei are also marked with numbers; in the 8-cell stage, nuclei of blastomeres in the background are highlighted with dashed circles. Sperm mitochondria, when stained (shown in green), are marked with arrowheads. Acquisition channels are indicated on top, and colours are the same as in Fig. 5.3. h: hinge; nu: oocyte nucleus. Scale bar: 20 µm.



Supplementary Figure S17 – Mito-Tracker staining and negative controls of Vasa immunolocalization in *M. galloprovincialis* embryos. Nuclei are shown in white. Sperm mitochondria (in green) are marked with arrowheads. pb: polar body. Scale bar: 20 μ m.

Supplementary tables

All the supplementary tables are available in a parsable version at the following GitHub repository:

https://github.com/filonomic/phd_thesis_tex

Supplementary Table S1 – Genomic and transcriptomic data of bivalves and other molluscs. For each species, the relative ID, taxonomic information, BUSCO statistics, NCBI accession number, and source publication are reported. Biv: Bivalvia; Ca: Caenogastropoda; Cep: Cephalopoda; Co: Coleoidea; Gas: Gastropoda; Gen: Genome; He: Heterobranchia; Im: Imparidentia; Ne: Neomphaliidae; Pa: Palaeoheterodontia; Pt: Pteriomorpha; Tra: Transcriptome; Ve: Vetigastropoda

Species	ID	Class	Group	Order	Type	Reduced dataset	BUSCO statistics (metazoa_odb10)	NCBI acc. no.	Reference	Annotation source
<i>Magallana (Crassostrea) angulata</i>	Cang	Biv	Pt	Ostreida	Gen	No	C:99.1%[S:97.1%D:2.0%], F:0.3%[M:0.6%]	GCF_025612915.1	Teng et al., 2023	NCBI
<i>Magallana (Crassostrea) gigas</i>	Cgig	Biv	Pt	Ostreida	Gen	Yes	C:98.2%[S:93.1%D:5.1%], F:0.4%[M:1.4%]	GCF_902806645.1	Penaloza et al., 2021	NCBI
<i>Magallana (Crassostrea) ariaeensis</i>	Cari	Biv	Pt	Ostreida	Gen	No	C:94.8%[S:91.2%D:3.6%], F:0.7%[M:4.5%]	GCA_020567875.1	Li et al., 2021	FigShare
<i>Crassostrea virginica</i>	Cvir	Biv	Pt	Ostreida	Gen	Yes	C:98.2%[S:73.1%D:25.1%], F:0.3%[M:1.5%]	GCF_002022765.2	Gómez-Chiarri et al., 2015	NCBI
<i>Ostrea edulis</i>	Oedu	Biv	Pt	Ostreida	Gen	Yes	C:98.7%[S:97.8%D:0.9%], F:0.5%[M:0.8%]	GCF_947568905.1	Darwin Tree of Life	NCBI
<i>Saccostrea glomerata</i>	Sglo	Biv	Pt	Ostreida	Gen	No	C:89.1%[S:85.5%D:3.6%], F:4.9%[M:6.0%]	GCA_003671525.1	Powell et al., 2018	dbSRG
<i>Atrina pectinata</i>	Apec	Biv	Pt	Ostreida	Tra	Yes	C:95.6%[S:93.1%D:2.5%], F:1.9%[M:2.5%]	DRR348924_-25_-26	Shimizu et al., 2022	–
<i>Pinctada margaritifera</i>	Pmar	Biv	Pt	Ostreida	Tra	Yes	C:94.3%[S:93.9%D:0.4%], F:1.7%[M:4.0%]	SRR1039667 SRR1041217	Teaniniuraitemoana et al., 2014	–
<i>Mytilus galloprovincialis</i>	Mgal	Biv	Pt	Mytilida	Gen	Yes	C:80.5%[S:50.4%D:30.1%], F:8.6%[M:10.9%]	GCA_900618805.1	Gerdol et al., 2020	NCBI
<i>Mytilus edulis</i>	Medu	Biv	Pt	Mytilida	Gen	No	C:83.8%[S:70.9%D:12.9%], F:5.1%[M:11.1%]	GCA_905397895.1	Corrochano-Fraile et al., 2022	NCBI
<i>Mytilus unguiculatus (coruscus)</i>	Mcor	Biv	Pt	Mytilida	Gen	No	C:80.8%[S:78.8%D:2.0%], F:4.3%[M:14.9%]	GCA_011752452.2	Yang et al., 2021	NCBI
<i>Mytilus californianus</i>	Mcal	Biv	Pt	Mytilida	Gen	No	C:96.2%[S:95.0%D:1.2%], F:0.4%[M:3.4%]	GCF_021869535.1	Paggeot et al., 2022	NCBI
<i>Perna viridis</i>	Pvir	Biv	Pt	Mytilida	Gen	Yes	C:99.4%[S:99.0%D:0.4%], F:0.2%[M:0.4%]	GCA_018327765.1	Inoue et al., 2021	Google Drive
<i>Modiolus modiolus</i>	Mmod	Biv	Pt	Mytilida	Gen	No	C:95.7%[S:92.3%D:3.4%], F:2.1%[M:2.2%]	SRR5043294	Meng et al., 2018	–
<i>Modiolus philippinarum</i>	Mphi	Biv	Pt	Mytilida	Gen	No	C:64.9%[S:63.0%D:1.9%], F:18.8%[M:16.3%]	GCA_002080025.1	Sun et al., 2017	Dryad
<i>Perumytilus purpuratus</i>	Ppur	Biv	Pt	Mytilida	Tra	Yes	C:84.2%[S:83.3%D:0.9%], F:11.8%[M:4.0%]	SRR4343820	Briones et al., 2018	–
<i>Argopecten irradians concentricus</i>	Airc	Biv	Pt	Pectinida	Gen	Yes	C:94.0%[S:94.0%D:0.9%], F:3.6%[M:1.5%]	GCA_004382765.1	Liu et al., 2020	Dryad
<i>Argopecten purpuratus</i>	Apur	Biv	Pt	Pectinida	Gen	No	C:89.2%[S:88.6%D:0.6%], F:5.0%[M:5.8%]	–	Liu et al., 2020	GigaDB

Tab. S1 continued from previous page

Species	ID	Class	Group	Order	Type	Reduced dataset	BUSCO statistics ('metazoa_odb10')	NCBI acc. no.	Reference	Annotation source
<i>Pecten maximus</i>	Pmax	Biv	Pt	Pectinida	Gen	Yes	C:98.5%[S:94.5%D:4.0%], F:0.4%[M:1.1%]	GCF_902652985.1	Kenny et al., 2020	NCBI
<i>Mizuhoplecten (Patinoplecten) yessoensis</i>	Pyes	Biv	Pt	Pectinida	Gen	Yes	C:98.3%[S:96.1%D:2.2%], F:0.5%[M:1.2%]	GCF_002113885.1	Wang, Zhang, et al., 2017	NCBI
<i>Ahadar (Scapharca) broughtonii</i>	Sbro	Biv	Pt	Arcida	Gen	Yes	C:91.2%[S:95.6%D:5.6%], F:2.6%[M:6.2%]	—	Bai et al., 2019	GigaDB
<i>Tegillarca granosa</i>	Tgra	Biv	Pt	Arcida	Gen	Yes	C:70.8%[S:61.3%D:9.3%], F:11.7%[M:17.7%]	GCA_029721355.1	—	NCBI
<i>Ruditapes decussatus</i>	Rdec	Biv	Im	Venerida	Tra	Yes	C:84.8%[S:94.1%D:0.7%], F:7.3%[M:7.9%]	SRR27740,-41,-43,-44,-47,-51,-52,-57	Ghiselli et al., 2018	—
<i>Ruditapes philippinarum</i>	Rphi	Biv	Im	Venerida	Gen	Yes	C:97.8%[S:85.5%D:12.3%], F:0.7%[M:1.5%]	GCF_026571515.1	Xu, Martelossi, et al., 2022	NCBI
<i>Mercenaria mercenaria</i>	Mmer	Biv	Im	Venerida	Gen	No	C:96.0%[S:89.8%D:6.2%], F:1.0%[M:3.0%]	GCF_021730395.1	Fahat et al., 2022	NCBI
<i>Cyclina sinensis</i>	Csin	Biv	Im	Venerida	Gen	Yes	C:94.1%[S:83.9%D:10.2%], F:1.8%[M:4.1%]	GCA_012932295.1	Wei et al., 2020	Dryad
<i>Calyptogena (Archivesicula) marissimica</i>	Amar	Biv	Im	Venerida	Tra	Yes	C:82.1%[S:80.1%D:2.0%], F:6.0%[M:11.9%]	GCA_014843695.1	Ip et al., 2021	FigShare
<i>Phreagena okutanii</i>	Poku	Biv	Im	Venerida	Gen	No	C:92.9%[S:85.8%D:7.1%], F:3.0%[M:4.1%]	SRR7156763,-64,-65,-66,-67,-68	Lan et al., 2019	—
<i>Corbicula fluminea</i>	Cflu	Biv	Im	Venerida	Tra	Yes	C:83.7%[S:79.9%D:3.8%], F:10.3%[M:6.0%]	SRR1559272,SRR5512046	González, et al., 2015 Zhu et al., 2019	—
<i>Macra chinensis</i>	Mchi	Biv	Im	Venerida	Tra	Yes	C:81.5%[S:80.8%D:0.7%], F:10.2%[M:8.3%]	SRR1263980	—	—
<i>Mya arenaria</i>	Mare	Biv	Im	Myida	Gen	Yes	C:98.5%[S:80.4%D:18.1%], F:0.4%[M:1.1%]	GCF_026914265.1	Hart et al., 2023	NCBI
<i>Dreissena polymorpha</i>	Dpol	Biv	Im	Myida	Gen	Yes	C:97.2%[S:80.1%D:17.1%], F:0.4%[M:2.4%]	GCF_020536995.1	McCartney et al., 2022	NCBI
<i>Pisidium coreanum</i>	Pcor	Biv	Im	Sphaeriida	Tra	Yes	C:94.5%[S:81.6%D:12.9%], F:3.6%[M:1.9%]	SRR6474597	—	—
<i>Solen grandis</i>	Sgra	Biv	Im	Adapedonta	Tra	Yes	C:92.7%[S:90.0%D:2.7%], F:2.5%[M:4.8%]	SRR5484647,SRR5485368,SRR5499447	Nie et al., 2018	—
<i>Sinonovacula constricta</i>	Scon	Biv	Im	Adapedonta	Gen	Yes	C:90.8%[S:79.2%D:11.6%], F:3.5%[M:5.7%]	GCA_007844125.1	Ran et al., 2019	Dryad
<i>Panopea generosa</i>	Pgen	Biv	Im	Adapedonta	Tra	Yes	C:84.1%[S:81.9%D:2.2%], F:9.7%[M:6.2%]	SRR12218869,-70	Putnam et al., 2022	—
<i>Tridacna squamosa</i>	Tsqu	Biv	Im	Cardida	Tra	Yes	C:89.7%[S:86.9%D:2.8%], F:3.5%[M:6.8%]	SRR10824662,-65	Li, Zhou, et al., 2020	—
<i>Loripes orbiculatus</i>	Lorb	Biv	Im	Lucinida	Tra	Yes	C:76.1%[S:74.9%D:1.2%], F:1.3%[M:9.6%]	SRR10002336,-38,-39,-47	Yuen et al., 2019	—
<i>Hyriopsis bivalata (Unio delphinus)</i>	Hbia	Biv	Pa	Unionida	Gen	Yes	C:97.5%[S:94.9%D:2.6%], F:2.0%[M:0.5%]	GCA_029339505.1	Gomes-dos-Santos et al., 2023	FigShare

Tab. S1 continued from previous page

Species	ID	Class	Group	Order	Type	Reduced dataset	BUSCO statistics ('metazoa_odb10')	NCBI acc. no.	Reference	Annotation source
<i>Cristaria plicata</i>	Cpli	Biv	Pa	Unionida	Tra	Yes	C:93.6%[S:92.8%D:0.8%], F:2.1%,M:4.3%	SRR2175868 SRR3095781	Patnaik et al., 2016 Wang, Liu, and Wu, 2017	—
<i>Megalonaia nervosa</i>	Mner	Biv	Pa	Unionida	Gen	Yes	C:65.0%[S:63.3%D:1.7%], F:14.0%,M:21.0%	GCA_016617855.1	Rogers et al., 2021	Dryad
<i>Potamius streckeri</i>	Pstr	Biv	Pa	Unionida	Gen	Yes	C:94.0%[S:93.4%D:1.5%], F:1.2%,M:3.9%	GCA_016746295.1	Smith, 2021	NCBI
<i>Margaritifera margaritifera</i>	Mmar	Biv	Pa	Unionida	Gen	Yes	C:92.6%[S:92.1%D:0.5%], F:3.0%,M:4.4%	GCA_015947965.1	Gomes-dos-Santos et al., 2021	FigShare
<i>Aplysia californica</i>	Acal	Gas	He	Aplysiida	Gen	No	C:97.8%[S:97.0%D:0.8%], F:0.7%,M:1.5%	GCF_000002075.1	Knudsen et al., 2006	NCBI
<i>Biomphalaria glabrata</i>	Bbla	Gas	He	—	Gen	No	C:98.9%[S:98.2%D:0.7%], F:0.1%,M:1.0%	GCF_947242115.1	—	NCBI
<i>Pomacea canaliculata</i>	Pcan	Gas	Ca	Architaenioglossa	Gen	No	C:98.2%[S:97.0%D:1.2%], F:0.4%,M:1.4%	GCF_003073045.1	Liu et al., 2018	NCBI
<i>Gigantopelta aegis</i>	Gaeg	Gas	Ne	Neomphalida	Gen	No	C:98.4%[S:94.2%D:4.2%], F:0.8%,M:0.8%	GCF_016097555.1	Lan et al., 2021	NCBI
<i>Haliotis rufescens</i>	Hruf	Gas	Ve	Lepetellida	Gen	No	C:99.0%[S:98.3%D:0.7%], F:0.0%,M:1.0%	GCF_023055435.1	—	NCBI
<i>Octopus bimaculoides</i>	Obim	Cep	Co	Octopoda	Gen	No	C:94.9%[S:94.4%D:0.5%], F:2.3%,M:2.8%	GCF_001194135.2	Albertin et al., 2015	NCBI
<i>Octopus sinensis</i>	Osin	Cep	Co	Octopoda	Gen	No	C:98.1%[S:96.9%D:1.2%], F:0.9%,M:1.0%	GCF_006345805.1	Li, Bian, et al., 2020	NCBI

Supplementary Table S2 – Dmrt, Sox, and Fox gene (DSFG) family and domain identifiers (IDs) in PANTHER and CDD, respectively. After having retrieved putative DSFGs on the basis of hidden Markov model (HMM) profiles, IDs have been used to retain only reliable hits.

Gene family	PANTHER/CDD	ID	Description
Dmrt	CDD	gn CDD 214606	Doublesex DNA-binding motif
	CDD	gn CDD 425850	DM DNA binding domain
	PANTHER	PTHR12322-SF115	DOUBLESEX AND MAB-3 RELATED TRANSCRIPTION FACTOR DMRT PROTEIN CRB-MAB-23
	PANTHER	PTHR12322-SF116	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR 1
	PANTHER	PTHR12322-SF118	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR DMD-4
	PANTHER	PTHR12322-SF123	DOUBLESEX- AND MAB-3 RELATED TRANSCRIPTION FACTOR 2
	PANTHER	PTHR12322-SF53	DOUBLESEX- AND MAB-3 RELATED TRANSCRIPTION FACTOR A1
	PANTHER	PTHR16897-SF2	STRESS RESPONSE PROTEIN NST1
	PANTHER	PTHR46888-SF11	RIBONUCLEASE H
Sox	CDD	gn CDD 432488	SOX transcription factor
	CDD	gn CDD 432558	Sox developmental protein N terminal
	CDD	gn CDD 438870	high mobility group (HMG)-box found in group B SOY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gn CDD 438837	PROTEIN CRB-MAB-23
	CDD	gn CDD 438838	high mobility group (HMG)-box found in sex-determining region Y (SRY)-box (SOX) family transcription factors
	CDD	gn CDD 438839	high mobility group (HMG)-box found in group C SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gn CDD 438840	high mobility group (HMG)-box found in group D SOY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gn CDD 438841	high mobility group (HMG)-box found in group E SOY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gn CDD 438842	high mobility group (HMG)-box found in group F SOY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gn CDD 438843	high mobility group (HMG)-box found in sex determining region Y protein (SRY) and similar proteins
Fox	CDD	gn CDD 438844	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 15 (SOX15) and similar proteins
	CDD	gn CDD 438845	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 4 (SOX4) and similar proteins
	CDD	gn CDD 438846	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 11 (SOX11) and similar proteins
	CDD	gn CDD 438847	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 12 (SOX12) and similar proteins
	CDD	gn CDD 438849	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 7 (SOX7) and similar proteins
	CDD	gn CDD 438851	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 17 (SOX17) and similar proteins
	PANTHER	PTHR10270-SF107	SOX DOMAIN CONTAINING PROTEIN DICHAETE-RELATED
	PANTHER	PTHR10270-SF161	SEX-DETERMINING REGION Y PROTEIN
	PANTHER	PTHR10270-SF199	PTHR10270-SF231
	PANTHER	PTHR10270-SF27	TRANSCRIPTION FACTOR SOX-2
Fox	PANTHER	PTHR10270-SF313	TRANSCRIPTION FACTOR SOX-4
	PANTHER	PTHR10270-SF315	PTHR10270-SF317
	PANTHER	PTHR10270-SF317	TRANSCRIPTION FACTOR SOX-1A-RELATED
	PANTHER	PTHR10270-SF322	PTHR10270-SF322
	PANTHER	PTHR10270-SF324	TRANSCRIPTION FACTOR SOX-3
	PANTHER	PTHR10270-SF326	TRANSCRIPTION FACTOR SOX-3
	PANTHER	PTHR48789P1	SOX TRANSCRIPTION FACTOR
	PANTHER	PTHR48789P2	F118025P1
	PANTHER	PTHR48803-SF1	TRANSCRIPTION FACTOR SOX-9
	PANTHER	PTHR48803-SF5	SOX10B
CDD	PANTHER	PTHR48803-SF8	TRANSCRIPTION FACTOR SOX-8
	PANTHER	PTHR48803-SF11	SOX10B
	PANTHER	PTHR47279-SF1	TRANSCRIPTION FACTOR SOX-30
	PANTHER	PTHR47279-SF30	TRANSCRIPTION FACTOR SOX-30
	CDD	gn CDD 410788	Forkhead (FH) domain found in Forkhead box (FOX) family of transcription factors and similar proteins
	CDD	gn CDD 410789	Forkhead (FH) domain found in the Forkhead box protein A (FOXA) subfamily
	CDD	gn CDD 410790	Forkhead (FH) domain found in the Forkhead box protein B (FOXB) subfamily
	CDD	gn CDD 410791	Forkhead (FH) domain found in the Forkhead box protein C (FOXC) subfamily
	CDD	gn CDD 410792	Forkhead (FH) domain found in the Forkhead box protein D (FOXD) subfamily
Fox	CDD	gn CDD 410793	Forkhead (FH) domain found in the Forkhead box protein E (FOXE) subfamily
	CDD	gn CDD 410794	Forkhead (FH) domain found in the Forkhead box protein F (FOXF) subfamily
	CDD	gn CDD 410795	Forkhead (FH) domain found in the Forkhead box protein G (FOGX) subfamily
	CDD	gn CDD 410796	Forkhead (FH) domain found in the Forkhead box protein H (FOXH) subfamily
	CDD	gn CDD 410797	Forkhead (FH) domain found in Forkhead box protein J (FOXJ1) and similar proteins
	CDD	gn CDD 410798	Forkhead (FH) domain found in Forkhead box protein K (FOXK) subfamily
	CDD	gn CDD 410799	Forkhead (FH) domain found in the Forkhead box protein L (FOXL) subfamily
	CDD	gn CDD 410800	Forkhead (FH) domain found in the Forkhead box protein M (FOXM) subfamily

Tab. S2 continued from previous page

Gene family	PANTHER/CDD	ID	Description
Fox			
PANTHER	CDD	gn CDD 410801	Forkhead (FH) domain found in Forkhead box protein L1 (FOXL1) and similar proteins
PANTHER	CDD	gn CDD 410802	Forkhead (FH) domain found in Forkhead box protein L2 (FOXL2) and similar proteins
PANTHER	CDD	gn CDD 410803	Forkhead (FH) domain found in the Forkhead box protein M (FOXFM) subfamily
PANTHER	CDD	gn CDD 410804	Forkhead (FH) domain found in Forkhead box protein N1 (FOXN1) and similar proteins
PANTHER	CDD	gn CDD 410805	Forkhead (FH) domain found in Forkhead box protein N2 (FOXN2) and similar proteins
PANTHER	CDD	gn CDD 410806	Forkhead (FH) domain found in the Forkhead box protein O (FOXO) subfamily
PANTHER	CDD	gn CDD 410807	Forkhead (FH) domain found in the Forkhead box protein P (FOXP) subfamily
PANTHER	CDD	gn CDD 410808	Forkhead (FH) domain found in Forkhead box protein Q1 (FOXQ1) and similar proteins
PANTHER	CDD	gn CDD 410809	Forkhead (FH) domain found in Forkhead box protein Q2 (FOXQ2) and similar proteins
PANTHER	CDD	gn CDD 410810	Forkhead (FH) domain found in the Forkhead box protein R (FOXQR) subfamily
PANTHER	CDD	gn CDD 410811	Forkhead (FH) domain found in Forkhead box protein S1 (FOXSL)
PANTHER	CDD	gn CDD 410812	Forkhead (FH) domain found in Forkhead box protein A1 (FOXA1) and similar proteins
PANTHER	CDD	gn CDD 410813	Forkhead (FH) domain found in Forkhead box protein A2 (FOXA2) and similar proteins
PANTHER	CDD	gn CDD 410814	Forkhead (FH) domain found in Forkhead box protein A3 (FOXA3) and similar proteins
PANTHER	CDD	gn CDD 410816	Forkhead (FH) domain found in Forkhead box protein B1 (FOXB1) and similar proteins
PANTHER	CDD	gn CDD 410817	Forkhead (FH) domain found in Forkhead box protein B2 (FOXB2) and similar proteins
PANTHER	CDD	gn CDD 410818	Forkhead (FH) domain found in Forkhead box protein C1 (FOXC1) and similar proteins
PANTHER	CDD	gn CDD 410819	Forkhead (FH) domain found in Forkhead box protein C2 (FOXC2) and similar proteins
PANTHER	CDD	gn CDD 410820	Forkhead (FH) domain found in Forkhead box proteins FOXD1, FOXD2, and similar proteins
PANTHER	CDD	gn CDD 410821	Forkhead (FH) domain found in Forkhead box protein D3 (FOXD3) and similar proteins
PANTHER	CDD	gn CDD 410822	Forkhead (FH) domain found in Forkhead box protein D4 (FOXD4) and similar proteins
PANTHER	CDD	gn CDD 410823	Forkhead (FH) domain found in Forkhead box protein F1 (FOXF1) and similar proteins
PANTHER	CDD	gn CDD 410824	Forkhead (FH) domain found in Forkhead box protein F2 (FOXF2) and similar proteins
PANTHER	CDD	gn CDD 410825	Forkhead (FH) domain found in Forkhead box protein J2 (FOXJ2) and similar proteins
PANTHER	CDD	gn CDD 410826	Forkhead (FH) domain found in Forkhead box protein J3 (FOXJ3) and similar proteins
PANTHER	CDD	gn CDD 410827	Forkhead (FH) domain found in Forkhead box protein I1 (FOXI1) and similar proteins
PANTHER	CDD	gn CDD 410828	Forkhead (FH) domain found in Forkhead box protein K1 (FOXK1) and similar proteins
PANTHER	CDD	gn CDD 410829	Forkhead (FH) domain found in Forkhead box protein K2 (FOXK2) and similar proteins
PANTHER	CDD	gn CDD 410830	Forkhead (FH) domain found in Forkhead box protein N1 (FOXNN1)
PANTHER	CDD	gn CDD 410831	Forkhead (FH) domain found in Forkhead box protein N4 (FOXNN4)
PANTHER	CDD	gn CDD 410832	Forkhead (FH) domain found in Forkhead box protein N2 (FOXNN2)
PANTHER	CDD	gn CDD 410833	Forkhead (FH) domain found in Forkhead box protein N3 (FOXNN3)
PANTHER	CDD	gn CDD 410834	Forkhead (FH) domain found in Forkhead box protein O1 (FOXO1)
PANTHER	CDD	gn CDD 410835	Forkhead (FH) domain found in Forkhead box protein O3 (FOXO3)
PANTHER	CDD	gn CDD 410836	Forkhead (FH) domain found in Forkhead box protein O4 (FOXO4) and similar proteins
PANTHER	CDD	gn CDD 410837	Forkhead (FH) domain found in Forkhead box protein O6 (FOXO6) and similar proteins
PANTHER	CDD	gn CDD 410838	Forkhead (FH) domain found in Forkhead box protein P1 (FOXP1)
PANTHER	CDD	gn CDD 410839	Forkhead (FH) domain found in Forkhead box protein P2 (FOXP2)
PANTHER	CDD	gn CDD 410840	Forkhead (FH) domain found in Forkhead box protein P3 (FOXP3) and similar proteins
PANTHER	CDD	gn CDD 410841	Forkhead (FH) domain found in Forkhead box protein P4 (FOXP4) and similar proteins
PANTHER	PTHR11829	PTHR11829-SF142	FORKHEAD BOX PROTEIN FOXQ2 PROTEIN
PANTHER	PTHR11829	PTHR11829-SF152	FORKHEAD BOX PROTEIN E3
PANTHER	PTHR11829	PTHR11829-SF206	FORKHEAD BOX PROTEIN Q1
PANTHER	PTHR11829	PTHR11829-SF335	FORKHEAD BOX PROTEIN B1
PANTHER	PTHR11829	PTHR11829-SF340	FORKHEAD BOX PROTEIN D2
PANTHER	PTHR11829	PTHR11829-SF342	FORKHEAD BOX PROTEIN H1
PANTHER	PTHR11829	PTHR11829-SF348	FORKHEAD BOX PROTEIN L2
PANTHER	PTHR11829	PTHR11829-SF361	FORKHEAD BOX PROTEIN D1
PANTHER	PTHR11829	PTHR11829-SF398	FORKHEAD BOX PROTEIN D3
PANTHER	PTHR11829	PTHR11829-SF399	FORKHEAD BOX PROTEIN PES-1
PANTHER	PTHR11829	PTHR11829-SF401	FORKHEAD TRANSCRIPTION FACTOR FKH-9
PANTHER	PTHR13962	PTHR13962-SF17	FORKHEAD BOX CLA-RELATED
PANTHER	PTHR13962	PTHR13962-SF19	FORKHEAD BOX PROTEIN N3-LIKE PROTEIN-RELATED
PANTHER	PTHR13962	PTHR13962-SF20	FORKHEAD BOX PROTEIN N2
PANTHER	PTHR13962	PTHR13962-SF22	FORKHEAD BOX PROTEIN N3-LIKE PROTEIN
PANTHER	PTHR13962	PTHR13962-SF26	FORKHEAD BOX PROTEIN N2
PANTHER	PTHR48767	PTHR48767-SF3	FORKHEAD BOX PROTEIN O
PANTHER	PTHR48767	PTHR48767-SF2	FORKHEAD BOX PROTEIN O
PANTHER	PTHR48766	PTHR48766-SF4	FORKHEAD BOX PROTEIN C
PANTHER	PTHR48766	PTHR48766-SF53	FORKHEAD BOX P, ISOFORM C
PANTHER	PTHR48766	PTHR48766-SF54	FORKHEAD BOX PROTEIN K2
PANTHER	PTHR48767	PTHR48767-SF3	FORKHEAD BOX PROTEIN K1
PANTHER	PTHR48767	PTHR48767-SF4	FORKHEAD BOX PROTEIN J2 FAMILY MEMBER
PANTHER	PTHR46078	PTHR46078	FORKHEAD BOX PROTEIN BINIQUET
PANTHER	PTHR46262	PTHR46262	FORKHEAD BOX PROTEIN BINIQUET

Tab. S2 continued from previous page

Gene family	PANTHER/CDD	ID	Description
Fox	PANTHER	PTHR4626_SF2	FORKHEAD BOX PROTEIN BIN1OU
	PANTHER	PTHR4617	FORKHEAD BOX PROTEIN G1
	PANTHER	PTHR4617_SF3	FORKHEAD BOX PROTEIN G1
	PANTHER	PTHR46721	FORKHEAD BOX PROTEIN N1
	PANTHER	PTHR46721_SF2	FORKHEAD BOX NL
	PANTHER	PTHR46805	FORKHEAD BOX PROTEIN J1
	PANTHER	PTHR46878	FORKHEAD BOX PROTEIN M1
	PANTHER	PTHR46878_SF1	FORKHEAD BOX PROTEIN M1
	PANTHER	PTHR47316	FORKHEAD BOX PROTEIN H1
	PANTHER	PTHR47316_SF1	FORKHEAD BOX PROTEIN H1

Supplementary Table S3 – List of DSFGs from reference species used to assess the identity of DSFGs in molluscs. NCBI accession numbers are reported in parenthesis. Each row represents an orthology group.

<i>Homo sapiens</i>	<i>Drosophila melanogaster</i>	<i>Caenorhabditis elegans</i>	Group
Dmrt gene family			
<i>DMRT1</i> (NP_068770.2)	–	–	1
<i>DMRT2</i> (NP_006548.1)	<i>dmrt11E</i> (NP_511146.2)	–	2
<i>DMRT3</i> (NP_067063.1)	<i>dmrt93B</i> (NP_524428.1)	<i>dmd-4</i> (NP_510466.1)	3
<i>DMRT4/A1</i> (NP_071443.2)	<i>dmrt99b</i> (NP_524549.1)	<i>dmd-5</i> (NP_495138.2)	A1/2
<i>DMRT5/A2</i> (NP_115486.1)	–	–	–
<i>DMRT6/B1</i> (NP_149056.1)	–	–	–
<i>DMRT7/C2</i> (NP_001035373.1)	–	–	–
<i>DMRT8/C1</i> (NP_149042.2)	–	–	–
–	<i>dsx</i> (NP_731197.1)	–	–
–	–	<i>mab-3</i> (NP_001256882.1)	–
–	–	<i>dmd-3</i> (NP_001256883.1)	–
–	–	<i>dmd-6</i> (NP_001370045.1)	–
–	–	<i>dmd-7</i> (NP_741551.1)	–
–	–	<i>dmd-8</i> (NP_503176.2)	–
–	–	<i>dmd-9</i> (NP_500305.1)	–
–	–	<i>dmd-11</i> (NP_001379162.1)	–
–	–	<i>mab-23</i> (NP_001041089.1)	–
Sox gene family			
<i>SRY</i> (NP_003131.1)	–	–	A
<i>SOX3</i> (NP_005625.2)	–	–	–
<i>SOX2</i> (NP_003097.1)	<i>dichaete</i> (NP_524066.1)	<i>sox3</i> (NP_510439.1)	B1
<i>SOX1</i> (NP_005977.2)	<i>soxN</i> (NP_524735.1)	<i>sox2</i> (NP_741836.1)	–
<i>SOX14</i> (NP_004180.1)	<i>sox21a</i> (NP_648694.1)	–	–
<i>SOX21</i> (NP_009015.1)	<i>sox21b</i> (NP_648695.1)	–	B2
<i>SOX11</i> (NP_003099.1)	–	–	–
<i>SOX12</i> (NP_008874.2)	<i>sox14</i> (NP_476894.1)	<i>sem-2</i> (NP_740846.1)	C
<i>SOX4</i> (NP_003098.1)	–	–	–
<i>SOX13</i> (NP_005677.2)	–	–	–
<i>SOX5</i> (NP_008871.3)	<i>sox102f</i> (NP_726612.1)	<i>egl-13</i> (NP_001024918.1)	D
<i>SOX6</i> (NP_001139291.2)	–	–	–
<i>SOX9</i> (NP_000337.1)	–	–	–
<i>SOX8</i> (NP_055402.2)	<i>sox110b</i> (NP_651839.1)	–	E
<i>SOX10</i> (NP_008872.1)	–	–	–
<i>SOX18</i> (NP_060889.1)	–	–	–
<i>SOX7</i> (NP_113627.1)	<i>sox15</i> (NP_523739.2)	–	F
<i>SOX17</i> (NP_071899.1)	–	–	–
<i>SOX15</i> (NP_008873.1)	–	–	G
<i>SOX30</i> (NP_848511.1)	–	–	H
Fox gene family			
<i>FOXA1/HNF-3α</i> (NP_004487.2)	–	–	–
<i>FOXA2/HNF-3β</i> (NP_068556.2)	<i>forkhead/fkh</i> (NP_524542.1)	<i>pha-4/Ce-fkh1</i> (NP_001041114.1)	A
<i>FOXA3/HNF-3γ</i> (NP_004488.2)	–	–	–
<i>FOXB1</i> (NP_036314.2)	<i>fd96Ca/fd4</i> (NP_524495.1)	–	–
<i>FOXB2</i> (NP_001013757.1)	<i>fd96Cb/fd5</i> (NP_524496.1)	<i>lin-31</i> (NP_494704.1)	B
<i>FOXC1/MF1/FKH-L7</i> (NP_001444.2)	–	–	–
<i>FOXC2/MFH1</i> (NP_005242.1)	<i>crocodile/fd1</i> (NP_524202.1)	–	C
<i>FOXD1/FREAC4</i> (NP_004463.1)	–	–	–
<i>FOXD2/FREAC9</i> (NP_004465.3)	–	–	–
<i>FOXD3</i> (NP_036315.1)	<i>fd59A/fd3</i> (NP_523814.1)	<i>unc-130</i> (NP_496411.1)	D
<i>FOXD4</i> (NP_997188.2)	–	–	–
<i>FOXE1/TITF2</i> (NP_004464.2)	–	–	E
<i>FOXE3</i> (NP_036318.1)	–	–	–
<i>FOXF1</i> (NP_001442.2)	<i>binious/FoxF</i> (NP_523950.2)	<i>let-381/F26B1.7</i> (NP_491826.1)	F
<i>FOXF2</i> (NP_001443.1)	–	–	–
<i>FOGX1/BF1/HBF2</i> (NP_005240.3)	<i>slp1</i> (NP_476730.1) <i>slp2</i> (NP_476834.1) <i>fd19B/cg9571</i> (NP_608369.1)	<i>fkh2/T14G12.4</i> (NP_508644.1)	G
<i>FOXH1/FAST1</i> (NP_003914.1)	–	–	H
<i>FOXI1/FREAC6/HFH3</i> (NP_036320.2)	–	–	I
<i>FOXJ1</i> (NP_001445.2)	–	–	J1
<i>FOXJ2</i> (XP_011519063.1)	–	–	J2
<i>FOXJ3</i> (XP_005270689.1)	–	–	J3
<i>FOXK1/ILF1</i> (NP_001032242.1)	<i>foxK/LD16137</i> (NP_001261701.1)	–	K
<i>FOXK2</i> (NP_004505.2)	–	–	–

Tab. S3 continued from previous page

<i>Homo sapiens</i>	<i>Drosophila melanogaster</i>	<i>Caenorhabditis elegans</i>	Group
Fox gene family			
<i>FOXL1</i> (NP_005241.1)	<i>foxL1/fd2</i> (NP_523912.1)	–	L1
<i>FOXL2</i> (NP_075555.1)	–	–	L2
<i>FOXM1</i> (NP_001400854.1)	–	–	M
<i>FOXN1/WHN</i> (NP_001356298.1)	<i>jumeau</i> (NP_524302.1)	–	N1/4
<i>FOXN4</i> (NP_998761.2)			
<i>FOXN2/HTLF</i> (NP_001362376.1)	<i>ches-1</i> (NP_511071.3)	–	N2/3
<i>FOXN3/CHES1</i> (NP_001078940.1)			
<i>FOXO1</i> (NP_002006.2)			
<i>FOXO3</i> (NP_963853.1)	–	<i>daf-16</i> (NP_001364785.1)	O
<i>FOXO3B</i> (NP_001355064.1)			
<i>FOXP1</i> (NP_001231739.1)			
<i>FOXP2</i> (NP_683696.2)	<i>foxP/cg16899</i> (NP_001247011.1)	<i>F26D12.1</i> (NP_001293813.1)	P
<i>FOXP3</i> (NP_054728.2)			
<i>FOXP4</i> (XP_011512591.1)			
<i>FOXQ/HFH11</i> (NP_150285.3)	–	–	Q1
–	<i>fd102C/cd11152</i> (NP_651951.1)	<i>fkh-10/C25A1.2</i> (NP_492676.2)	Q2
<i>FOXSI/FREAC10</i> (NP_004109.1)	–	–	S
–	–	<i>PES-1</i> (NP_001023406.1)	–
–	–	<i>B0286.5/FKH-6</i> (NP_494775.1)	–
–	–	<i>F40H3.4/FKH-8</i> (NP_001254107.1)	–
–	–	<i>C29F7.4/FKH-3</i> (NP_001294822.1)	–
–	–	<i>K03C7.2/FKH-9</i> (NP_001024760.1)	–

Supplementary Table S4 – Genomic data of mammals used to retrieve DSFGs and compute amino acid sequence divergence (AASD) of single-copy orthogroups (SCOs). For each species, the relative ID, taxonomic information, BUSCO statistics, NCBI accession number, and source publication are reported.

Species	ID	Class	Group	Order	Type	BUSCO statistics ('mammalia_odb10')	NCBI acc. no.	Reference
<i>Gallus gallus</i>	Ggal	Aves	Neognathae	Galliformes	Genome	C:99.0%[S:98.6%,D:0.4%],F:0.2%,M:0.8%	GCF_016699485.2	Vertebrate Genome Project
<i>Chrysocloris asiatica</i>	Casi	Mammalia	Afrotheria	Afrosciricida	Genome	C:98.0%[S:97.4%,D:0.0%],F:1.1%,M:0.9%	GCF_000296735.1	Murata et al., 2003
<i>Elephas maximus indicus</i>	Emax	Mammalia	Afrotheria	Proboscidea	Genome	C:98.9%[S:98.3%,D:0.7%],F:0.4%,M:0.7%	GCF_024166365.1	Vertebrate Genome Project
<i>Trichechus manatus latirostris</i>	Tman	Mammalia	Afrotheria	Sirenia	Genome	C:96.1%[S:95.7%,D:0.4%],F:1.8%,M:2.1%	GCF_000243295.1	Foote et al., 2015
<i>Orycteropus afer afer</i>	Oafe	Mammalia	Afrotheria	Tubulidentata	Genome	C:96.5%[S:96.0%,D:0.5%],F:1.9%,M:1.6%	GCF_000298275.1	–
<i>Ochotonota princeps</i>	Opri	Mammalia	Euarctontogires	Lagomorpha	Genome	C:98.3%[S:96.4%,D:1.9%],F:0.5%,M:1.2%	GCF_030435755.1	Vertebrate Genome Project
<i>Cebus imitator</i>	Cimi	Mammalia	Euarctontogires	Primates	Genome	C:97.3%[S:95.1%,D:2.2%],F:1.7%,M:1.0%	GCF_001664975.1	Orkin et al., 2021
<i>Homo sapiens</i>	Hsap	Mammalia	Euarctontogires	Primates	Genome	C:99.6%[S:97.3%,D:2.3%],F:0.2%,M:0.2%	GCF_000001405.40	Genome Reference Consortium
<i>Lemur catta</i>	Lcat	Mammalia	Euarctontogires	Primates	Genome	C:98.3%[S:97.2%,D:1.1%],F:0.4%,M:1.3%	GCF_020740050.2	Vertebrate Genome Project
<i>Cavia porcellus</i>	Cpor	Mammalia	Euarctontogires	Rodentia	Genome	C:96.4%[S:95.7%,D:0.7%],F:1.7%,M:1.9%	GCF_000151735.1	The Genome Sequencing Platform
<i>Mus musculus</i>	Mmus	Mammalia	Euarctontogires	Rodentia	Genome	C:99.4%[S:98.7%,D:0.7%],F:0.2%,M:0.4%	GCF_000001035.27	Genome Reference Consortium
<i>Sciurus carolinensis</i>	Scar	Mammalia	Euarctontogires	Rodentia	Genome	C:99.1%[S:96.9%,D:2.2%],F:0.3%,M:0.6%	GCF_902666445.1	Mead et al., 2020
<i>Bubalus bubalis</i>	Bbub	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.7%[S:97.0%,D:1.7%],F:0.6%,M:0.7%	GCF_019923935.1	Deng et al., 2016
<i>Balaenoptera musculus</i>	Bmus	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.4%[S:95.7%,D:2.7%],F:0.6%,M:1.0%	GCF_009873245.2	Genome 10K
<i>Camelus dromedarius</i>	Cdro	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.7%[S:98.3%,D:0.4%],F:0.7%,M:0.6%	GCF_000803125.2	Elbers et al., 2019
<i>Hippopotamus amphibius kiboko</i>	Hamp	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.7%[S:95.2%,D:3.5%],F:0.5%,M:0.8%	GCF_030028045.1	Vertebrate Genome Project
<i>Phacochoerus africanus</i>	Paff	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.8%[S:98.3%,D:0.5%],F:0.6%,M:0.6%	GCF_016966955.1	–
<i>Tursiops truncatus</i>	Ttru	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:97.3%[S:95.2%,D:2.1%],F:1.1%,M:1.6%	GCF_011722959.1	Xiong et al., 2009
<i>Alliropoda melanoleuca</i>	Amel	Mammalia	Laurasiatheria	Carnivora	Genome	C:97.3%[S:96.8%,D:0.7%],F:1.3%,M:1.4%	GCF_002007445.2	Fan et al., 2019
<i>Canis lupus familiaris</i>	Clup	Mammalia	Laurasiatheria	Carnivora	Genome	C:98.5%[S:96.7%,D:1.8%],F:0.6%,M:0.9%	GCF_0111010685.1	Wang et al., 2021
<i>Mirounga angustirostris</i>	Maing	Mammalia	Laurasiatheria	Carnivora	Genome	C:96.7%[S:94.5%,D:2.2%],F:1.1%,M:1.4%	GCF_021288785.2	Moreno et al., 2024
<i>Panthera tigris</i>	Ptig	Mammalia	Laurasiatheria	Carnivora	Genome	C:99.4%[S:98.9%,D:0.5%],F:0.3%,M:0.3%	GCF_018350195.1	Bredemeyer et al., 2023
<i>Desmodus rotundus</i>	Drot	Mammalia	Laurasiatheria	Chiroptera	Genome	C:98.2%[S:97.2%,D:1.0%],F:0.5%,M:1.3%	GCF_022662495.1	Bat 1K
<i>Pteropus giganteus</i>	Pgig	Mammalia	Laurasiatheria	Chiroptera	Genome	C:97.2%[S:96.9%,D:0.3%],F:1.1%,M:1.7%	GCF_902729225.1	Foruret et al., 2020
<i>Rhinolophus ferrumequinum</i>	Rfer	Mammalia	Laurasiatheria	Chiroptera	Genome	C:99.2%[S:97.9%,D:1.3%],F:0.3%,M:0.5%	GCF_004115265.2	Vertebrate Genome Project
<i>Ceratotherium simum simum</i>	Csim	Mammalia	Laurasiatheria	Perissodactyla	Genome	C:98.8%[S:98.6%,D:0.2%],F:0.9%,M:0.3%	GCF_000283155.1	–
<i>Equus quagga</i>	Equa	Mammalia	Laurasiatheria	Perissodactyla	Genome	C:98.5%[S:95.9%,D:3.5%],F:0.5%,M:1.0%	GCF_021613505.1	Vilstrup et al., 2013
<i>Manis javanica</i>	Mjav	Mammalia	Laurasiatheria	Pholidota	Genome	C:95.7%[S:93.7%,D:2.0%],F:1.9%,M:2.4%	GCF_014570535.1	–
<i>Sarcophilus harrisii</i>	Shar	Mammalia	Metatheria	Dasyuromorphia	Genome	C:95.5%[S:94.5%,D:1.0%],F:0.9%,M:3.6%	GCF_90265505.1	Stammitz et al., 2023
<i>Monodelphis domestica</i>	Mdom	Mammalia	Metatheria	Didelphimorphia	Genome	C:95.1%[S:92.3%,D:2.8%],F:0.9%,M:4.0%	GCF_027887165.1	Vertebrate Genome Project
<i>Omnithorhynchus anatinus</i>	Oana	Mammalia	Prototheria	Monotremata	Genome	C:92.3%[S:91.2%,D:1.1%],F:1.4%,M:6.3%	GCF_004115215.2	Zhou et al., 2021
<i>Dasypus novemcinctus</i>	Dnov	Mammalia	Xenarthra	Cingulata	Genome	C:96.9%[S:94.3%,D:2.6%],F:0.7%,M:2.4%	GCF_030445035.1	Vertebrate Genome Project
<i>Choloepus didactylus</i>	Cdid	Mammalia	Xenarthra	Pilosa	Genome	C:97.8%[S:91.9%,D:5.9%],F:0.7%,M:1.5%	GCF_015220235.1	Vertebrate Genome Project

Supplementary Table S5 – Genomic data of *Drosophila* used to retrieve DSFGs and compute AASD of SCOs. For each species, the relative ID, taxonomic information, BUSCO statistics, NCBI accession number, and source publication are reported.

Species	ID	Family	Subgenus	Type	BUSCO statistics ('diptera_odb10')	NCBI acc. no.	Reference
<i>Anopheles gambiae</i>	Agam	Culicidae	Celilia	Genome	C:99.4%[S: 99.1%,D:0.3%],F:0.1%,M:0.5%	GCF_943734735.2	Habtewold et al., 2023
<i>Drosophila sechellia</i>	Dsec	Drosophilidae	Sophophora	Genome	C:99.9%[S: 99.3%,D:0.6%],F:0.0%,M:0.1%	GCF_004382195.2	Chakraborty et al., 2021
<i>Drosophila melanogaster</i>	Dmel	Drosophilidae	Sophophora	Genome	C:100.0%[S: 99.7%,D:0.3%],F:0.0%,M:0.0%	GCF_000001215.4	Hoskins et al., 2015
<i>Drosophila erecta</i>	Dere	Drosophilidae	Sophophora	Genome	C:99.9%[S: 99.5%,D:0.4%],F:0.0%,M:0.1%	GCF_003286155.1	Dong et al., 2022
<i>Drosophila suzukii</i>	Dsuz	Drosophilidae	Sophophora	Genome	C:99.7%[S: 96.5%,D:3.2%],F:0.1%,M:0.2%	GCF_013340165.1	Paris et al., 2020
<i>Drosophila elegans</i>	Dele	Drosophilidae	Sophophora	Genome	C:99.8%[S: 99.5%,D:0.3%],F:0.1%,M:0.1%	GCF_018152505.1	Kim et al., 2021
<i>Drosophila serrata</i>	Dser	Drosophilidae	Sophophora	Genome	C:99.9%[S: 97.5%,D:2.4%],F:0.0%,M:0.1%	GCF_002093755.2	Allen et al., 2017
<i>Drosophila kikkawai</i>	Dkik	Drosophilidae	Sophophora	Genome	C:100.0%[S: 99.1%,D:0.9%],F:0.0%,M:0.0%	GCF_018152535.1	Kim et al., 2021
<i>Drosophila bipectinata</i>	Dbip	Drosophilidae	Sophophora	Genome	C:99.9%[S: 99.2%,D:0.7%],F:0.0%,M:0.1%	GCF_018153845.1	Kim et al., 2021
<i>Drosophila ananassae</i>	Dana	Drosophilidae	Sophophora	Genome	C:99.6%[S: 99.3%,D:0.3%],F:0.0%,M:0.4%	GCF_017639315.1	Tvedte et al., 2021
<i>Drosophila pseudoobscura</i>	Dpse	Drosophilidae	Sophophora	Genome	C:99.7%[S: 98.8%,D:0.9%],F:0.1%,M:0.2%	GCF_009870125.1	Liao et al., 2021
<i>Drosophila miranda</i>	Dmir	Drosophilidae	Sophophora	Genome	C:99.8%[S: 85.6%,D:14.2%],F:0.1%,M:0.1%	GCF_003369915.1	Mahajan et al., 2018
<i>Drosophila willistoni</i>	Dwil	Drosophilidae	Sophophora	Genome	C:99.6%[S: 98.4%,D:1.2%],F:0.0%,M:0.4%	GCF_018902025.1	Ranz et al., 2023
<i>Drosophila arizonae</i>	Dari	Drosophilidae	Drosophila	Genome	C:95.7%[S: 95.3%,D:0.4%],F:1.2%,M:3.1%	GCF_001654025.1	Sanchez-Flores et al., 2016
<i>Drosophila hydei</i>	Dhyd	Drosophilidae	Drosophila	Genome	C:99.7%[S: 97.5%,D:2.2%],F:0.1%,M:0.2%	GCF_003285905.1	Dong et al., 2022
<i>Drosophila grimshawi</i>	Dgri	Drosophilidae	Drosophila	Genome	C:99.9%[S: 99.2%,D:0.7%],F:0.0%,M:0.1%	GCF_018153295.1	Kim et al., 2021
<i>Drosophila albomicans</i>	Dalb	Drosophilidae	Drosophila	Genome	C:99.9%[S: 99.1%,D:0.8%],F:0.0%,M:0.1%	GCF_009650485.2	Mai et al., 2020
<i>Drosophila busckii</i>	Dbus	Drosophilidae	Drosophila	Genome	C:98.1%[S: 97.4%,D:0.7%],F:0.3%,M:1.6%	GCF_011750605.1	Renschler et al., 2019

Supplementary Table S6 – Complete set of DSFGs in bivalves. For each gene, the species ID (Sp. ID) as in Tab. S1, the accession number (Gene ID), the Possvm-based annotation, and the CDD domains (including their Psm-ID) are indicated.

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Aicr	Contig6_279	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
	scaffold_235_403	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Apur	XP_052698016.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Cang	EVM0027346.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Cari	XP_001441049.2	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Ceig	XP_022333989.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Cirf	XP_0014358115.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Gaeg	M00000045261	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Hibia	XP_046372382.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Hiruf	XP_046335704.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Ncal	XP_052068518.1	Dmrt	Dmrt-OG4 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Mcor	CA5397186.1	Dmrt	Dmrt-OG4 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Medu	CA6232556.1	Dmrt	Dmrt-OG4 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Ngal	VD103798.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Nmar	MMAM0000008302	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Nmer	g120437_11	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Obim	XP_048778618.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Oedu	XP_048736174.2	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Csin	XP_0363889111	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Csin	XP_0363866461	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Csin	XP_02964701.2	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fcan	XP_025090051.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fmar	DN30587_cg.i1.i1.p1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fmax	XP_033733655.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Ppur	DN2292_cg.i1.i1.p1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fstr	KAK3590448.1	Dmrt	Dmrt-OG4 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fvir	XP_018598168	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Fyes	EV/M020695.1	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Sbro	Sgi01295	Dmrt	Dmrt-OG0 / NA	Doublesex DNA-binding motif (214606)	-	Annotated as Dmrt-1L
Sgio	Contig172_04	Dmrt	Dmrt-2	-	-	-
Airc	DN21321_c1_cg.i1.i1.p1	Dmrt	Dmrt-2	-	-	-
Apec	XP_05887190.1	Dmrt	Dmrt-2	-	-	-
Bgia	DN116054_cg.i1.i1.p1	Dmrt	Dmrt-2	-	-	-
Cphi	XP_041359979.1	Dmrt	Dmrt-2	-	-	-
Gaeg	XP_048255484.1	Dmrt	Dmrt-2	-	-	-
Hiruf	XP_033738864.1	Dmrt	Dmrt-2	-	-	-
Mcor	CAC404148.1	Dmrt	Dmrt-2	-	-	-
Medu	CA6232366.1	Dmrt	Dmrt-2	-	-	-
Ngal	VD142071.1	Dmrt	Dmrt-2	-	-	-
Nmer	XP_045156965.2	Dmrt	Dmrt-2	-	-	-
Nphi	scf_3.04_771.9	Dmrt	Dmrt-2	-	-	-
Obim	XP_052832484.1	Dmrt	Dmrt-2	-	-	-
Oedu	XP_029650766.2	Dmrt	Dmrt-2	-	-	-
Fcan	XP_05111540.1	Dmrt	Dmrt-2	-	-	-
Fcor	DN16532_cg.i1.i1.p1	Dmrt	Dmrt-2	-	-	-
Fmar	DN280523_cg.i1.i1.p1	Dmrt	Dmrt-2	-	-	-
Fmax	XP_033738864.1	Dmrt	Dmrt-2	-	-	-
Fstr	KAK3603675.1	Dmrt	Dmrt-2	-	-	-
Fvir	s0009/g235	Dmrt	Dmrt-2	-	-	-
Fyes	XP_021368788.1	Dmrt	Dmrt-2	-	-	-
Rphi	XP_060589226.1	Dmrt	Dmrt-2	-	-	-
Sbro	EVM001645.1	Dmrt	Dmrt-2	-	-	-
Scor	Ch8_1365	Dmrt	Dmrt-2	-	-	-
Tgra	KA18306274.1	Dmrt	Dmrt-2	-	-	-
Tsqu	DNS1730_cg.i1.i1.p1	Dmrt	Dmrt-2	-	-	-
Acal	XP_05096932.1	Dmrt	Dmrt-3	-	-	-
Aicr	Contig349_40	Dmrt	Dmrt-3	-	-	-
Amar	Amal264	Dmrt	Dmrt-3	-	-	-
Apur	scaffold_95_76	Dmrt	Dmrt-3	-	-	-
Bgia	XP_013077145.1	Dmrt	Dmrt-3	-	-	-
Cang	XP_0526687934.1	Dmrt	Dmrt-3	-	-	-

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Nmar	MMA00000008515	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmar	MMA00000016566	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmar	MMA00000047004	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmer	XP_045157593.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmer	XP_045157053.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Nmer	XP_048159713.2	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Nmer	DN116328.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Nmer	834603.t1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmer	g58531.t1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Nmer	g241174.t1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nmer	g192820.t1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Nphi	scf68796.0.3	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Oedeu	XP_025110328.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fcan	DN12587.c0.g3.i3.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fcor	DN12587.o.g2.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fcor	DN12587.c0.g1.i4.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fgen	DN68344.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fgen	DN35556.o.g1.i2.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fmax	XP_033737154.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Foku	DN35178.c0.g1.i2.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fstr	KAK3612677.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fstr	KAK3583105.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fstr	KAK3583112.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fstr	KAK3583110.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fstr	KAK3583109.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Fvir	s007362	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Fyes	XP_021377274.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Redec	DN26973.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Rphi	XP_060600638.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Rphi	XP_060600746.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Rphi	XP_060578674.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Sbro	EV/M0004335.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600)	
Sbro	EV/M0021940.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Scon	Chr8.1999	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Scon	Chr8.2143	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Siglo	Saf006992	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Sgra	DN54078.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Sgra	DN6659.o.g1.i3.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Tgra	KA18305799.1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Tsqu	DN75749.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Apec	DN93773.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Nchi	DN34711.c0.g1.i2.p1	Dmrt	Dmrt-4/5	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270600, 270602)	
Scon	Chr8.738	Dmrt	-			
Acal	XP_005007243.2	Fox-A	Fox-A			
Contig-157	Amo88751	Fox-A	Fox-A			
Amar	DN1107972.c0.g1.i1.p1	Fox-A	Fox-A			
Apec	scf00124.7	Fox-A	Fox-A			
Apur	XP_013061342.2	Fox-A	Fox-A			
Bgia	XP_052701295.1	Fox-A	Fox-A			
Cang	EV/M0004613.1	Fox-A	Fox-A			
Cari	DN1101169.c0.g1.i1.p1	Fox-A	Fox-A			
Cflu	XP_011413445.1	Fox-A	Fox-A			
Cgag	DN47094.c0.g1.i1.p1	Fox-A	Fox-A			
Cqli	Hic.asm.10.638	Fox-A	Fox-A			
Csin	XP_00510.437	Fox-A	Fox-A			
Cvir	XP_02233332.1	Fox-A	Fox-A			
Cvir	XP_022334050.1	Fox-A	Fox-A			
Dpol	XP_052272379.1	Fox-A	Fox-A			
Gaeg	XP_04135454.1	Fox-A	Fox-A			
Hibia	M00000018167	Fox-A	Fox-A			
Hruf	XP_046371021.1	Fox-A	Fox-A			
Mare	XP_052769228.1	Fox-A	Fox-A			
Mcal	XP_052106467.1	Fox-A	Fox-A			
Nchi	DN23553.c0.g1.i1.p1	Fox-A	Fox-A			

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Ncor	CAC5374046.1	Fox	Fox-A	Forkhead domain A1 (410812)	HNF3 C-terminal domain (430552; partial)	-
Neclu	CAG2201348.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872); HNF3 C-terminal domain (430552; partial)	-
Ngal	VDI17457.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872); HNF3 C-terminal domain (430552; partial)	-
Nmar	MMA0000008663	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872); HNF3 C-terminal domain (430552; partial)	-
Nmer	XP_045173733.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872); HNF3 C-terminal domain (430552)	-
Nmmod	DN103780_c0_g1.i1.p1_g19227.tl	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; partial)	-
Nmer	scat4682.0	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Obim	XP_014788201.2	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Oetu	XP_048735259.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pcan	DN3042_c0_g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pcor	DN104637_c0_g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; partial)	-
Pgen	DNS0866_c0_g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pfmr	XP_033734080.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pmax	KAK3597847.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pstr	s00068_g447	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Pvir	XP_021361791.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Rphi	XP_060590755.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; partial)	-
Sbro	EV/M0003194.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Scon	Chrd4_2670	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Sgio	Scg000464	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; HNF3 C-terminal domain (430552))	-
Sgra	DNF8052_c0_g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (349872; partial)	-
Acal	XP_05089018.1	Fox	Fox-B	Forkhead domain B2 (410812)	HNF3 C-terminal domain (430552)	-
Airc	Contig36_38	Fox	Fox-B	Forkhead domain B2 (410812)	-	-
Apur	Scatfold313.50	Fox	Fox-B	Forkhead domain B2 (410812)	-	-
Bglia	XP_013078204.1	Fox	Fox-B	Forkhead domain B2 (410812)	-	-
Cang	XP_052700333.1	Fox	Fox-B	Forkhead domain B2 (410812)	-	-
Cari	EV/M0003536.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Cflu	DN9813_c0_g1.i1.p1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Cegig	XP_011445364.2	Fox	Fox-B	Forkhead domain B (410817)	-	-
Cfin	Hic-asn_16_1347	Fox	Fox-B	Forkhead domain B (410817)	-	-
Cvrl	XP_022334612.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Dpol	DPol_052233250.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Dpol	XP_052256324.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Dpol	XP_052281977.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Gaeg	XP_04136159.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Hibia	M00000029836	Fox	Fox-B	Forkhead domain B (410817)	-	-
Hruf	XP_046358590.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Lorb	DN5589_c3_g1.i1.p1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nare	XP_052791461.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Ncal	XP_052100219.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Ncor	CAc5382565.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nmedu	CAG229716.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Ngal	VD16670.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nmar	MMA00000015629	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nmer	XP_045215505.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nmer	g250725.tl	Fox	Fox-B	Forkhead domain B (410817)	-	-
Nphi	scat10920_0.0	Fox	Fox-B	Forkhead domain B (410817)	-	-
Obim	XP_052832317.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Oedu	XP_048722871.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Osin	XP_029653697.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Pcan	XP_02078261.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Scon	Ch5_12	Fox	Fox-B	Forkhead domain B (410817)	-	-
Sglo	Sg1012012	Fox	Fox-B	Forkhead domain B (410817)	-	-
Tgra	KA18304921.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Pstr	s00189_g215	Fox	Fox-B	Forkhead domain B (410817)	-	-
Pvir	XP_021357620.1	Fox	Fox-B	Forkhead domain B (410817)	-	-
Rphi	XP_060564412.1	Fox	Fox-C	Forkhead domain L1 (410801)	-	-
Scon	Contig58_63	Fox	Fox-C	Forkhead domain L1 (410801)	-	-
Amar	Ama17094	Fox	Fox-C	Forkhead domain L1 (410801)	-	-
Apar	scafoid577_50	Fox	Fox-C	Forkhead domain L1 (410801)	-	-
Bglia	XP_055890240.1	Fox	Fox-C	Forkhead domain L1 (410801)	-	-

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Cang	XP_052715579.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Cari	EV/M0002771.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain C (410791)	
Cflu	DN96516_c0_g1.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Cgig	XP_011417585.2	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Cpli	DN157725_c0_g1.i1.p1.	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Csin	Hic-asn_17.1357	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Csin	Hic-asn_17.1443	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Cvir	XP_022346235.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Dpol	XP_052247387.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Gaeg	XP_041377087.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Hibia	M0000031058	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Hruf	XP_046372770.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlare	XP_052819073.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlcal	XP_052063314.1	Fox	Fox-C	Forkhead domain L1 (410791)	Forkhead domain L1 (410801)	
Nlchi	DN13809_c0_g1.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlcor	CAC5374004.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nledu	CAC2206344.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlgai	VD12482.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlmar	MMAM0000035616	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlmer	XP_045194706.2	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlmer	g82158.i1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Nlphi	scat69950.1.0	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Obim	XP_014786040.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Octeu	XP_048762038.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Osin	XP_029653806.2	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pcan	XP_0281115697.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pcor	DN14158_c0_g2.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pmax	DN14158_c0_g5.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pstr	XP_033755061.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pvir	s02023e12	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Pyes	XP_021346967.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Rphi	XP_066597004.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Sbro	EVM0022192.1	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Scon	Chr11.448	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Sgio	Sg009485	Fox	Fox-C	Forkhead domain L1 (410801)	Forkhead domain L1 (410801)	
Tgra	KA183035351.1	Fox	Fox-C	Forkhead domain C2 (410819)	Forkhead domain D4 (410822)	
Acal	XP_035824261.1	Fox	Fox-D	Forkhead domain L1 (410801)	Forkhead domain D4 (410822)	
Aaic	Contig1003.15	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D3 (410821)	
Amar	Amal186	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Apec	DN87882_c0_g1.i1.p1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Apur	scaffold13962.11	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Bglia	XP_013096936.2	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cang	XP_052668370.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cari	EVM0005770.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cgig	XP_011446328.2	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cpli	DN23774_c0_g1.i1.p1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Csin	Hic-asn_11.1425	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cvir	XP_022316146.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Dpol	XP_052256035.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Dpol	XP_052256591.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Gaeg	XP_041356731.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Hibia	M0000030583	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Hruf	XP_046329290.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Lorb	DN24803_c0_g1.i1.p1.	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlare	XP_052777467.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlcal	XP_053095292.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlchi	XP_052075808.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlcor	CAC5382691.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nledu	CAC5407497.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlgai	CAC2204666.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlmar	CAG2248150.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlmer	CA62203362.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlmer	VD10735.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D3 (410821)	
Nlmer	VDH93066.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlmer	MMAM0000002467	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nlmer	XP_045157253.2	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Mnher	g192986.t1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Nphi	scf_69489_1.2	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Oblim	scf_42856_0.4	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Cediu	XP_052826256.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Pcan	XP_025110533.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Pcor	DN15187_c0.g1.i5.p1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Fmar	DN51265_c0.g1.i1.p1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Fmax	XP_033737945.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Pstr	KAK352139.1	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Fvir	s01405.g91	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain D4 (410822)	
Fvir	s00997.g340	Fox	Fox-D	Forkhead domain D4 (410822)	Forkhead domain E (410793)	
Fyes	XP_021345225.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Rphi	XP_060582828.1	Fox	Fox-D	Forkhead domain E (410793)	Forkhead domain E (410793)	
Sbro	EVM0002351.1	Fox	Fox-D	Forkhead domain E (410793)	Forkhead domain E (410793)	
Scon	Chr8.2069	Fox	Fox-D	Forkhead domain E (410793)	Forkhead domain E (410793)	
Sgio	Sg013024	Fox	Fox-D	Forkhead domain E (410793)	Forkhead domain E (410793)	
Tgra	KA18306624.1	Fox	Fox-D	Forkhead domain E (410793)	Forkhead domain E (410793)	
Airc	Contig89.19	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Amar	Ama12250	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Apur	scaffold_253_23	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cang	XP_052668878.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cari	EVM0003839.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cflu	DN108936_c2.g1.i1.p1.	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cgig	XP_011444776.2	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cvrr	XP_022319236.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Dpol	XP_052286560.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Hiba	XP_0000038943	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Hruf	XP_046353578.2	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nare	XP_052778423.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Ncal	XP_052075782.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Ncor	CAE5384260.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nedu	CAG2217852.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nedu	CAG2194171.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Ngai	CAG2199036.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nmer	MMA00000033504	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nmod	DN117568_c0.g1.i1.p1.	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nmer	g241620.t1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Nphi	scf_31587_0.4	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Cediu	XP_048762291.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Fmar	DN7017_c0.g1.i1.p1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Fmax	XP_033737819.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain E (410793)	
Fstr	KAK353103.1	Fox	Fox-E	Forkhead domain E (410793)	Forkhead domain F1 (410823)	
Fvir	s00149854	Fox	Fox-F	Forkhead domain F1 (410794)	Forkhead domain F (410794)	
Fyes	XP_021378858.1	Fox	Fox-F	Forkhead domain F1 (410823)	Forkhead domain F (410794)	
Rdec	DN24595_c4.g1.i1.p1	Fox	Fox-F	Forkhead domain F1 (410794)	Forkhead domain F (410794)	
Rphi	XP_060578637.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Sbro	EVM0010028.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Sgio	Sg009305	Fox	Fox-F	Forkhead domain F1 (410823)	Forkhead domain F1 (410823)	
Acal	XP_058802380.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Airc	Contig1133.18	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Amar	Am339500	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Amar	Am32615	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Apec	DN75342_c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Apur	scaffold_860_37	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Bglia	XP_05880969.2	Fox	Fox-F	Forkhead domain F1 (410823)	Forkhead domain F (410794)	
Cang	Contig11190.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Cari	XP_011445317.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Cgig	DN7628_c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Cpli	Hic.asm.17.158	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Csin	Cvrr	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Dpol	XP_022335664.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Gaeg	XP_041375666.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Hiba	M00000007664	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F (410794)	
Hruf	XP_046372691.1	Fox	Fox-F	Forkhead domain F (410794)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Nlare	XP_052815084.1	Fox	Fox-F	Forkhead domain F1 (410823)	Forkhead domain F	
Nlcal	XP_050060477.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlcior	CAE5387332.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nledu	CAG2252875.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlgal	VDI121852.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlmar	MMA000000030848	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlmer	XP_045194642.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlmod	DN104261.c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlmer	g1061129.t1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Nlphi	scaf.40546.0.2	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Obim	XP_014777539.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Oledu	XP_048732202.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pcan	XP_025116010.1	Fox	Fox-F	Forkhead domain F1 (410823)	Forkhead domain F	
Pcor	DN180003.c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pfmar	DN129940.c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pmax	Chr11.927	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pstr	XP_033755005.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pvir	KAK3601654.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Pyes	sl133835g10	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Rphi	XP_021358008.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Sbro	XP_060601663.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Scon	EVN0015186.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Scor	Chr11.810	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Sgio	Sg002627	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Tgra	KA18302829.1	Fox	Fox-F	Forkhead domain F (410794)	Forkhead domain F	
Tsqu	DN137576.c0.g1.i1.p1	Fox	Fox-F	Forkhead domain G (410795)	Forkhead domain G	
Acal	XP_005099252.2	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Acal	XP_005099253.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Airc	Contig625.38	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Amar	Amal0381	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Apec	DN10836.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Apur	sccaffold.36470.28	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Bejia	XP_055879295.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cang	XP_052699015.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cari	EVN0001891.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cflu	DN104980.c0.g1.i2.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cgig	XP_011427689.2	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cipli	DNS8419.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Cisin	Hic-asn.10.1034	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Civr	XP_022334541.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Dpol	XP_052270224.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Dpol	XP_052270147.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Graeg	XP_041354930.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Hbia	M00000035830	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Hlrf	XP_046371537.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlcal	DN60568.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlmer	g133265.t1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlphi	scaf.15017.0.3	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Medu	CAE219343.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlgal	VDI2297.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlmar	MMA000000030730	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlmer	XP_045162348.2	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Nlmod	DN60568.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Pcor	XP_025105724.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Poku	DN81635.c0.g2.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Fgen	DN112984.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Fmar	DN28516.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	
Pmax	XP_033734631.1	Fox	Fox-G	Forkhead domain G (410795)	Forkhead domain G	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Pstr	KAK3604690.1	Fox	Fox-G	Forkhead domain G (410795)		
Fvir	s00383:835	Fox	Fox-G	Forkhead domain G (410795)		
Fyes	XP_021363790.1	Fox	Fox-G	Forkhead domain G (410795)		
Rphi	XP_060589805.1	Fox	Fox-G	Forkhead domain G (410795)		
Sbro	EV/M0011335.1	Fox	Fox-H	Forkhead domain H (410796)		
Sbro	EV/M0012606.1	Fox	Fox-G	Forkhead domain G (410795)		
Scon	Chr3:2805	Fox	Fox-G	Forkhead domain G (410795)		
Sgio	Sg014601	Fox	Fox-G	Forkhead domain G (410795)		
Sgra	DNA9488_c0_g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)		
Arc	Contig18244.2	Fox	Fox-H	Forkhead domain H (410796)		
Arc	Contig178.106	Fox	Fox-H	Forkhead domain H (410796)		
Amar	Ana16564	Fox	Fox-H	Forkhead domain H (410796)		
Amar	Amp-05266	Fox	Fox-H	Forkhead domain H (410796)		
Cang	XP_052684756.1	Fox	Fox-H	Forkhead domain H (410796)		
Cari	EV/M0021377.1	Fox	Fox-H	Forkhead domain H (410796)		
Cflu	DN1:38466_c0_g1.i1.p1	Fox	Fox-H	Forkhead domain H (410796)		
Cgig	XP_034313225.1	Fox	Fox-H	Forkhead domain H (410796)		
Hic-as1	Chr.asm.14.811	Fox	Fox-H	Forkhead domain H (410796)		
Csin	XP_022314590.1	Fox	Fox-H	Forkhead domain H (410796)		
Cvir	M00000018729	Fox	Fox-H	Forkhead domain H (410796)		
Hibia	XP_048254913.1	Fox	Fox-H	Forkhead domain H (410796)		
Htuf	XP_048255113.1	Fox	Fox-H	Forkhead domain H (410796)		
Ncor	CAC5408624.1	Fox	Fox-H	Forkhead domain H (410796)		
Ncor	CAC5397897.1	Fox	Fox-H	Forkhead domain H (410796)		
Ncor	CAC5397896.1	Fox	Fox-H	Forkhead domain H (410796)		
Ncor	CAC5403969.1	Fox	Fox-H	Forkhead domain H (410796)		
Nedu	CAG2228903.1	Fox	Fox-H	Forkhead domain H (410796)		
Nedu	CAG2188004.1	Fox	Fox-H	Forkhead domain H (410796)		
Nedu	CAG2292853.1	Fox	Fox-H	Forkhead domain H (410796)		
Nedu	CAG2202596.1	Fox	Fox-H	Forkhead domain H (410796)		
Ngal	VD162725.1	Fox	Fox-H	Forkhead domain H (410796)		
Ngal	VDH43947.1	Fox	Fox-H	Forkhead domain H (410796)		
Ngal	VDI20844.1	Fox	Fox-H	Forkhead domain H (410796)		
Nlmar	MMAM00000022684	Fox	Fox-H	Forkhead domain H (410796)		
Nlmer	XP_053378216.1	Fox	Fox-H	Forkhead domain H (410796)		
Nlmer	XP_045194303.2	Fox	Fox-H	Forkhead domain H (410796)		
Nlmer	XP_045198985.2	Fox	Fox-H	Forkhead domain H (410796)		
Nlmer	R213542.tl	Fox	Fox-H	Forkhead domain H (410796)		
Nphi	scaf.17325.0.4	Fox	Fox-H	Forkhead domain H (410796)		
Nphi	scaf.28666.1.1	Fox	Fox-H	Forkhead domain H (410796)		
Oedu	XP_048759429.2	Fox	Fox-H	Forkhead domain H (410796)		
Fcan	XP_025075954.1	Fox	Fox-H	Forkhead domain H (410796)		
Fcor	DN1:16957_c0_g1.i1.p1	Fox	Fox-H	Forkhead domain H (410796)		
Fmax	XP_033755807.1	Fox	Fox-H	Forkhead domain H (410796)		
Pstr	KAK360389.1	Fox	Fox-H	Forkhead domain H (410796)		
Fvir	s75598g33	Fox	Fox-H	Forkhead domain H (410796)		
Rphi	s00238_g1.31	Fox	Fox-H	Forkhead domain H (410796)		
Rphi	XP_060558970.1	Fox	Fox-H	Forkhead domain H (410796)		
Rphi	XP_060567331.1	Fox	Fox-H	Forkhead domain H (410796)		
Sbro	XP_060604067.1	Fox	Fox-H	Forkhead domain H (410796)		
Sbro	EV/M0016618.1	Fox	Fox-H	Forkhead domain H (410796)		
Scon	Chr11:1359	Fox	Fox-H	Forkhead domain H (410796)		
Sgio	Sg013003	Fox	Fox-H	Forkhead domain H (410796)		
Acal	XP_005108651.1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Arc	Contig775.5	Fox	Fox-J1	Forkhead domain J1 (410797)		
Amar	Ama0222	Fox	Fox-J1	Forkhead domain J1 (410797)		
Apec	DN20109_c0_g1.i6.p1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Apur	scaffold.797.10	Fox	Fox-J1	Forkhead domain J1 (410797)		
Bglia	XP_013064514.2	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cang	XP_052689275.1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cari	EV/M0003558.1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cflu	DN1:27407_c0_g2.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cgig	XP_011445234.1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cpli	DN55792_c0_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)		
Hic-as1	0.1540	Fox	Fox-J1	Forkhead domain J1 (410797)		
Cvir	XP_022319181.1	Fox	Fox-J1	Forkhead domain J1 (410797)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Cvir	XP_022319268.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Dpol	XP_052265485.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Gaeq	XP_041362703.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Hbiba	M00000003225	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Htruf	XP_046330175.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Lorb	DN146217_c0_g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_0527164656.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_052816834.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Ncor	CAC5405074.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Medu	XP_052764667.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_052775202.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_052775217.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_052775230.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Ncal	XP_05068038.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nichi	DN13639_c0_g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nher	g198765.t1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nphi	scar_33310.2.9	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Obim	XP_052824622.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Oetu	XP_048763213.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Csin	XP_029638410.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fcan	XP_028095423.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fcor	DN184891_c0_g1.i2.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fcor	DN18480_c0_g2.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fgen	DN18481_c0_g2.i2.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fgen	DN1399_c8.g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fmar	DN12837_c1.g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Fmax	XP_033752519.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Poku	DN19777_c2.g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Ppur	DN2521_c0_g1.i4.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Pstr	KAK3579229.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Pvir	s01693.g10	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Pyes	XP_021351058.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Rdec	DN22834_c9.g7.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Rphi	XP_060587750.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Sbro	EVM0018668.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Sccon	Chrl_3201	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Sccon	Chrl_3198	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Sgio	Sgi000050	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Sgra	DN10939_c0.g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Tgra	KA18318321.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Tsqu	DN0625_c2.g1.i2.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Acal	XP_05611247.3	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Arc	Contig53.201	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Amar	Am34942	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Apur	scaffold_id.11801.24	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Bglia	XP_03070049.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cang	XP_052716138.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cari	EV/M008910.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cflu	DNS8808_c0_g1.i1.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cgig	XP_011422959.2	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cpli	DN17831_c5.g1.i2.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Csin	Hic_asm_4.381	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Cvir	XP_022341777.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Dpol	XP_052285372.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Gaeq	XP_041378546.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Hbiba	M00000010754	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Htruf	XP_048247606.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nare	XP_052759824.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Ncal	XP_052759802.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Nchi	DN13497_c0.g1.i2.p1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Ncor	CAC5378041.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		
Medu	CAG2221519.1	Fox	Fox-JJ	Forkhead domain J1 (410797)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Ngal	VDI57447.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Ngal	VDI57448.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Nmar	MMA00000017129	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Nmer	XP_053379821.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Nmer	XP_053379966.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Nmmod	DN38610_c0_g1_i1.p1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Nphi	scar40576.04	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Obim	XP_052832979.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Oeclu	XP_048739234.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Csin	XP_029651657.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Fcan	XP_025081218.1	Fox	Fox	Fox-12/3	Forkhead domain J3 (410826)	
Fcor	DN2942_c0_g1_i5.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fcor	DN2942_c0_g1_i3.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fcor	DN2942_c0_g1_i7.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fgen	DN5381_c0_g1_i3.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fmar	DN41364_c0_g1_i3.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fmax	XP_033763328.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Foku	DN14771_c0_g2_i3.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fpur	DN2181_c0_g1_i5.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fstr	KAK3533417.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fvir	s0001986	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Fyes	XP_021374633.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Rphi	XP_060577634.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Sbro	EV/M0013081.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Scon	Chrl2.800	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Sgio	Sg003279	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Sgra	DN1042_c0_g1_i2.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Tgra	KA1832020.1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Tsqu	DN10376_c0_g1_i2.p1	Fox	Fox	Fox-12/3	Forkhead domain FOX12_FOX13 (410798)	
Acal	XP_005092494.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Apur	scaffold_855_43	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Bglg	XP_0130902951	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Cang	XP_052688140.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Cari	EV/M000967.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Cegig	XP_011416099.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Cpli	DN64350_c0_g1_i1.p1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Cvrl	XP_022316096.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Gaeg	XP_041362451.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Hbia	M0000008333	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Hruf	XP_048248693.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Nmar	MMA00000012630	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828; partial)	
scatf14580.0.11	Fox	Fox-K	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828; partial)	
Nphi	OBim	XP_014772374.1	Fox	Fox-K	Forkhead associated (FHA) domain (410828)	
Oedu	XP_048761057.2	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Osin	XP_029646877.1	Fox	Fox-K	Fox-K	Forkhead associated (FHA) domain (410828)	
Rdec	Sg00589	Fox	Fox-K	Fox-L1	Forkhead associated (FHA) domain (410828; partial)	
Sgio	XP_01294028.1	Fox	Fox-L1	Fox-L1	Forkhead associated (FHA) domain (410828)	
Acal	Contig58.64	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Airc	Amal7914	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Apec	DN74037_c0_g1_i1.p1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Apur	scaffold122.2	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Bglg	XP_055890278.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Cang	XP_052718866.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Cari	EV/M0019009.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Cegig	XP_041375667.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Hbia	M0000031057	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Hruf	XP_046343957.2	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Csin	Hicam17.1225	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Cvrl	DPol	XP_022346240.1	Fox	Fox-L1	Forkhead domain L1 (410801)	
Gaeg	XP_041375667.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Hbia	DN157469_c0_g1_i1.p1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Hruf	DN104458_c0_g1_i1.p1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Lorb	XP_052817971.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Mare	XP_052817977.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Mcal	XP_052063315.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	
Mcor	CAC5374405.1	Fox	Fox-L1	Fox-L1	Forkhead domain L1 (410801)	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Nmedu	CAG2206845_1	Fcor	Fox-LI	Forkhead domain L1 (410801)		
Mgal	VDI22484_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Ngal	VDH97507_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Nmar	MMA00000028776	Fox	Fox-LI	Forkhead domain L1 (410801)		
Nmer	XP_053402988_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Nmod	DN51324_c0.g1.i1.p1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Nmer	g268924.tl	Fox	Fox-LI	Forkhead domain L1 (410801)		
Nphi	scaf69950_0.0	Fox	Fox-LI	Forkhead domain L1 (410801)		
Obib	XP_014785001_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Oedu	XP_048762096_2	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fcan	XP_028076243_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fcor	DN28326_c0.g1.i1.p1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fcor	DN187497_c0.g1.i1.p1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fmar	DN50135_c0.g1.i1.p1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fmax	XP_03375354_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fpur	DN73831_c0.g1.i1.p1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fstr	KAK3509991_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fvir	s02023@1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Fyes	XP_021346965_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Rphi	XP_060608039_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Sbro	EV/M0016190_1	Fox	Fox-LI	Forkhead domain L1 (410801)		
Scon	Chr1.1.868	Fox	Fox-LI	Forkhead domain L1 (410801)		
Sgio	Sg009486	Fox	Fox-LI	Forkhead domain L1 (410801)		
Acal	XP_005101910_2	Fox	Fox-L2	Forkhead domain L2 (410802)		
Airc	Contig51_34	Fox	Fox-L2	Forkhead domain L2 (410802)		
Amar	Ama34673	Fox	Fox-L2	Forkhead domain L2 (410802)		
Apur	scaffold84_159	Fox	Fox-L2	Forkhead domain L2 (410802)		
Bglg	XP_0558865110_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cang	XP_0527178506_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cari	EVM0021728_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cflu	DN127322_c6.g2.i2.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cgig	NP_01295827_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cpli	DN7506_c5.g1.i2.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Csin	Hc_asm.4.274	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cvir	XP_022345405_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Cvir	XP_022345173_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Dpol	XP_052212727_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Gaesg	XP_041378252_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Hbia	M00000035173	Fox	Fox-L2	Forkhead domain L2 (410802)		
Hruf	XP_048250285_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Lorb	DN129129_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nare	XP_052760962_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Ncal	XP_050862415_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Ncor	CAc5401149_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nmedu	CAG2239672_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Mgal	VDI438865_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Ngal	VDI44864_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nmar	MMA0000016212	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nmer	XP_045161614_2	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nmod	DN2410_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nmer	g3235_11	Fox	Fox-L2	Forkhead domain L2 (410802)		
Nphi	scaf510301_0.3	Fox	Fox-L2	Forkhead domain L2 (410802)		
Obib	XP_014785648_2	Fox	Fox-L2	Forkhead domain L2 (410802)		
Oedu	XP_048729555_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fcan	XP_025083514_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fcor	DN35937_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fgen	DN134171_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fmar	DN32846_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fmax	XP_033724433_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fstr	KAK3602726_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fvir	s00248e193	Fox	Fox-L2	Forkhead domain L2 (410802)		
Fyes	XP_021353421_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Rdec	DN21003_c0.g1.i2.p1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Rphi	XP_060566301_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Sbro	EV/M0017513_1	Fox	Fox-L2	Forkhead domain L2 (410802)		
Scon	Chr12.1684	Fox	Fox-L2	Forkhead domain L2 (410802)		
Sgio	Sg0033363	Fox	Fox-L2	Forkhead domain L2 (410802)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Tsqu	DN37_c29_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410804)		
Acal	XP_005091040.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Airc	Contig281.47	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Amar	Am23426	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Apec	DN7027_c0.g1.i2.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Apur	scaffold17_i163	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Bglia	XP_055899100.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Cang	XP_052711314.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Carie	EV/M0024311.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Cgig	XP_034303195.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Cpli	DN78931_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Csin	Hic.asm.2.101	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Cvir	XP_022927871	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Dpol	XP_052270473.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Gaeg	XP_041365083.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Hbia	M000002027642	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Hruf	XP_048241610.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Mare	XP_052801997.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ncal	XP_052066630.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Nchi	DN25972_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ncor	CAC5383890.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Nedu	CAG2287106.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ngal	VDH33464.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ngal	VDH33462.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ngal	VDH33463.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Nmar	MMA00000018109	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Nmer	XP_045177340.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Mmod	DN2607_c0.g2.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Nphi	scaf69935_0.10	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Obim	XP_052825413.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Oediu	XP_055996035.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Osin	XP_029638459.2	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pcan	XP_025080495.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pcor	DN5558_c0.g2.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pgen	DN145626_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pjgen	DN3078_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pmar	XP_033751425.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Poku	DN42531_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Sbro	DN195653_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Ppur	KAK3587366.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pstr	s24333645	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pvir	XP_021371548.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Pyes	DN20122_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Rdec	XP_060606622.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Rphi	EV/M009578.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Scon	Chr14_2061	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Sgio	Sgo04456	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Sgra	DN11938_c0.g2.i3.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Tgra	KA18298705.1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Tsqu	DN22139_c0.g1.i1.p1	Fox	Fox-NL/4	Forkhead domain N1 (410804)		
Acal	XP_052099217.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Airc	Contig17_i153	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Amar	Am09979	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Apec	DN11918_c0.g1.i10.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Apur	scaffold489_22	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Bglia	XP_013084232.2	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Cang	XP_053668143.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Carie	EV/M016469.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Cflu	DN125734_c1.g1.i8.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Cgig	XP_034324225.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Cpli	DN79231_c0.g1.i17.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Csin	Hic.asm.10.136	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Cvir	XP_022331167.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Dpol	XP_052270965.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Gaeg	XP_041363111.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Hbia	M0000030949	Fox	Fox-N2/3	Forkhead domain N3 (410833)		
Hruf	XP_046351344.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Mare	XP_052767715.1	Fox		Fox-N2/3		
Ncal	XP_052107190.1	Fox		Fox-N2/3		
Nichi	DN^39446_c1.g1.i1.p1	Fox		Fox-N2/3		
CAC5378437.1				Fox-N2/3		
Ncor	CAG2235611.1	Fox		Fox-N2/3		
Nedu	VDI802961	Fox		Fox-N2/3		
Ngal	VDI802891	Fox		Fox-N2/3		
Ngal	VDI8287.1	Fox		Fox-N2/3		
Nmar	MMA00000016688	Fox		Fox-N2/3		
Nmer	XP_053376884.1	Fox		Fox-N2/3		
Nmod	DN^2418_c0.g1.i31.p1	Fox		Fox-N2/3		
Nmer	g156153.i2	Fox		Fox-N2/3		
Nphi	scf_37509_0_5	Fox		Fox-N2/3		
Obim	XP_052822674.1	Fox		Fox-N2/3		
Oedu	XP_048735022.1	Fox		Fox-N2/3		
Osin	XP_029633348.1	Fox		Fox-N2/3		
Pcan	XP_025106088.1	Fox		Fox-N2/3		
Pcor	DN19540_c0.g1.i1.p1	Fox		Fox-N2/3		
Pcor	DN5191_c0.g2.i7.p1	Fox		Fox-N2/3		
Pcor	DN18451_c0.g2.i3.p1	Fox		Fox-N2/3		
Pgen	DN14328_c0.g1.i3.p1	Fox		Fox-N2/3		
Pmar	DN^2157_c1.g1.i3.p1	Fox		Fox-N2/3		
Pmax	DN17347_c4.g1.i2.p1	Fox		Fox-N2/3		
Poku	DN17429_c4.g1.i2.p1	Fox		Fox-N2/3		
Ppur	DN5075_c0.g1.i1.p1	Fox		Fox-N2/3		
Pstr	KAK359595.3.1	Fox		Fox-N2/3		
Pvir	s00410g95	Fox		Fox-N2/3		
Pyes	XP_021366964.1	Fox		Fox-N2/3		
Rdec	DN^2296_c2.g1.i1.p1	Fox		Fox-N2/3		
Rphi	XP_060552999.1	Fox		Fox-N2/3		
Sgio	Se013452	Fox		Fox-N2/3		
Sgra	DN13133_c0.g2.i8.p1	Fox		Fox-N2/3		
Tgra	KA1830864.1	Fox		Fox-N2/3		
Tsau	DN^75347_c0.g1.i2.p1	Fox		Fox-N2/3		
Acal	XP_05112460.1	Fox		Fox-O		
Airc	Contig16.24	Fox		Fox-O		
Amar	Ana07814	Fox		Fox-O		
Apec	DN23636_c0.g1.i1.p1	Fox		Fox-O		
Apur	scaffold_712.6	Fox		Fox-O		
Egia	XP_013095255.2	Fox		Fox-O		
Cang	XP_052708272.1	Fox		Fox-O		
Cari	EV/M000968.1	Fox		Fox-O		
Cflu	DN112955_c5.g3.i5.p1	Fox		Fox-O		
Cgag	XP_011414359.1	Fox		Fox-O		
Cgpli	DN72415_c5.g1.i1.p1	Fox		Fox-O		
Cvir	XP_022287692.1	Fox		Fox-O		
Cvir	XP_022228423.1	Fox		Fox-O		
Dpol	DN^29098_c0.g1.i3.p1	Fox		Fox-O		
Gaeq	XP_041351714.1	Fox		Fox-O		
Hbia	M0009024236	Fox		Fox-O		
Hruif	XP_046374293.1	Fox		Fox-O		
Lorb	DN142512_c0.g1.i1.p1	Fox		Fox-O		
Nmer	DN^45195791.2	Fox		Fox-O		
Ncal	XP_052808484.1	Fox		Fox-O		
Nichi	g187283.i1	Fox		Fox-O		
Obim	XP_052822486.1	Fox		Fox-O		
Oedu	XP_04876690.1	Fox		Fox-O		
Osin	XP_029637040.1	Fox		Fox-O		
Pcan	DN^2008_c0.g1.i2.p1	Fox		Fox-O		
Pcor	DN^7549_c0.g3.i2.p1	Fox		Fox-O		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Fgen	DN24871_c0.g1.i3.p1	Fox	Fox-O	Forkhead domain O (410806)		
Fgen	DN24871_c0.g1.i3.p2	Fox	Fox-O	Forkhead domain O (410806)		
Fmar	DN44399_c0.g5.i1.p1	Fox	Fox-O	Forkhead domain O (410806)		
Fmax	XP_033740844.1	Fox	Fox-O	Forkhead domain O (410806)		
Foku	DN10962_c1.g2.i2.p1	Fox	Fox-O	Forkhead domain O (410806)		
Fpur	DN72510_c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)		
Fstr	KAK357695.1	Fox	Fox-O	Forkhead domain O (410806)		
Fvir	s00079e102	Fox	Fox-O	Forkhead domain O (410806)		
Fyes	XP_021377366.1	Fox	Fox-O	Forkhead domain O (410806)		
Rdec	DN30721_c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)		
Rphi	XP_605389384.1	Fox	Fox-O	Forkhead domain O (410806)		
Scon	Chr2_491	Fox	Fox-O	Forkhead domain O (410806)		
Sgio	Sgi018351	Fox	Fox-O	Forkhead domain O1 (410834)		
Sgra	DN3576_c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)		
Tsqu	DN138852_c0.g1.i1.p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Airc	Contig3461.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Airc	Contig330_72	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Apur	Sccaffold576_108	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Cang	EV/M0015778.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Cari	XP_011412452.2	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Ccgig	XP_022300144.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Cvrl	XP_052234997.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Dpol	XP_052237166.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Gaeleg	XP_041362088.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Hibia	M0000010651	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Hruf	XP_048236781.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Hruf	XP_0463288651.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Ncal	XP_052089402.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Ncor	CAC5388114.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Medu	CAG2290347.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Mgal	VDI055631	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Mgal	VDI05564.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Nlmar	MMAM0000027087	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Nmer	XP_0451829632.3	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Nmmod	DN9753_c0.g1.i1.p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Nphi	scat6119_21	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Oedu	XP_048731527.2	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Pmar	DN32892_c0.g1.i1.p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Pmax	XP_033727511.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Pstr	KAk3609024.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Pvir	s01298e51	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Pyes	XP_021354438.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Rdec	DN23702_c0.g1.i2.p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Rphi	XP_060566633.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Sbro	EV/M009544.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Scon	Chr1_409	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Sgio	Se004484	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Tgra	KA18322379.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)		
Tsqu	DN207442_c0.g1.i1.p1	Fox	Fox-OG13/NA	Forkhead domain Q (410809)		
Acal	XP_051069163	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Airc	Contig85_21	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Amar	Am19770	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Apur	Sccaffold360_14	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Bglia	XP_013071662.2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Cang	XP_052676257.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Cari	EV/M001823.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Ccgig	XP_011439389.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Cpli	DN35749_c0.g1.i1.p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Csin	Hic_asm.12_159	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Cvrl	XP_022296913.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Dpol	XP_052253290.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Gaeleg	XP_05225240.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Hibia	XP_041347345.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Hruf	M0000015843	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Mare	XP_046382017.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		
Mcal	XP_052062481.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Nlcor	CAC5419615.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Nlcor	CAG2299155.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Nlgal	VD150805.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Nlmar	MMA00000012410	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Nlmer	XP_0451066371.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Nlphi	scrf67833.0.2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Obim	XP_04772941.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Oedu	XP_048742700.2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pcan	XP_025096321.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pcan	DN7667..0..g..i1..p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pcor	DN54039..c0..g..i1..p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pcor	DN7667..c0..g..i1..p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pmax	XP_033744896.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pstr	KAR589497.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Pyes	XP_021371037.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Rphi	XP_060586724.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Sbro	EV/M0000506.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Sbro	EV/M0006433.1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Scn	Chr..16.24	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Sgio	Sgl024307	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b	
Airc	Contig4307.7	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	-	
Apur	scrf604..173	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	-	
Cang	XP_052676026.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cari	EV/M0023364.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cflu	DN107758..c5..g1..i2..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cgig	XP_019926571.2	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cphi	DNF669..c0..g1..i2..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cvir	XP_022321288.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Cvir	XP_0222295893.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Dpol	XP_052226263.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Gaeg	XP_041365712.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Hba	M0000030826	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	-	
Hruf	XP_04636140..2	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	-	
Lorb	DN243786..c0..g1..i1..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlare	XP_0523802309.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlcor	CAC5370465.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlcor	CAG2202185.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nledu	CAG2246856.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlgal	VD124665.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlmar	MMA00000032793	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlmer	XP_045177123.2	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Nlphi	scrf70200..0..4	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Obim	XP_014771053.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Oedu	XP_048728661.2	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Osin	XP_036362897.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Pmax	DNA3205..c4..g1..i1..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Ppur	XP_033752233.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Pstr	DN4666..c0..g1..i2..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Pvir	KAK381527.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Pvir	s00194..451	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Pyes	DN2384..c1..g1..i1..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Rdec	DN2502..c0..g4..i1..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Rphi	XP_060599562.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Sbro	EV/M002125.1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Scn	Chr..4..16.28	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Tsgu	DN2384..c1..g1..i1..p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	-	
Airc	Contig438..5	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Apec	DN2340..c1..g1..i1..p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Apur	scrf133..74	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Cang	XP_052712245.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Cari	EV/M0012682.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Cflu	DN113069..c3..g1..i4..p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Cphi	DN76173..c0..g1..i2..p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Csin	Hic..asm..2..1802	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Csin	Hic..asm..2..1455	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Cvir	XP_022286391.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	
Dpol	XP_052282333.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Dpol	XP_052234537.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Gaeg	XP_041366058.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Hbba	M0000014061	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Hruf	XP_046335487.2	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nfare	XP_052798013.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Ncal	XP_05064572.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Ncor	CAC53715062.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nedu	CAG224977.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nmar	MMA00000037791	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nmer	XP_053384136.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nmod	DN27089_c0.g1.i1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Nmer	SL44243.t1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Ceclu	XP_048731286.2	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pcor	DN8679_c0.g1.i1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pcor	DN13056_c0.i1.l1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pcor	DN13056_0.g2.i2.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pgen	DN28413_c1.i1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pmar	DN44947_c1.i2.i3.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pmax	XP_033751305.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Poku	DN37223_c1.g1.i8.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Ppur	DN5451_c0.g1.i1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pvir	s00219e11	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Pyes	XP_021377259.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Rdec	DN22152_c4.g4.i1.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Rphi	XP_060569990.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Sbro	EV/M0023670.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Sgra	Sgi0005561	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Fstr	KAK3597624.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Tgira	KA1829935.1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Tsqu	DN5434_c0.g1.i7.p1	Fox	Fox	Fox-OG2/NA	Forkhead domain M (410803)	Annotated as Fox-M
Arc	Contig65_41	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Am25953	Fox	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Amar	scaffold_867_41	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Apur	scaffold_381_16	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Cang	EV/M0004465.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Cari	XP_052700156.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ccgig	XP_011435457.2	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Csin	Hic-asn_16.939	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Cvrr	XP_022334283.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Dpol	XP_052278575.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Dpol	XP_052278576.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Dpol	XP_052278604.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Hruf	XP_046341176.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nfare	XP_052791887.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nfare	XP_052791890.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nfare	XP_052791888.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nfare	XP_052791891.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncal	XP_052098761.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncal	XP_052098555.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5419382.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5380823.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5379920.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5419389.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5419386.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ncor	CAC5419388.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2214460.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2214467.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2198066.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2198058.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2198055.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nedu	CAG2214461.1	Fox	Fox	Fox-OG2/NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Nmedu	CAG2198060_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198057_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198065_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198059_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198056_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAC2234548_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198063_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198061_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198064_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmedu	CAG2198062_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI02350_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI139859_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI15906_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI02348_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI15903_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI02347_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI15905_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Ngal	VDI15904_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nmer	XP_0532405097_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nphip	DN5982_20_g1.i3.p1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Nphip	scat.1544_0.1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Cediu	XP_0560201213_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Fmar	DN5963_c0.g1.l1.p1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Fmax	XP_033751006_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Pvir	XP_033749723_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Pys	s0343_648	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Rphi	XP_021360588_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Sbro	XP_060585377_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Scon	EVM0013029_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Sglo	Ch5.396_1	Fox	Fox-OG28	NA	Forkhead domain Q2 (410809)	Annotated as Fox-Q2c
Acal	Sg01_3625	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-Q2c
Airn	XP_005102249_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-Q2c
Airn	Contig63_41	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-Q2c
Amar	Am26012	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Apur	scaffold_313_52	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Beigia	XP_058874345_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Cang	EV/M0018541_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Cari	XP_011441298_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Cigig	XP_022334408_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Cvir	XP_022334077_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Dpol	XP_052278569_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Graeg	XP_041375984_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hibria	M0000015535	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046330707_2	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Hruf	XP_046380710_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Hruf	XP_046350686_2	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046330709_2	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Hruf	XP_046330688_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046350660_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046350687_2	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046380712_2	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Hruf	XP_046350708_2	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Nare	XP_052102496_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Ncal	CAC5414394_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Ncor	XP_046350687_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Nmedu	CAG2193762_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Ngal	VDI35942_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Nmar	MMA00000023830	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Nmer	XP_045215157_2	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Nphi	scat.7111_0.1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Cediu	XP_048737442_2	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Fcan	XP_025078030_1	Fox	Fox-OG28	NA	Forkhead domain E (410793)	Annotated as Fox-AB
Fmax	XP_033750900_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Fstr	KAK3601439_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB
Fstr	KAK3601419_1	Fox	Fox-OG28	NA	Forkhead domain L1 (410801)	Annotated as Fox-AB

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Pvir	s001189e393	Fox	Fox-OC39/NA	Forkhead domain E (410793)	Annotated as Fox-AB	
Pys	XP_021357612.1	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Rphi	XP_060589031.1	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Sbro	EVM0003782.1	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Scn	Chr5.18	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Sgio	Sg004401	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Tgra	KAJ8304916.1	Fox	Fox-OC39/NA	Forkhead domain L1 (410801)	Annotated as Fox-AB	
Airc	Contig120.132	Fox	Fox-P	Forkhead domain P (410801)	Annotated as Fox-AB	
Apur	scf0ffid.507.3	Fox	Fox-P	Fox-P	-	
BglA	XP_055882746.1	Fox	Fox-P	Fox-P	-	
Cang	XP_052680529.1	Fox	Fox-P	Fox-P	-	
Cari	EV/M0015348.1	Fox	Fox-P	Fox-P	-	
Cflu	DN101403.c0.g1.13.p1	Fox	Fox-P	Fox-P	-	
Cgg	XP_011419330.2	Fox	Fox-P	Fox-P	-	
Cpfl	DN80534.c0.g1.14.p1	Fox	Fox-P	Fox-P	-	
Csin	Hic_asm.12.158	Fox	Fox-P	Fox-P	-	
Cvir	XP_022295865.1	Fox	Fox-P	Fox-P	-	
Dpol	XP_052252868.1	Fox	Fox-P	Fox-P	-	
Gaeq	XP_048239143.1	Fox	Fox-P	Fox-P	-	
Htrf	DN59682.c1.g1.14.p1	Fox	Fox-P	Fox-P	-	
Lorb	XP_052780914.1	Fox	Fox-P	Fox-P	-	
Nare	XP_052101037.1	Fox	Fox-P	Fox-P	-	
Ncal	XP_052083749.1	Fox	Fox-P	Fox-P	-	
Ncal	XP_052100999.1	Fox	Fox-P	Fox-P	-	
Ncal	XP_052061774.1	Fox	Fox-P	Fox-P	-	
Nchi	DNA6200.c0.g1.11.p1	Fox	Fox-P	Fox-P	-	
Nmer	CAC5419517.1	Fox	Fox-P	Fox-P	-	
Nmod	DN215.0.g1.i2.p1	Fox	Fox-P	Fox-P	-	
Ngal	VD114555.1	Fox	Fox-P	Fox-P	-	
Ngal	VD150808.1	Fox	Fox-P	Fox-P	-	
Ngal	VD150806.1	Fox	Fox-P	Fox-P	-	
Ngal	VD150807.1	Fox	Fox-P	Fox-P	-	
Nmar	MMAM0000012411	Fox	Fox-P	Fox-P	-	
Nmer	XP_053376718.1	Fox	Fox-P	Fox-P	-	
Nmod	DN215.0.g1.i2.p1	Fox	Fox-P	Fox-P	-	
Nphr	g33547.tl	Fox	Fox-P	Fox-P	-	
Nphi	scat3947.0.7	Fox	Fox-P	Fox-P	-	
Obin	XP_052822821.1	Fox	Fox-P	Fox-P	-	
Oedu	XP_055998788.1	Fox	Fox-P	Fox-P	-	
Osin	XP_036357838.1	Fox	Fox-P	Fox-P	-	
Pcan	XP_025106713.1	Fox	Fox-P	Fox-P	-	
Fcor	DN1820.0.g1.i26.p1	Fox	Fox-P	Fox-P	-	
Pgen	DN35611.c0.g1.i6.p1	Fox	Fox-P	Fox-P	-	
Pmar	DN1268.c0.g1.11.p1	Fox	Fox-P	Fox-P	-	
Pmax	XP_033745371.1	Fox	Fox-P	Fox-P	-	
Poku	DN88526.c2.g1.11.p1	Fox	Fox-P	Fox-P	-	
Ppur	DN59408.c0.g1.i2.p1	Fox	Fox-P	Fox-P	-	
Pstr	KAK3589495.1	Fox	Fox-P	Fox-P	-	
Pvir	s01329g124	Fox	Fox-P	Fox-P	-	
Pyes	XP_02133304.1	Fox	Fox-P	Fox-P	-	
Rdec	DN8028.c0.g1.11.p1	Fox	Fox-P	Fox-P	-	
Rphi	XP_060586741.1	Fox	Fox-P	Fox-P	-	
Sbro	EVM0004295.1	Fox	Fox-P	Fox-P	-	
Scn	Chr7.2129	Fox	Fox-P	Fox-P	-	
Sgio	Chr7.2133	Fox	Fox-P	Fox-P	-	
Sgio	Sg011345	Fox	Fox-P	Fox-P	-	
Sgra	DN17101.c0.g1.113.p1	Fox	Fox-P	Fox-P	-	
Tgra	KAJ8302344.1	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Acal	XP_005089459.2	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Aicr	Contig1420.28	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Amar	Ama29905	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Apur	scf0ffid.832.35	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
BglA	XP_058853671	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Cang	XP_052696920.1	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Cari	EVM002665.1	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Cgg	XP_011425762.2	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	
Cpli	DN105612.c0.g1.11.p1	Fox	Fox-Q2	Fox-Q2	Annotated as Fox-Q2a	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Csin	Hic.asm.16.4	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cvir	XP_022333968.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Dpol	XP_052280896.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Gaeg	XP_041363029.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Haia	XP_041363041.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
M00000035328	M00000035328	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Hruf	XP_046337359.2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Ncal	XP_052101305.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Nchi	DN22486_c0_g2_i1.p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Ncor	CAC5388792.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Medu	CAG2191193.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Ngal	VD174621.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Nmar	MMA0000000686	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Nmer	XP_045215524.2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Nphi	g20053.tl	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Nphi	scaf22910.1.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Obim	XP_014767584.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pcan	XP_025078472.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pcor	DN99235_c0_g1_i1.p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pfar	DN9446_c0_g1_i1.p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pmax	XP_033751003.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pstr	KAK3595133.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pvir	s00115g23	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pyes	XP_021343668.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Rphi	XP_060571531.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Sbro	EVM023378.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Scn5	Chr5.1974	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Scn5	Chr5.2105	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Sglo	Sg1091183	Fox	Fox-Q2	Forkhead domain Q2 (410788)	Annotated as Fox-Q2a	
Acal	XP_02109004.3	Fox	-	-	-	
Airc	Contig879.9	Fox	-	-	-	
Amar	Amz25052	Fox	-	-	-	
Amar	Amz23735	Fox	-	-	-	
Cang	XP_052676682.1	Fox	-	-	-	
Cang	XP_052680288.1	Fox	-	-	-	
Cang	XP_052677368.1	Fox	-	-	-	
Cari	EVM0011935.1	Fox	-	-	-	
Cari	EVM027332.1	Fox	-	-	-	
Cgig	XP_034306826.1	Fox	-	-	-	
Cgig	XP_0521447567.2	Fox	-	-	-	
Cphi	DN157619_c0_g1_i1.p1	Fox	-	-	-	
Cvir	XP_022300767.1	Fox	-	-	-	
Dpol	XP_052277921.1	Fox	-	-	-	
Dpol	XP_052227296.1	Fox	-	-	-	
Gaeg	XP_041366967.1	Fox	-	-	-	
Gaeg	XP_041378820.1	Fox	-	-	-	
Gaeg	XP_041347225.1	Fox	-	-	-	
Gaeg	XP_041375925.1	Fox	-	-	-	
Gaeg	XP_041375913.1	Fox	-	-	-	
Gaeg	XP_041379015.1	Fox	-	-	-	
Hbia	M00000018946	Fox	-	-	-	
Ngal	VD115902.1	Fox	-	-	-	
Ngal	VD152978.1	Fox	-	-	-	
Nmar	MMA00000049704	Fox	-	-	-	
Nmer	XP_0452166362	Fox	-	-	-	
Nmer	g159704.i1	Fox	-	-	-	
Nphi	scaf46189.0.0	Fox	-	-	-	
Nphi	scaf15444.0.2	Fox	-	-	-	
Nphi	scaf27787.1.10	Fox	-	-	-	
Obim	XP_014777604.1	Fox	-	-	-	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Cediu	XP_048739629.2	Fox				
Csin	XP_036359188.1	Fox				
Csin	XP_029655092.1	Fox				
Fcor	DN89866_c4_g1.i1.p1	Fox				
Fcor	DN55206_c2_g1.i1.p2	Fox				
Fcor	DN15905_c0.g1.i1.p1	Fox				
Fmax	XP_033750561.1	Fox				
Fstr	KAK3285306.1	Fox				
Pvir	S00582648	Fox				
Fyes	XP_021348419.1	Fox				
Rdec	DN2325_c0.g1.i1.p1	Fox				
Rphi	XP_060585776.1	Fox				
Rphi	XP_000511311.1	Fox				
Scon	Ch5_397	Fox				
Sgio	Sg02575	Fox				
Tsqu						
Acal	DN23960_c0.g1.i1.p1	Fox				
Acal	XP_035824685.1	Fox				
Acal	XP_012946205.1	Fox				
Apec	DN48806_c0.g1.i1.p1	Fox				
Apec	DN108003_c0.g1.i1.p1	Fox				
Apur	scaffold_391_70	Fox				
Begl	XP_013078241.2	Fox				
Begl	XP_013078156.1	Fox				
Cang	XP_052697278.1	Fox				
Cang	XP_052713692.1	Fox				
Cari	EV/M0018891.1	Fox				
Cari	EV/M005567.1	Fox				
Craig	XP_011425869.2	Fox				
Craig	XP_034335819.1	Fox				
Cvir	XP_022339078.1	Fox				
Cvir	XP_022339079.1	Fox				
Dpol	DP02271004.1	Fox				
Gaeg	XP_041377139.1	Fox				
Hbia	M00000038049	Fox				
Hbia	M00000004998	Fox				
Hruf	XP_046329595.1	Fox				
Mcru	XP_048239511.1	Fox				
Medu	CAc5384832.1	Fox				
Medu	CAG2253429.1	Fox				
Ngal	VD178477.1	Fox				
Nhnod	DN113112_c0.g1.i1.p1	Fox				
Oblim	XP_014776519.1	Fox				
Cediu	XP_048738250.2	Fox				
Cediu	XP_048752144.2	Fox				
Csin	XP_03638200.1	Fox				
Csin	XP_029654541.1	Fox				
Csin	XP_029636568.1	Fox				
Csin	XP_029637644.1	Fox				
Csin	XP_029636220.1	Fox				
Csin	XP_029657644.1	Fox				
Csin	XP_029657644.1	Fox				
Csin	XP_029657644.1	Fox				
Fcan	DN32781_c0.g1.i1.p1	Fox				
Fcor	DN21964_c0.g1.i2.p1	Fox				
Fmar	DN33290_c0.g1.i2.p1	Fox				
Fmar	DN35008_c0.g1.i4.p1	Fox				
Fmax	XP_033755821.1	Fox				
Fpur	DN4784_c0.g1.i4.p1	Fox				
Fyes	XP_021347051.1	Fox				
Sbro	EV/M0018224.1	Fox				
Sgio	Sg019175	Fox				
Sgio	Sg012029	Fox				
Tsqu	DN039_c0.g1.i1.p1	Fox				

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Acal	XP_005108230.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Acal	XP_03824438.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Airc	Contig449.126	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Airc	Contig44.115	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Amar	Ama33032	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Amar	Ama33828	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Apec	DN29410_c0.g1.i12.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Apec	DN112297_c0.g1.i13.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Apur	scaffold_15489.10	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Apur	scaffold_865.4	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Bglia	XP_013075432.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Bglia	XP_055888106.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cang	XP_052706368.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cang	XP_052705551.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cari	EVM026792.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cari	EVM0013965.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cflu	DN118670_c2.g1.i2.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cflu	DN99542_c1.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cgig	XP_011433975.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cgig	XP_011455662.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cphi	DN31343_c0.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cphi	DN98511_c0.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Csin	Hic.asm.6.930	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Csin	Hic.asm.6.233.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cvir	XP_022286516.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Cvir	XP_022234320.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Dpol	XP_052214544.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Dpol	XP_052217420.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Gaeg	XP_041353075.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Gaeg	XP_041357874.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Hbia	M0000002798	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Hbia	M0000033682	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Hruf	XP_046370193.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Hruf	XP_046326733.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Lorb	DN80278_c0.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Lorb	DN14_c4.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nare	XP_052784720.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nare	XP_052105617.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ncal	XP_052104911.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nchi	DN33632_c0.g1.i2.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nchi	DN45716_c0.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ncor	CAC5401077.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ncor	CAC5413203.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ncor	CAG2229644.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nedu	CAG2206403.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ngal	VD133296.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Ngal	VD165660.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmar	MMAM00000041532	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmar	MMAM00000023253	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmer	XP_045210594.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmer	XP_045201080.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmod	DN1981_c0.g1.i2.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmod	DN78279_c0.g1.i1.p1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmer	g140596.t1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nmer	g157489.t1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Nphi	scaf_63249.0.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Oblim	XP_01478971.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Oblim	XP_014780771.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Oeclu	XP_048746651.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Oeclu	XP_048746653.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Osin	XP_029654000.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Osin	XP_029655838.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Fcan	XP_025079293.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Fcan	XP_025078588.1	Sox	Sox-Bl/2	High mobility group box B	(438790)	
Fcor	DN9087_c0.g1.i2.p1	Sox	Sox-Bl/2	High mobility group box B, A, B and G (438837)		
Fcor	DN14753_c0.g1.i2.p1	Sox	Sox-Bl/2			

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Fcor	DN5688_c0_g1.i10.p2	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fcor	DN5688_c0_g3.i1.p1	Sox		Sox-Bl/2	High mobility group box A, B and G (438837)	
Fcor	DN9742_c0_g1.i2.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fgen	DN5232_c0_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fgen	DN1199_c0_g1.i2.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fmar	DN30477_c0_g2.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fmar	DN30439_c0_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fmax	XP_033760067.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fmax	XP_033759382.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Foku	DN51870_c1_g1.i2.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Foku	DN11067_c0_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fpur	DN202737_c0_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fstr	KAK586311	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fstr	KAK589936.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fvir	s00319g159	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fvir	s00037g281	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fyes	XP_021356125.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fyes	XP_021344413.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Fyes	XP_021372128.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Rdec	DN14777_c2_g7.i2.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Rphi	XP_06056101.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Rphi	XP_060561544.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sbro	EV/M0016386.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sbro	EV/M0007529.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Scon	Chr9_1352	Sox		Sox-Bl/2	High mobility group box B (438790)	
Scon	Chr9_1522	Sox		Sox-Bl/2	High mobility group box B (438790)	
Scon	Chr9_1514	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sgio	Sg010100	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sgio	Sg020107	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sgra	DN3782_c0_g1.i3.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Sgra	DN357_c1_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Tgra	KA18310104.1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Tsqu	DN97880_c0_g1.i1.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Tsqu	DN50556_c0_g1.i2.p1	Sox		Sox-Bl/2	High mobility group box B (438790)	
Aic	Contig80_70	Sox		Sox-C	High mobility group box C (438838)	
Amar	Ama1226	Sox		Sox-C	High mobility group box C (438838)	
Apec	DN12286_c0_g3.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Apur	scffold_16.61	Sox		Sox-C	High mobility group box C (438838)	
Cang	XP_052689209.1	Sox		Sox-C	High mobility group box C (438838)	
Cari	EV/M0025846.1	Sox		Sox-C	High mobility group box C (438838)	
Chlu	DN126276_c0_g1.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Cegig	XP_011445203.1	Sox		Sox-C	High mobility group box C (438838)	
Ccli	DN19112_c0_g1.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Csin	Hic-asn_11_1009	Sox		Sox-C	High mobility group box C (438838)	
Cvrr	XP_022317619.1	Sox		Sox-C	High mobility group box C (438838)	
Dpol	XP_052297395.1	Sox		Sox-C	High mobility group box C (438838)	
Gaeg	XP_041388324.1	Sox		Sox-C	High mobility group box C (438838)	
Hihra	M00000037669	Sox		Sox-C	High mobility group box C (438838)	
Hihra	XP_046350504.1	Sox		Sox-C	High mobility group box C (438838)	
Lorb	DN14941_c0_g1.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Mare	XP_052777703.1	Sox		Sox-C	High mobility group box C (438838)	
Ncal	XP_05208802.1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	DN104308_c0_g4.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	DN44798_c0_g4.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	g26404_t1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	CAC524030_1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	CAG2159937.1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	VD141453.1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	VD114462.1	Sox		Sox-C	High mobility group box C (438838)	
Nlmer	MMA00000036315	Sox		Sox-C	High mobility group box C (438838)	
Oclu	XP_045156937.1	Sox		Sox-C	High mobility group box C (438838)	
Oclu	XP_029634195.1	Sox		Sox-C	High mobility group box C (438838)	
Oclu	XP_025110204.1	Sox		Sox-C	High mobility group box C (438838)	
Oclu	DN2429_c2_g1.i2.p1	Sox		Sox-C	High mobility group box C (438838)	
Oclu	DN929_c2_g1.i1.p1	Sox		Sox-C	High mobility group box C (438838)	
Oclu	DN2572_c0_g1.i1.p1	Sox		Sox-C	High mobility group box C (438838)	

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Pcor	DN353_c2_g3.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Fgen	DN788_c0_g1.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Fmar	DN29124_c0_g2.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Fmax	XP_033737425.1	Sox	Sox-C	High mobility group box C (438838)		
Foku	DN71015_c0_g2.i17.p1	Sox	Sox-C	High mobility group box C (438838)		
Ppur	DN8989_c0_g1.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Pstr	KAK3610995.1	Sox	Sox-C	High mobility group box C (438838)		
Fvir	s00145c243	Sox	Sox-C	High mobility group box C (438838)		
Fyes	XP_021356242.1	Sox	Sox-C	High mobility group box C (438838)		
Rdec	DN52924_c0_g1.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Rphi	XP_060553827.1	Sox	Sox-C	High mobility group box C (438838)		
Sbro	EVM006311.1	Sox	Sox-C	High mobility group box C (438838)		
Scon	Chr8.1790	Sox	Sox-C	High mobility group box C (438838)		
Sgio	Sgi00072	Sox	Sox-C	High mobility group box C (438838)		
Sgra	DN6210_c0_g1.i1.p1	Sox	Sox-C	High mobility group box C (438838)		
Tgra	KA18306264.1	Sox	Sox-C	High mobility group box C (438838)		
Tgra	KA18306266.1	Sox	Sox-C	High mobility group box C (438838)		
Tsqu	DN11669_c1.i12.p1	Sox	Sox-C	High mobility group box C (438838)		
Acal	XP_03824396.1	Sox	Sox-D	High mobility group box (438838)		
Ainc	Contig290.5.1	Sox	Sox-D	High mobility group box (438838)		
Amar	Am23921	Sox	Sox-D	High mobility group box (438838)		
Apec	DN11990_c0_g1.i10.p1	Sox	Sox-D	High mobility group box (438838)		
Apur	scaffold_393.10	Sox	Sox-D	High mobility group box (438838)		
Bela	XP_055899647.1	Sox	Sox-D	High mobility group box (438838)		
Cari	EVM0012405.1	Sox	Sox-D	High mobility group box (438838)		
Chlu	DN124882_c0_g1.i15.p1	Sox	Sox-D	High mobility group box (438838)		
Cfig	XP_011425377.1	Sox	Sox-D	High mobility group box (438838)		
Cipli	DN64448_c0_g1.i1.p1	Sox	Sox-D	High mobility group box (438838)		
Hic.asm	2.1656	Sox	Sox-D	High mobility group box (438838)		
Csin	Hic.asm.2.1600.2	Sox	Sox-D	High mobility group box (438838)		
Cvir	XP_022302926.1	Sox	Sox-D	High mobility group box (438838)		
Dpol	XP_052233125.1	Sox	Sox-D	High mobility group box (438838)		
Gaeg	XP_04136101.1	Sox	Sox-D	High mobility group box (438838)		
Hibra	M00000144098	Sox	Sox-D	High mobility group box (438838)		
Hruf	XP_046329046.1	Sox	Sox-D	High mobility group box (438838)		
Lorb	DN15537_c0_g2.i3.p1	Sox	Sox-D	High mobility group box (438838)		
Ncar	XP_052800695.1	Sox	Sox-D	High mobility group box (438838)		
Ncal	XP_052065962.1	Sox	Sox-D	High mobility group box (438838)		
Nchi	DN138691_c1.g14.p1	Sox	Sox-D	High mobility group box (438838)		
Ncor	CAC5366270.1	Sox	Sox-D	High mobility group box (438838)		
Neudu	CAG2197887.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47525.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47529.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47528.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47527.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47526.1	Sox	Sox-D	High mobility group box (438838)		
Ngal	VDI47530.1	Sox	Sox-D	High mobility group box (438838)		
Nmar	MMA0000004319	Sox	Sox-D	High mobility group box (438838)		
Nmer	XP_053384959.1	Sox	Sox-D	High mobility group box (438838)		
Nmod	DN588_c0_g1.i9.p1	Sox	Sox-D	High mobility group box (438838)		
Nmer	g103147.12	Sox	Sox-D	High mobility group box (438838)		
Nphi	scaf42181.1.3	Sox	Sox-D	High mobility group box (438838)		
Oblim	XP_0528828391.1	Sox	Sox-D	High mobility group box (438838)		
Cedu	XP_0487779633.1	Sox	Sox-D	High mobility group box (438838)		
Osin	XP_029644081.1	Sox	Sox-D	High mobility group box (438838)		
Fcan	DN24654_c0_g1.i2.p1	Sox	Sox-D	High mobility group box (438838)		
Fcor	DN1386.c0_g1.i4.p1	Sox	Sox-D	High mobility group box (438838)		
Fgen	DN371.c1.g1.i1.p1	Sox	Sox-D	High mobility group box (438838)		
Fmar	DN40112_c0_g1.i4.p1	Sox	Sox-D	High mobility group box (438838)		
Fmax	XP_03375161.1	Sox	Sox-D	High mobility group box (438838)		
Foku	DN371.c1.g1.i1.p1	Sox	Sox-D	High mobility group box (438838)		
Fpur	DN3319_c0_g1.i1.p1	Sox	Sox-D	High mobility group box (438838)		
Fvir	KAK3605548.1	Sox	Sox-D	High mobility group box (438838)		
Fyes	s00219ig102	Sox	Sox-D	High mobility group box (438838)		
	XP_021388061.1	Sox	Sox-D	High mobility group box (438838)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psm-ID)	Additional domains (Psm-ID)	Notes
Rdec	DN8093_c0_g1.i1.p1	Sox	Sox-D	High mobility group box (438839)	-	-
Rphi	XP_060604110.1	Sox	Sox-D	High mobility group box (438839)	-	-
Sbro	EV/M0000795.1	Sox	Sox-D	High mobility group box (438839)	-	-
Scon	Chr4.562.1	Sox	Sox-D	High mobility group box (438839)	-	-
Sgra	DN6138_c0_g1.i6.p1	Sox	Sox-D	High mobility group box (438839)	-	-
Tgra	KA1820878.1	Sox	Sox-D	High mobility group box (438839)	-	-
Tsqu	DN55031_c0_g1.i1.p1	Sox	Sox-D	High mobility group box (438839)	-	-
Acal	XP_05102100.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Contig52_209	Sox	Sox-E	High mobility group box E (438840)	-	-	-
Amar	Cari	Sox	Sox-E	High mobility group box E (438840)	-	-
Apec	DNA330_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Apur	scaffold488_7	Sox	Sox-E	High mobility group box E (438840)	-	-
Begl	XP_03091187.2	Sox	Sox-E	High mobility group box E (438840)	-	-
Cang	XP_052668335.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Cari	DNU0407_c5_g2.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Cflu	NP_001295801.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Cgig	NP_013092100.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Gaeq	DN71393_c0_g2.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Hbia	M0000012324	Sox	Sox-E	High mobility group box E (438840)	-	-
Hbia	M0000012325	Sox	Sox-E	High mobility group box E (438840)	-	-
Hfri	XP_04635336.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nare	XP_022312890.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nare	XP_052264587.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nica	XP_041362638.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nichi	M0000012324	Sox	Sox-E	High mobility group box E (438840)	-	-
Nicor	CAC5402442.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Needu	CAC221021.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nigai	VD182092.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nilar	MMAM00000042410	Sox	Sox-E	High mobility group box E (438840)	-	-
Nmer	XP_045213795.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Nimod	DNF8330_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Niphi	scat25414.0.6	Sox	Sox-E	High mobility group box E (438840)	-	-
Cetu	XP_056019113.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fcan	XP_025091282.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fcor	DNA2474_c0_g1.i3.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fcor	DN96098_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fcor	DNA274_c0_g3.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fmar	DN30335_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Fmax	XP_033739301.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Poku	DN87807_c0_g1.i7.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Ppur	DNA6000_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Pstr	KAK3600863.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Pvir	s136484g74	Sox	Sox-E	High mobility group box E (438840)	-	-
Pyes	XP_021348843.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Rphi	XP_060604697.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Sbro	EV/M002110.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Scon	Chr1.75	Sox	Sox-E	High mobility group box E (438840)	-	-
Sgio	Sgi02297	Sox	Sox-E	High mobility group box E (438840)	-	-
Sgra	DN22463_c0_g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	-	-
Tgra	KA18317914.1	Sox	Sox-E	High mobility group box E (438840)	-	-
Tsqu	DN8973_c2_g1.i2.p1	Sox	Sox-F	High mobility group box F (438841)	-	-
Acal	Contig80.101	Sox	Sox-F	High mobility group box F (438841)	-	-
Amar	Ama1616	Sox	Sox-F	High mobility group box F (438841)	-	-
Apur	scaffold546_32	Sox	Sox-F	High mobility group box F (438841)	-	-
Begl	XP_051074628.2	Sox	Sox-F	High mobility group box F (438841)	-	-
Cang	XP_052665434.1	Sox	Sox-F	High mobility group box F (438841)	-	-
Cflu	EV/M0006823.1	Sox	Sox-F	High mobility group box F (438841)	-	-
Cgig	DN139006_c0_g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	-	-
Cpi	XP_011448074.2	Sox	Sox-F	High mobility group box F (438841)	-	-
	DN4414_c0_g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	-	-

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Csin	Hic.asm.11.549	Sox	Sox-F	High mobility group box F (438841)		
Cvir	XP_0223199621	Sox	Sox-F	High mobility group box F (438841)		
Cvir	XP_022314364_1	Sox	Sox-F	High mobility group box F (438841)		
Dpol	XP_052274104_1	Sox	Sox-F	High mobility group box F (438841)		
Gaeg	XP_041359436_1	Sox	Sox-F	High mobility group box F (438841)		
Hbia	M0000015459	Sox	Sox-F	High mobility group box F (438841)		
Hruf	XP_046357912_1	Sox	Sox-F	High mobility group box F (438841)		
Nmar	XP_052774544_1	Sox	Sox-F	High mobility group box F (438841)		
Nmar	XP_052774361_1	Sox	Sox-F	High mobility group box F (438841)		
Ncal	XP_052061059_1	Sox	Sox-F	High mobility group box F (438841)		
Ncor	CAc5414609_1	Sox	Sox-F	High mobility group box F (438841)		
Nedu	CAG22422031_1	Sox	Sox-F	High mobility group box F (438841)		
Nedu	CAG22422050_1	Sox	Sox-F	High mobility group box F (438841)		
Ngal	VD150271_1	Sox	Sox-F	High mobility group box F (438841)		
Ngal	VD150270_1	Sox	Sox-F	High mobility group box F (438841)		
Nmar	MIMM00000025810	Sox	Sox-F	High mobility group box F (438841)		
Nmar	XP_053305054_1	Sox	Sox-F	High mobility group box F (438841)		
Nmod	DN80495_c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Nner	g115494_t1	Sox	Sox-F	High mobility group box F (438841)		
Nphi	scaf61114_0.13	Sox	Sox-F	High mobility group box F (438841)		
Obim	XP_052825684_1	Sox	Sox-F	High mobility group box F (438841)		
Oedu	XP_048764319_2	Sox	Sox-F	High mobility group box F (438841)		
Fcan	XP_025109598_1	Sox	Sox-F	High mobility group box F (438841)		
Fcor	DN11375_c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Fcor	DN19649_c0.g1.i3.p1	Sox	Sox-F	High mobility group box F (438841)		
Fgen	DN144332_c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Fmar	DN24748_c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Fmax	XP_033738287_1	Sox	Sox-F	High mobility group box F (438841)		
Foku	DN11229_c1.g2.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Fstr	KAK3533243_1	Sox	Sox-F	High mobility group box F (438841)		
Fvir	500137g284	Sox	Sox-F	High mobility group box F (438841)		
Fyes	XP_021378109_1	Sox	Sox-F	High mobility group box F (438841)		
Rdec	DN53443_c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)		
Rphi	XP_06059438_1	Sox	Sox-F	High mobility group box F (438841)		
Sbro	EVIM000861_1	Sox	Sox-F	High mobility group box F (438841)		
Scsn	Chr8_397	Sox	Sox-F	High mobility group box F (438841)		
Sgio	Saf003442	Sox	Sox-F	High mobility group box F (438841)		
Sgra	DNS013_c3.g1.i7.p1	Sox	Sox-F	High mobility group box F (438841)		
Arc	Contig1525_38	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Amar	Amz26724	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Apec	DN93182_c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Apur	scaffold_768_3	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Cang	XP_052703370_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Cari	EVMM018164_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Cegig	XP_011415859_3	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Cili	DN80002_c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Hica	Hic.asm.15.1471	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Cvir	XP_022338738_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Dpol	XP_052226448_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Gaeg	XP_041370217_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Gaeg	XP_041369137_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Hbia	M0000001184	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Hruf	XP_046358520_2	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Ncal	XP_052099860_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Ncor	CAC5406014_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nedu	CAG2257203_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Ngal	VD130824_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Ngal	VD130823_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nmar	MIMM0000015662	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nmar	XP_053407277_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nner	g125234_t1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nphi	scaf12010_0.4	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Nphi	scaf59202_0.9	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Obim	XP_052832677_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Oedu	XP_056006679_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Osin	XP_036368794_1	Sox	Sox-H	High mobility group box Sox-30 (438842)		
Pcor	DN186456_c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)		

Tab. S6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Psmn-ID)	Additional domains (Psmn-ID)	Notes
Fmar	DN40950_c1.g1.i2.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Fmax	XP_033756818_1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Foku	DN16718_c0.g1.i6.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Fpur	DNF268_c0.g1.i11.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Fstr	KAK3522760_1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Fvir	s00451_g108	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Fyes	XP_021340986_1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Rdec	DN22482_c4.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Rphi	XP_060578490_1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Scon	Ch15_1899	Sox	Sox-H	High mobility group box Sox-30 (438820)	-	-
Sgio	Sg010047	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-
Tsqu	DN874_c7.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	-	-

Supplementary Table S7 – Proportions of missing data in both bivalve species and DSFGs. Bivalve species represented by transcriptomic data are highlighted with an asterisk (*).

Species	Species		Genes	
	Species	% missing data (out of 33 DSFGs)	Group	% missing data (out of 43 species)
<i>A. irradians concentricus</i>		0.000000	<i>Dmrt-1L</i>	48.837209
<i>A. marissinica</i>		21.212121	<i>Dmrt-3</i>	30.232558
<i>A. pectinata*</i>		48.484848	<i>Dmrt-2</i>	55.813953
<i>A. purpuratus</i>		6.060606	<i>Dmrt-4/5</i>	6.976744
<i>C. angulata</i>		6.060606	<i>Fox-A</i>	13.953488
<i>C. ariakensis</i>		3.030303	<i>Fox-B</i>	30.232558
<i>C. fluminea*</i>		42.424242	<i>Fox-C</i>	23.255814
<i>C. gigas</i>		6.060606	<i>Fox-D</i>	20.930233
<i>C. plicata*</i>		21.212121	<i>Fox-E</i>	30.232558
<i>C. sinensis</i>		21.212121	<i>Fox-F</i>	18.604651
<i>C. virginica</i>		3.030303	<i>Fox-G</i>	16.279070
<i>D. polymorpha</i>		9.090909	<i>Fox-H</i>	41.860465
<i>H. bialata</i>		9.090909	<i>Fox-J1</i>	0.000000
<i>L. orbiculatus*</i>		63.636364	<i>Fox-J2/3</i>	9.302326
<i>M. arenaria</i>		21.212121	<i>Fox-L1</i>	18.604651
<i>M. californianus</i>		9.090909	<i>Fox-L2</i>	13.953488
<i>M. chinensis*</i>		57.575758	<i>Fox-N1/4</i>	6.976744
<i>M. coruscus</i>		0.000000	<i>Fox-N2/3</i>	6.976744
<i>M. edulis</i>		3.030303	<i>Fox-O</i>	13.953488
<i>M. galloprovincialis</i>		6.060606	<i>Fox-P</i>	9.302326
<i>M. margaritifera</i>		6.060606	<i>Fox-Q2</i>	30.232558
<i>M. mercenaria</i>		3.030303	<i>Fox-OG13/NA</i>	32.558140
<i>M. modiolus*</i>		36.363636	<i>Fox-OG15/NA</i>	34.883721
<i>M. nervosa</i>		27.272727	<i>Fox-OG16/NA</i>	30.232558
<i>M. philippinarum</i>		9.090909	<i>Fox-OG2/NA</i>	16.279070
<i>O. edulis</i>		6.060606	<i>Fox-OG28/NA</i>	39.534884
<i>P. coreanum*</i>		18.181818	<i>Fox-OG39/NA</i>	37.209302
<i>P. generosa*</i>		54.545455	<i>Sox-B1/2</i>	0.000000
<i>P. margaritifera*</i>		21.212121	<i>Sox-C</i>	0.000000
<i>P. maximus</i>		0.000000	<i>Sox-D</i>	4.651163
<i>P. okutanii*</i>		54.545455	<i>Sox-E</i>	9.302326
<i>P. purpuratus*</i>		54.545455	<i>Sox-F</i>	13.953488
<i>P. streckersoni</i>		6.060606	<i>Sox-H</i>	20.930233
<i>P. viridis</i>		3.030303		
<i>P. yessoensis</i>		3.030303		
<i>R. decussatus*</i>		51.515152		
<i>R. philippinarum</i>		3.030303		
<i>S. broughtonii</i>		12.121212		
<i>S. constricta</i>		12.121212		
<i>S. glomerata</i>		9.090909		
<i>S. grandis*</i>		54.545455		
<i>T. granosa</i>		42.424242		
<i>T. squamosa*</i>		48.484848		

Tab. S8 continued from previous page

Sp. ID	Gene ID	Group	Annotation	Sp. ID	Gene ID	Group	Annotation
Scar	XP_047375686.1	Fox	Fox-E3	Bmus	XP_036717651.1	Sox	Sox-6
Oana	XP_001516678.2	Fox	Fox-D4	Mjav	XP_036882886.1	Sox	Sox-6
Pgig	XP_039744744.1	Fox	Fox-D4	Amel	XP_034501998.1	Sox	Sox-6
Hamp	XP_057576439.1	Fox	Fox-D4	Equa	XP_046493708.1	Sox	Sox-6
Pafr	XP_047625506.1	Fox	Fox-D4	Mdom	XP_007497087.2	Sox	Sox-6
Hsap	NP_001119806.1	Fox	Fox-D4	Cpor	XP_023419042.1	Sox	Sox-6
Emax	XP_049752498.1	Fox	Fox-D4	Pgig	XP_039707456.1	Sox	Sox-6
Cdid	XP_037706680.1	Fox	Fox-D4	Oafe	XP_007955001.1	Sox	Sox-6
Hsap	NP_954714.2	Fox	Fox-D4	Cdid	XP_037695265.1	Sox	Sox-6
Rfer	XP_032979406.1	Fox	Fox-D4	Casi	XP_006865846.1	Sox	Sox-6
Ttru	XP_033714036.1	Fox	Fox-D4	Cimi	XP_017381104.1	Sox	Sox-6
Ptig	XP_042820135.1	Fox	Fox-D4	Mmus	XP_036008726.1	Sox	Sox-6
Cdro	XP_010998361.2	Fox	Fox-D4	Ptig	XP_042815591.1	Sox	Sox-6
Mang	XP_045739597.1	Fox	Fox-D4	Pafr	XP_047632165.1	Sox	Sox-6
Scar	XP_047380997.1	Fox	Fox-D4	Tman	XP_023593283.1	Sox	Sox-6
Drot	XP_053785855.1	Fox	Fox-D4	Drot	XP_053780580.1	Sox	Sox-6
Hsap	NP_001078945.1	Fox	Fox-D4	Mang	XP_054368007.1	Sox	Sox-6
Amel	XP_034503132.1	Fox	Fox-D4	Csim	XP_014642191.1	Sox	Sox-6
Oafe	XP_007954771.1	Fox	Fox-D4	Scar	XP_047372506.1	Sox	Sox-6
Equa	XP_046519849.1	Fox	Fox-D4	Lcat	XP_045413549.1	Sox	Sox-6
Bbub	XP_025137372.3	Fox	Fox-D4	Cdro	XP_010975526.2	Sox	Sox-6
Dnov	XP_023444807.2	Fox	Fox-D4	Clup	XP_038425054.1	Sox	Sox-6
Bmus	XP_036712053.1	Fox	Fox-D4	Emax	XP_049725376.1	Sox	Sox-18
Hsap	NP_036316.1	Fox	Fox-D4	Casi	XP_006873390.1	Sox	-
Opri	XP_004591937.2	Fox	Fox-D4	Oafe	XP_007952883.1	Sox	-
Cimi	XP_017398902.1	Fox	Fox-D4	Mmus	NP_035577.1	Sox	Sox-8
Mdom	XP_001373972.1	Fox	Fox-D4	Opri	XP_004596857.2	Sox	-
Mmus	NP_032048.1	Fox	Fox-D4	Cpor	XP_003478479.1	Sox	-
Hsap	NP_997188.2	Fox	Fox-D4	Oana	XP_028913325.1	Sox	Sox-8
Hsap	NP_954586.4	Fox	Fox-D4	Ggal	NP_990062.2	Sox	-
Csim	XP_014650927.1	Fox	Fox-D4	Cpor	XP_023421073.1	Sox	-
Cpor	XP_013010584.1	Fox	Fox-D4	Cimi	XP_017378281.2	Sox	-
Mjav	XP_036847234.1	Fox	Fox-D4	Cdid	XP_037702077.1	Sox	-
Casi	XP_006863895.1	Fox	Fox-D4	Shar	XP_031794321.1	Sox	-
Mang	XP_045757341.1	Fox	-	Mdom	XP_016280870.1	Sox	-
Clup	XP_038401447.1	Fox	-	Ggal	XP_040553151.1	Sox	-
Amel	XP_019655962.1	Fox	-	Oana	XP_028913966.1	Sox	Sox-5
Clup	XP_038439326.1	Fox	-	Lcat	XP_045422920.1	Sox	-
Ggal	NP_001382914.1	Fox	Fox-C1				

Tab. S9 continued from previous page

Species ID	Gene ID	Group	Annotation	Species ID	Gene ID	Group	Annotation
Dana	XP_001958356.2	Fox	Biniou	Dkik	XP_017034345.1	Sox	Sox-21A
Dari	XP_017864430.1	Fox	Biniou	Dbus	XP_0178640298.2	Sox	Sox-21A
Dbus	XP_017841230.1	Fox	Biniou	Dari	XP_017862178.1	Sox	Sox-21A
Dwil	XP_002061803.1	Fox	Biniou	Dalb	XP_034108985.1	Sox	Sox-21A
Dgri	XP_001983803.1	Fox	Biniou	Dser	XP_020816311.1	Sox	Sox-21A
Dere	XP_001971514.2	Fox	Biniou	Agam	XP_061513404.1	Sox	Sox-21A
Dsuz	XP_036672243.1	Fox	Biniou	Dele	XP_017128154.1	Sox	Sox-21A
Dpse	XP_002134730.3	Fox	Biniou	Dana	XP_032310341.1	Sox	Sox-21A
Dhyd	XP_030081482.1	Fox	Biniou	Dhyd	XP_023173058.2	Sox	Sox-21A
Dbib	XP_017095285.2	Fox	Biniou	Dere	XP_001972711.1	Sox	Sox-21A
Dmir	XP_017139097.1	Fox	Biniou	Dbib	XP_043070009.1	Sox	Sox-21A
Dsec	XP_002035641.1	Fox	Biniou	Dwil	XP_023033998.1	Sox	Sox-21A
Agam	XP_061503012.1	Fox	Fd-2/Fox-L1	Dmel	NP_001261827.1	Sox	Sox-21A
Dmel	NP_001246609.1	Fox	Fd-2/Fox-L1	Dpse	XP_033239763.1	Sox	Sox-21A
Dhyd	XP_023178350.2	Fox	Fd-2/Fox-L1	Dwil	XP_002066518.2	Sox	Sox-N
Dbus	XP_017840363.1	Fox	Fd-2/Fox-L1	Dser	XP_020807900.1	Sox	Sox-N
Dkik	XP_017019394.1	Fox	Fd-2/Fox-L1	Dsuz	XP_036678170.1	Sox	Sox-N
Dser	XP_020807718.1	Fox	Fd-2/Fox-L1	Dbus	XP_033150665.1	Sox	Sox-N
Dari	XP_017863042.1	Fox	Fd-2/Fox-L1	Dsec	XP_002036270.2	Sox	Sox-N
Dsec	XP_002035260.1	Fox	Fd-2/Fox-L1	Dgri	XP_001993541.2	Sox	Sox-N
Dere	XP_001971815.2	Fox	Fd-2/Fox-L1	Dele	XP_017121213.1	Sox	Sox-N
Dalb	XP_051860220.1	Fox	Fd-2/Fox-L1	Dari	XP_017860249.1	Sox	Sox-N
Dana	XP_001956070.1	Fox	Fd-2/Fox-L1	Dalb	XP_051858933.1	Sox	Sox-N
Dsuz	XP_016932589.2	Fox	Fd-2/Fox-L1	Dkik	XP_017024634.1	Sox	Sox-N
Dele	XP_017126783.1	Fox	Fd-2/Fox-L1	Agam	XP_061516479.1	Sox	Sox-N
Dgri	XP_001983216.1	Fox	Fd-2/Fox-L1	Dmir	XP_017154597.1	Sox	Sox-N
Dpse	XP_001352427.1	Fox	Fd-2/Fox-L1	Dbib	XP_017099296.2	Sox	Sox-N
Dwil	XP_002062380.1	Fox	Fd-2/Fox-L1	Dere	XP_026834908.1	Sox	Sox-N
Dbib	XP_01709814.2	Fox	Fd-2/Fox-L1	Dmel	NP_001260269.1	Sox	Sox-N
Dmir	XP_017137846.1	Fox	Fd-2/Fox-L1	Dhyd	XP_023172210.2	Sox	Sox-N
Dere	XP_001977848.1	Fox	Fd-19B	Dpse	XP_001355808.4	Sox	Sox-N
Dhyd	XP_030079337.1	Fox	Fd-19B	Dana	XP_001962613.2	Sox	Sox-N
Dmel	NP_608369.	Fox	Fd-19B	Dbus	XP_017845374.1	Sox	Sox-100B
Dsec	XP_002039497.1	Fox	Fd-19B	Dari	XP_017874440.1	Sox	Sox-100B
Dsuz	XP_016923819.2	Fox	Fd-19B	Agam	XP_061509035.1	Sox	Sox-100B
Dele	XP_041565246.1	Fox	Fd-19B	Dhyd	XP_023168522.1	Sox	Sox-100B
Dpse	XP_033239022.1	Fox	Fd-19B	Dkik	XP_017037336.1	Sox	Sox-100B
Dmir	XP_017134931.2	Fox	Fd-19B	Dana	XP_001954720.1	Sox	Sox-100B
Dari	XP_017869476.1	Fox	Fd-19B	Dele	XP_017122478.1	Sox	Sox-100B
Dgri	XP_043072169.1	Fox	Fd-19B	Dser	XP_020811779.1	Sox	Sox-100B
Dalb	XP_034101064.2	Fox	Fd-19B	Dbib	XP_017088796.2	Sox	Sox-100B
Dwil	XP_046867097.1	Fox	Fd-19B	Dmir	XP_017145320.1	Sox	Sox-100B
Dser	XP_020808848.1	Fox	Fd-19B	Dalb	XP_051863654.1	Sox	Sox-100B
Dkik	XP_017017955.1	Fox	Fd-19B	Dmel	NP_651839.1	Sox	Sox-100B
Dbus	XP_033150229.1	Fox	Fd-19B	Dsec	XP_002037434.1	Sox	Sox-100B
Agam	XP_061512060.1	Fox	-	Dere	XP_001981122.1	Sox	Sox-100B
Dalb	XP_034098689.1	Fox	-	Dsuz	XP_016935593.2	Sox	Sox-100B
Dgri	XP_032595015.1	Fox	-	Dwil	XP_023035523.1	Sox	Sox-100B
Dalb	XP_034098925.1	Fox	Slp-1	Dgri	XP_032597554.1	Sox	Sox-100B
Dele	XP_017110574.1	Fox	Slp-1	Dpse	XP_001357577.3	Sox	Sox-100B

Supplementary Table S10 – All the enriched GO terms for Group 1 and Group 2 genes of bivalves, mammals, and *Drosophila*.

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1 + Group 2	GO:0060255	regulation of macromolecule metabolic process	737	59	31.91	0.0453
Bivalvia	Group 1 + Group 2	GO:0080990	regulation of primary metabolic process	673	53	29.14	0.0182
Bivalvia	Group 1 + Group 2	GO:0019219	regulation of nucleobase-containing compound metabolic process	541	41	23.42	0.0239
Bivalvia	Group 1 + Group 2	GO:0063551	DNA-templated transcription	571	39	24.72	0.0377
Bivalvia	Group 1 + Group 2	GO:0032774	RNA biosynthetic process	579	39	25.07	0.0449
Bivalvia	Group 1 + Group 2	GO:0051522	regulation of RNA metabolic process	517	37	22.38	0.0272
Bivalvia	Group 1 + Group 2	GO:0006355	regulation of DNA-templated transcription	490	35	21.22	0.0375
Bivalvia	Group 1 + Group 2	GO:2001141	regulation of RNA biosynthetic process	491	35	21.26	0.0384
Bivalvia	Group 1 + Group 2	GO:0066950	response to stress	370	33	16.02	0.0195
Bivalvia	Group 1 + Group 2	GO:0032502	developmental process	261	27	11.30	0.0445
Bivalvia	Group 1 + Group 2	GO:0064688	protein phosphorylation	345	23	14.94	0.0248
Bivalvia	Group 1 + Group 2	GO:0031325	positive regulation of cellular metabolic process	125	17	5.41	0.0080
Bivalvia	Group 1 + Group 2	GO:0016064	positive regulation of macromolecule metabolic process	151	17	6.54	0.0405
Bivalvia	Group 1 + Group 2	GO:0051172	negative regulation of nitrogen compound metabolic process	117	16	5.07	0.0081
Bivalvia	Group 1 + Group 2	GO:0051173	positive regulation of nitrogen compound metabolic process	137	15	5.93	0.0245
Bivalvia	Group 1 + Group 2	GO:0063110	DNA recombination	66	14	2.86	0.0009
Bivalvia	Group 1 + Group 2	GO:0048513	animal organ development	83	12	3.59	0.0409
Bivalvia	Group 1 + Group 2	GO:0010629	negative regulation of gene expression	78	11	3.38	0.0005
Bivalvia	Group 1 + Group 2	GO:0023051	regulation of signaling	133	11	5.76	0.0287
Bivalvia	Group 1 + Group 2	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	64	11	2.77	0.0364
Bivalvia	Group 1 + Group 2	GO:0069605	response to external stimulus	90	11	3.90	0.0454
Bivalvia	Group 1 + Group 2	GO:0044119	biological process involved in interspecies interaction between organisms	63	11	2.73	0.0476
Bivalvia	Group 1 + Group 2	GO:0069195	apoptotic process	95	10	4.11	0.0077
Bivalvia	Group 1 + Group 2	GO:0099666	regulation of signal transduction	120	10	5.20	0.0345
Bivalvia	Group 1 + Group 2	GO:00646417	regulation of translation	52	9	2.25	0.0003
Bivalvia	Group 1 + Group 2	GO:0045892	negative regulation of DNA-templated transcription	59	9	2.55	0.0297
Bivalvia	Group 1 + Group 2	GO:1926779	negative regulation of RNA biosynthetic process	59	9	2.55	0.0297
Bivalvia	Group 1 + Group 2	GO:0096007	response to biotic stimulus	55	9	2.38	0.0321
Bivalvia	Group 1 + Group 2	GO:0051253	negative regulation of RNA metabolic process	61	9	2.64	0.0372
Bivalvia	Group 1 + Group 2	GO:0069522	regulation of response to stimulus	58	9	2.51	0.0416
Bivalvia	Group 1 + Group 2	GO:006302	double-strand break repair	52	9	2.25	0.0486
Bivalvia	Group 1 + Group 2	GO:0080134	regulation of response to stress	52	9	2.25	0.0486
Bivalvia	Group 1 + Group 2	GO:0010564	regulation of cell cycle process	43	8	1.86	0.0067
Bivalvia	Group 1 + Group 2	GO:0042981	regulation of apoptotic process	70	8	3.03	0.0102
Bivalvia	Group 1 + Group 2	GO:0043067	regulation of programmed cell death	72	8	3.12	0.0121
Bivalvia	Group 1 + Group 2	GO:0048834	positive regulation of response to stimulus	61	8	2.64	0.0400
Bivalvia	Group 1 + Group 2	GO:0013110	cellular response to organic substance	58	9	2.25	0.0486
Bivalvia	Group 1 + Group 2	GO:0010628	positive regulation of gene expression	34	7	1.47	0.0266
Bivalvia	Group 1 + Group 2	GO:0045944	positive regulation of transcription by RNA polymerase II	38	6	1.65	0.0054
Bivalvia	Group 1 + Group 2	GO:1901987	regulation of cell cycle phase transition	29	6	1.26	0.0237
Bivalvia	Group 1 + Group 2	GO:2000779	regulation of double-strand break repair	11	6	0.48	0.0243
Bivalvia	Group 1 + Group 2	GO:0051247	positive regulation of protein metabolic process	54	6	2.34	0.0282
Bivalvia	Group 1 + Group 2	GO:0051248	negative regulation of protein metabolic process	55	6	2.38	0.0305
Bivalvia	Group 1 + Group 2	GO:0088657	import into cell	56	6	2.42	0.0330
Bivalvia	Group 1 + Group 2	GO:1902531	regulation of intracellular signal transduction	59	6	2.55	0.0412
Bivalvia	Group 1 + Group 2	GO:0044770	cell cycle phase transition	35	6	1.52	0.0467
Bivalvia	Group 1 + Group 2	GO:0000122	negative regulation of transcription by RNA polymerase II	31	5	1.34	0.0099
Bivalvia	Group 1 + Group 2	GO:006402	mRNA catabolic process	35	5	1.52	0.0164
Bivalvia	Group 1 + Group 2	GO:1901980	regulation of mitotic cell cycle phase transition	18	5	0.78	0.0244
Bivalvia	Group 1 + Group 2	GO:1901698	response to nitrogen compound	41	5	1.78	0.0307
Bivalvia	Group 1 + Group 2	GO:006601	RNA catabolic process	44	5	1.91	0.0401
Bivalvia	Group 1 + Group 2	GO:005155	regulation of cell adhesion	11	4	0.48	0.0009
Bivalvia	Group 1 + Group 2	GO:0048568	embryonic organ development	12	4	0.52	0.0013
Bivalvia	Group 1 + Group 2	GO:007517	muscle organ development	13	4	0.56	0.0018
Bivalvia	Group 1 + Group 2	GO:0051607	defense response to virus	13	4	0.56	0.0018
Bivalvia	Group 1 + Group 2	GO:0010569	regulation of double-strand break repair via homologous recombination	5	4	0.22	0.0054
Bivalvia	Group 1 + Group 2	GO:0042274	ribosomal small subunit biogenesis	21	4	0.91	0.0115
Bivalvia	Group 1 + Group 2	GO:0043066	negative regulation of apoptotic process	28	4	1.21	0.0310
Bivalvia	Group 1 + Group 2	GO:0043069	negative regulation of programmed cell death	29	4	1.26	0.0348
Bivalvia	Group 1 + Group 2	GO:0016477	cell migration	29	4	1.26	0.0348
Bivalvia	Group 1 + Group 2	GO:0052101	regulation of response to external stimulus	30	4	1.30	0.0388
Bivalvia	Group 1 + Group 2	GO:005769	positive regulation of neurogenesis	5	3	0.22	0.0008
Bivalvia	Group 1 + Group 2	GO:007368	determination of left/right symmetry	30	3	0.30	0.0025
Bivalvia	Group 1 + Group 2	GO:0001819	positive regulation of cytokine production	7	3	0.30	0.0025
Bivalvia	Group 1 + Group 2	GO:0070192	chromosome organization involved in meiotic cell cycle	7	3	0.35	0.0038
Bivalvia	Group 1 + Group 2	GO:0045132	meiotic chromosome segregation	8	3	0.35	0.0038

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1 + Group 2	GO:0042266	negative regulation of phosphorylation	10	3	0.43	0.0077
Bivalvia	Group 1 + Group 2	GO:0068285	negative regulation of cell population proliferation	10	3	0.43	0.0077
Bivalvia	Group 1 + Group 2	GO:0026044	regulation of cell morphogenesis	10	3	0.43	0.0077
Bivalvia	Group 1 + Group 2	GO:0001894	tissue morphogenesis	10	3	0.43	0.0077
Bivalvia	Group 1 + Group 2	GO:0003007	heart morphogenesis	10	3	0.43	0.0077
Bivalvia	Group 1 + Group 2	GO:0051093	negative regulation of developmental process	11	3	0.48	0.0102
Bivalvia	Group 1 + Group 2	GO:0001501	skeletal system development	11	3	0.48	0.0102
Bivalvia	Group 1 + Group 2	GO:0042327	positive regulation of phosphorylation	12	3	0.52	0.0132
Bivalvia	Group 1 + Group 2	GO:0010562	positive regulation of phosphorus metabolic process	13	3	0.56	0.0166
Bivalvia	Group 1 + Group 2	GO:0005010	active nucleation	13	3	0.56	0.0166
Bivalvia	Group 1 + Group 2	GO:0045937	positive regulation of phosphate metabolic process	13	3	0.56	0.0166
Bivalvia	Group 1 + Group 2	GO:000727	meiosis I	14	3	0.61	0.0205
Bivalvia	Group 1 + Group 2	GO:0031400	negative regulation of protein modification process	14	3	0.61	0.0205
Bivalvia	Group 1 + Group 2	GO:0001982	meiosis I cell cycle process	14	3	0.61	0.0205
Bivalvia	Group 1 + Group 2	GO:0057190	apoptotic signaling pathway	16	3	0.69	0.0296
Bivalvia	Group 1 + Group 2	GO:0040008	regulation of growth	16	3	0.69	0.0296
Bivalvia	Group 1 + Group 2	GO:0051345	positive regulation of hydrolase activity	17	3	0.74	0.0348
Bivalvia	Group 1 + Group 2	GO:0010257	NADH dehydrogenase complex assembly	17	3	0.74	0.0348
Bivalvia	Group 1 + Group 2	GO:0032981	mitochondrial respiratory chain complex I assembly	17	3	0.74	0.0348
Bivalvia	Group 1 + Group 2	GO:0032880	regulation of protein localization	18	3	0.78	0.0405
Bivalvia	Group 1 + Group 2	GO:0005976	polysaccharide metabolic process	18	3	0.78	0.0405
Bivalvia	Group 1 + Group 2	GO:0048729	tissue morphogenesis	19	3	0.82	0.0466
Bivalvia	Group 1 + Group 2	GO:0018022	peptidyl-lysine methylation	19	3	0.82	0.0466
Bivalvia	Group 1 + Group 2	GO:0000041	transition metal ion transport	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:0032488	Cdc42 protein signal transduction	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:0026260	digestive system process	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:0007097	nuclear migration	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:0032232	negative regulation of actin filament bundle assembly	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:1905168	positive regulation of double-strand break repair via homologous recombination	2	2	0.09	0.0019
Bivalvia	Group 1 + Group 2	GO:0002664	epithelial cell development	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0061383	trabecula morphogenesis	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0010830	regulation of myotube differentiation	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0010833	telomere maintenance via telomere lengthening	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0009595	mitochondrial RNA metabolic process	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0036117	mitochondrial cytochrome c oxidase assembly	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	CO:2000179	positive regulation of neural precursor cell proliferation	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0050777	negative regulation of immune response	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0070955	mitotic G2 DNA damage checkpoint signaling	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:200736	regulation of stem cell differentiation	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0010833	telomere capping	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0045910	negative regulation of DNA recombination	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0051701	biological process involved in interaction with host	3	2	0.13	0.0054
Bivalvia	Group 1 + Group 2	GO:0046677	response to antibiotic	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0046620	regulation of organ growth	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0035014	negative regulation of BMP signaling pathway	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:1901678	iron coordination entity transport	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0003181	atrioventricular valve morphogenesis	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0007519	skeletal muscle tissue development	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0032507	maintenance of protein location in cell	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0074116	synapse assembly	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0087288	ncRNA transcription	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0035023	regulation of Rho protein signal transduction	4	2	0.17	0.0106
Bivalvia	Group 1 + Group 2	GO:0003190	atrioventricular valve formation	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0003181	negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:007162	protein refolding	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0010001	glial cell differentiation	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0007129	negative regulation of cellular response to growth factor stimulus	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0035010	regulation of BMP signalling pathway	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0061371	determination of heart left/right asymmetry	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0042063	gliogenesis	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0007162	negative regulation of cell adhesion	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0007129	homologous chromosome pairing at meiosis	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0001947	heart looping	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0035050	embryonic heart tube development	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0035265	organ growth	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0030968	endoplasmic reticulum unfolded protein response	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0003188	heart valve formation	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0003171	atrioventricular valve development	5	2	0.22	0.0171

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1 + Group 2	GO:0003179	heart valve morphogenesis	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0003143	embryonic heart tube morphogenesis	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0050287	regulation of cellular response to growth factor stimulus	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0005338	skeletal muscle organ development	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0034620	cellular response to unfolded protein	5	2	0.22	0.0171
Bivalvia	Group 1 + Group 2	GO:0030509	BMP signaling pathway	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0006360	transcription by RNA polymerase I	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0045185	maintenance of protein location	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0045143	homologous chromosome segregation	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0001889	liver development	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0071772	response to BMP	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0071773	cellular response to BMP stimulus	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0060092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0003170	heart valve development	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0003012	muscle system process	6	2	0.26	0.0250
Bivalvia	Group 1 + Group 2	GO:0030490	maturity of SSU-rRNA	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0018105	peptidyl-serine phosphorylation	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0032465	regulation of cytokinesis	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0061448	connective tissue development	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0018209	peptidyl-serine modification	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0032102	negative regulation of response to external stimulus	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0061008	hepatocyte system development	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0048638	regulation of developmental growth	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0010212	response to ionizing radiation	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:0034329	cell junction assembly	7	2	0.30	0.0340
Bivalvia	Group 1 + Group 2	GO:1901632	response to peptide	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0048732	gland development	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0033157	regulation of intracellular protein transport	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0051302	regulation of cell division	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0001822	kidney development	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0050808	synapse organization	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0035067	cellular response to topologically incorrect protein	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0031032	actomyosin structure organization	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0001503	osification	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0002071	polysaccharide biosynthetic process	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0008933	regulation of Notch signalling pathway	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:001822	negative regulation of cell differentiation	8	2	0.35	0.0440
Bivalvia	Group 1 + Group 2	GO:0045536	regulation of defense response to virus by host	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0050691	semaphorin-plexin signaling pathway involved in bone trabecula morphogenesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0002020	negative regulation of organ growth	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0046621	animal organ maturation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048799	O antigen metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0046402	response to host immune response	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0052572	malate transport	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0015743	regulation of extrinsic apoptotic signaling pathway via death domain receptors	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0043628	inhibitory synapse assembly	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1904862	mitochondrial RNA processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:000963	positive regulation of defense response to virus by host	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0002320	regulation of skeletal muscle fiber development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048742	positive regulation of osteoblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:003688	negative regulation of osteoblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:003689	protein poly-ADP-ribosylation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:005136	response to host	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1902285	semaphorin-plexin signaling pathway involved in neuron projection guidance	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:000963	osteoblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:003687	lipopolysaccharide metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1902287	semaphorin-plexin signaling pathway involved in axon guidance	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:003688	regulation of osteoblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1902041	negative regulation of exit from mitosis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0071423	protein auto-ADP-ribosylation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:008653	malate transmembrane transport	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0030653	beta-lactam antibiotic catabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0030655	beta-lactam antibiotic catabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0001100	negative regulation of exit from mitosis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0068625	extrinsic apoptotic signaling pathway via death domain receptors	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:003144	sno(s)RNA processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0031848	protection from non-homologous end joining at telomere	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0010526	retrotransposon silencing	1	1	0.04	0.0433

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1 + Group 2	GO:0061668	mitochondrial ribosome assembly	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0035418	protein localization to synapse	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:001430	bone trabecula morphogenesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0006356	regulation of transcription by RNA polymerase I	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1901269	lipopolysaccharide metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0035622	intrahepatic bile duct development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0016444	somatic cell DNA recombinant	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1901271	lipopolysaccharide biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0052173	response to defenses of other organism	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0003382	epithelial cell morphogenesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0070977	bone maturation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0031571	mitotic G1 DNA damage checkpoint signaling	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0043247	telomere maintenance in response to DNA damage	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0010669	epithelial structure maintenance	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0093301	snRNA transcription	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0042149	cellular response to glucose starvation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1902914	regulation of protein polyubiquitination	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032196	negative regulation of protein polyubiquitination	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032197	transposition	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0034964	retrotransposition	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0007168	receptor guanylyl cyclase signaling pathway	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0017001	antibiotic catabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0030277	maintenance of gastrointestinal epithelium	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0000495	box H/ACA sno(s)RNA 3'-end processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0000495	box H/ACA sno(s)RNA 3'-end processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1903513	endoplasmic reticulum to cytosol transport	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1900481	mRNA pseudouridine synthesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0044819	mitotic G1/S transition checkpoint signaling	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0051155	regulation of protein transport, ER to cytosol	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0007140	regulation of striated muscle cell differentiation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0050722	male meiotic nuclear division	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0050722	positive regulation of axonogenesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0060646	mitochondrial tRNA processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0091772	presynapse organization	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032065	maintenance of protein location in cell cortex	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1904152	regulation of retrograde protein transport, ER to cytosol	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0097222	mitochondrial mRNA polyadenylation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0051149	positive regulation of muscle cell differentiation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0020033	antigenic variation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0052200	response to host defenses	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0016074	syn(s)RNA metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0043931	osification involved in bone maturation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0035279	mRNA-mediated gene silencing by mRNA destabilization	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0061009	common bile duct development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0090544	presynapse assembly	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048641	regulation of skeletal muscle tissue development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048643	positive regulation of skeletal muscle tissue development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1901861	regulation of muscle tissue development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032783	evasion of host immune response	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0030970	retrograde protein transport, ER to cytosol	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048634	regulation of muscle organ development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0031220	snRNA pseudouridine synthesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0031126	sn(s)RNA 3'-end processing	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0033979	box H/ACA sno(s)RNA metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0071966	fungal cell wall polysaccharide metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0042783	regulation of skeletal muscle tissue development	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0030970	evasion of host immune response	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048634	retrograde protein transport, ER to cytosol	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0051274	beta-glucan biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0051278	fungal-type cell wall polysaccharide biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0099103	lipopolysaccharide biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0052922	nuclear matrix anchoring at nuclear membrane	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0071947	protein deubiquitination involved in ubiquitin-dependent protein catabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0045943	positive regulation of transcription by RNA polymerase I	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0081144	fibroblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048145	regulation of fibroblast proliferation	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0072340	lactam catabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0006074	(1->3)-beta-D-glucan metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0006075	(1->3)-beta-D-glucan biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0016999	antibiotic metabolic process	1	1	0.04	0.0433

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1 + Group 2	GO:0072238	lactam metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0044650	adhesion of symbiont to host cell	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:192414	protein localization to cell junction	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0009272	fungal-type cell wall biogenesis	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0069243	O antigen biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0043578	nuclear matrix organization	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0069245	lipid A biosynthetic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0044406	adhesion of symbiont to host	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032978	protein insertion into membrane from inner side	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0032979	protein insertion into mitochondrial inner membrane from matrix	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0046493	Lipid A metabolic process	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0072695	regulation of DNA recombination at telomere	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1900044	regulation of protein K63-linked ubiquitination	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0035871	protein K11-linked deubiquitination	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:1900045	negative regulation of protein K63-linked ubiquitination	1	1	0.04	0.0433
Bivalvia	Group 1 + Group 2	GO:0048239	negative regulation of DNA recombination at telomere	1	1	0.04	0.0433
Drosophila	Group 1 + Group 2	GO:0045132	meiotic chromosome segregation	64	11	2.63	0.0015
Drosophila	Group 1 + Group 2	GO:0008119	sister chromatid segregation	140	11	5.75	0.0293
Drosophila	Group 1 + Group 2	GO:007192	chromosome organization involved in meiotic cell cycle	54	9	2.22	0.0085
Drosophila	Group 1 + Group 2	GO:007131	reciprocal meiotic recombination	37	7	1.52	0.0007
Drosophila	Group 1 + Group 2	GO:0007143	female meiotic nuclear division	54	6	2.22	0.0227
Drosophila	Group 1 + Group 2	GO:0058967	cellular response to topologically incorrect protein	44	5	1.81	0.0333
Drosophila	Group 1 + Group 2	GO:0058666	response to topologically incorrect protein	47	5	1.93	0.0427
Drosophila	Group 1 + Group 2	GO:007141	male meiosis I	13	4	0.53	0.0015
Drosophila	Group 1 + Group 2	GO:0140543	positive regulation of piRNA transcription	3	3	0.12	0.0001
Drosophila	Group 1 + Group 2	GO:0010526	retrotransposon silencing	8	3	0.33	0.0033
Drosophila	Group 1 + Group 2	GO:0067330	synaptonemal complex assembly	10	3	0.41	0.0067
Drosophila	Group 1 + Group 2	GO:0030719	P granule organization	11	3	0.45	0.0089
Drosophila	Group 1 + Group 2	GO:0071218	cellular response to misfolded protein	12	3	0.49	0.0115
Drosophila	Group 1 + Group 2	GO:0051788	response to misfolded protein	12	3	0.49	0.0115
Drosophila	Group 1 + Group 2	GO:0071315	meiosis II	15	3	0.62	0.0217
Drosophila	Group 1 + Group 2	GO:0034508	centromere complex assembly	19	3	0.78	0.0409
Drosophila	Group 1 + Group 2	GO:0048136	male germ-line cyst formation	2	2	0.08	0.0017
Drosophila	Group 1 + Group 2	GO:0061964	negative regulation of entry into reproductive diapause	5	2	0.21	0.0155
Drosophila	Group 1 + Group 2	GO:0051382	kinetochore assembly	5	2	0.21	0.0155
Drosophila	Group 1 + Group 2	GO:0051116	entry into reproductive diapause	6	2	0.25	0.0226
Drosophila	Group 1 + Group 2	GO:0071712	ER-associated misfolded protein catabolic process	6	2	0.25	0.0226
Drosophila	Group 1 + Group 2	GO:0061963	regulation of entry into reproductive diapause	6	2	0.25	0.0226
Drosophila	Group 1 + Group 2	GO:0043984	histone H4-K16 acetylation	6	2	0.25	0.0226
Drosophila	Group 1 + Group 2	GO:0055115	entry into diapause	7	2	0.29	0.0308
Drosophila	Group 1 + Group 2	GO:1900834	response to odorant	7	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0051177	meiotic sister chromatid cohesion	8	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0042795	snRNA transcription by RNA polymerase II	8	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0022611	dormancy process	8	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0069301	snRNA transcription	8	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0051144	meiotic sister chromatid segregation	8	2	0.33	0.0400
Drosophila	Group 1 + Group 2	GO:0001015	snRNA transcription by RNA polymerase II	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0010778	meiotic DNA repair synthesis involved in reciprocal meiotic recombination	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0072265	centromere localization	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0093932	sn(s)RNA transcription	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0051308	male meiosis chromosome separation	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0088653	centromere clustering	1	1	0.04	0.0411
Drosophila	Group 1 + Group 2	GO:0051415	microtubule nucleation by interphase microtubule organizing center	1	1	0.04	0.0411
Mammalia	Group 1 + Group 2	GO:0069595	immune response	1,297	145	48.02	0.0006
Mammalia	Group 1 + Group 2	GO:0058342	defense response to other organism	853	112	31.58	0.0207
Mammalia	Group 1 + Group 2	GO:0050587	innate immune response	647	82	23.95	0.0000
Mammalia	Group 1 + Group 2	GO:0001817	regulation of cytokine production	630	51	23.33	0.0466
Mammalia	Group 1 + Group 2	GO:0042742	defense response to bacterium	233	45	8.63	0.0000
Mammalia	Group 1 + Group 2	GO:0060554	inflammatory response	642	45	23.77	0.0174
Mammalia	Group 1 + Group 2	GO:0019221	cytokine-mediated signaling pathway	382	44	14.14	0.0000
Mammalia	Group 1 + Group 2	GO:0022520	adaptive immune response	342	44	12.66	0.0000
Mammalia	Group 1 + Group 2	GO:0001819	positive regulation of cytokine production	402	41	14.88	0.0272
Mammalia	Group 1 + Group 2	GO:0002697	regulation of immune effector process	308	37	11.40	0.0443
Mammalia	Group 1 + Group 2	GO:0042110	T cell activation	432	35	15.99	0.0256
Mammalia	Group 1 + Group 2	GO:0051607	defense response to virus	257	34	9.52	0.0000
Mammalia	Group 1 + Group 2	GO:008232	male gamete generation	491	32	18.18	0.0226

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1 + Group 2	GO:002920	regulation of humoral immune response	36	8	1.33	0.0051
Mammalia	Group 1 + Group 2	GO:0043303	mast cell degranulation	43	8	1.59	0.0055
Mammalia	Group 1 + Group 2	GO:002251	organ or tissue specific immune response	22	8	0.81	0.0065
Mammalia	Group 1 + Group 2	GO:002886	regulation of myeloid leukocyte mediated immunity	45	8	1.67	0.0078
Mammalia	Group 1 + Group 2	GO:0043030	regulation of macrophage activation	46	8	1.70	0.0092
Mammalia	Group 1 + Group 2	GO:002526	acute inflammatory response	83	8	3.07	0.0116
Mammalia	Group 1 + Group 2	GO:0032649	regulation of type II interferon production	88	8	3.26	0.0162
Mammalia	Group 1 + Group 2	GO:002609	type II interferon production	88	8	3.26	0.0162
Mammalia	Group 1 + Group 2	GO:0050691	regulation of defense response to virus by host	37	8	1.37	0.0498
Mammalia	Group 1 + Group 2	GO:0002367	natural killer cell mediated cytotoxicity	44	7	1.63	0.0011
Mammalia	Group 1 + Group 2	GO:0042447	lymphocyte chemotaxis	49	7	1.81	0.0020
Mammalia	Group 1 + Group 2	GO:0042119	neutrophil activation	32	7	1.18	0.0079
Mammalia	Group 1 + Group 2	GO:002945	negative regulation of mononuclear cell proliferation	63	7	2.33	0.0085
Mammalia	Group 1 + Group 2	GO:0027220	positive regulation of cytokine production involved in immune response	66	7	2.44	0.0108
Mammalia	Group 1 + Group 2	GO:005456	response to interferon-beta	26	7	0.96	0.0125
Mammalia	Group 1 + Group 2	GO:0050853	B cell receptor signaling pathway	45	7	1.67	0.0168
Mammalia	Group 1 + Group 2	GO:0069886	cellular response to virus	78	7	2.89	0.0252
Mammalia	Group 1 + Group 2	GO:0046425	regulation of receptor signaling pathway via JAK-STAT	78	7	2.89	0.0252
Mammalia	Group 1 + Group 2	GO:002385	mucoosal immune response	19	7	0.70	0.0325
Mammalia	Group 1 + Group 2	GO:0032755	positive regulation of interleukin-6 production	83	7	3.07	0.0340
Mammalia	Group 1 + Group 2	GO:1901222	regulation of non-canonical NF-kappaB signal transduction	89	7	3.30	0.0468
Mammalia	Group 1 + Group 2	GO:0048245	eosinophil chemotaxis	17	6	0.63	0.0000
Mammalia	Group 1 + Group 2	GO:0048240	sperm capacitation	25	6	0.93	0.0002
Mammalia	Group 1 + Group 2	GO:0070269	pyronotic inflammatory response	26	6	0.96	0.0003
Mammalia	Group 1 + Group 2	GO:0022320	positive regulation of defense response to virus by host	27	6	1.00	0.0004
Mammalia	Group 1 + Group 2	GO:0031295	T cell costimulation	27	6	1.00	0.0004
Mammalia	Group 1 + Group 2	GO:0028225	regulation of T-helper 1 type immune response	29	6	1.07	0.0006
Mammalia	Group 1 + Group 2	GO:0062691	regulation of cellular extravasation	38	6	1.41	0.0025
Mammalia	Group 1 + Group 2	GO:0140374	antiviral innate immune response	41	6	1.52	0.0037
Mammalia	Group 1 + Group 2	GO:0026359	positive regulation of immunoglobulin production	44	6	1.63	0.0054
Mammalia	Group 1 + Group 2	GO:0032731	positive regulation of interleukin-1 beta production	46	6	1.70	0.0067
Mammalia	Group 1 + Group 2	GO:1904994	positive regulation of receptor signaling pathway via STAT	46	6	1.70	0.0067
Mammalia	Group 1 + Group 2	GO:0029413	heterotypic cell-cell adhesion	52	6	1.74	0.0074
Mammalia	Group 1 + Group 2	GO:002548	monocyte chemotaxis	50	6	1.85	0.0100
Mammalia	Group 1 + Group 2	GO:001961	positive regulation of cytokine-mediated signaling pathway	50	6	1.85	0.0100
Mammalia	Group 1 + Group 2	GO:0030449	antiviral innate immune response	54	6	1.85	0.0100
Mammalia	Group 1 + Group 2	GO:0042531	positive regulation of immunoglobulin production	44	6	1.63	0.0054
Mammalia	Group 1 + Group 2	GO:0060337	positive regulation of interleukin-1 beta production	46	6	1.70	0.0067
Mammalia	Group 1 + Group 2	GO:0071357	positive regulation of receptor signaling pathway	46	6	1.70	0.0067
Mammalia	Group 1 + Group 2	GO:0034113	cellular response to type I interferon	52	6	1.74	0.0074
Mammalia	Group 1 + Group 2	GO:002548	regulation of T cell mediated cytotoxicity	50	6	1.85	0.0100
Mammalia	Group 1 + Group 2	GO:001961	establishment of spindle localization	54	6	1.85	0.0100
Mammalia	Group 1 + Group 2	GO:0051293	regulation of complement activation	23	6	0.67	0.0108
Mammalia	Group 1 + Group 2	GO:0051873	killing by host of symbiont cells	57	6	2.11	0.0184
Mammalia	Group 1 + Group 2	GO:0060760	positive regulation of tyrosine phosphorylation of STAT protein	51	6	1.89	0.0110
Mammalia	Group 1 + Group 2	GO:0150077	type I interferon-mediated signaling pathway	52	6	1.93	0.0121
Mammalia	Group 1 + Group 2	GO:0034340	cellular response to type I interferon	52	6	1.74	0.0074
Mammalia	Group 1 + Group 2	GO:0051653	spindle localization	58	6	2.15	0.0215
Mammalia	Group 1 + Group 2	GO:001914	retina homeostasis	59	6	2.18	0.0232
Mammalia	Group 1 + Group 2	GO:0051293	regulation of T cell mediated cytotoxicity	60	6	2.22	0.0232
Mammalia	Group 1 + Group 2	GO:0071260	establishment of spindle localization	60	6	2.22	0.0232
Mammalia	Group 1 + Group 2	GO:0050672	negative regulation of lymphocyte proliferation	62	6	2.30	0.0268
Mammalia	Group 1 + Group 2	GO:0032279	positive regulation of type II interferon production	62	6	2.30	0.0268
Mammalia	Group 1 + Group 2	GO:0042609	regulation of tyrosine phosphorylation of STAT protein	62	6	2.30	0.0268
Mammalia	Group 1 + Group 2	GO:0046596	tyrosine phosphorylation of STAT protein	66	6	2.44	0.0350
Mammalia	Group 1 + Group 2	GO:007340	innate immune response in mucosa	11	5	0.41	0.0060
Mammalia	Group 1 + Group 2	GO:0062830	positive regulation of type 2 immune response	14	5	0.52	0.0001
Mammalia	Group 1 + Group 2	GO:0053455	response to interferon-alpha	14	5	0.52	0.0001
Mammalia	Group 1 + Group 2	GO:0033005	positive regulation of mast cell activation	17	5	0.63	0.0003
Mammalia	Group 1 + Group 2	GO:0061760	antifungal innate immune response	17	5	0.63	0.0003
Mammalia	Group 1 + Group 2	GO:0035658	cellular response to interferon-beta	21	5	0.78	0.0009
Mammalia	Group 1 + Group 2	GO:0046596	acute-phase response	31	5	1.15	0.0052
Mammalia	Group 1 + Group 2	GO:0066053	canonic inflammasome complex assembly	32	5	1.18	0.0060
Mammalia	Group 1 + Group 2	GO:0032653	acute-phase response	34	5	1.26	0.0078
Mammalia	Group 1 + Group 2	GO:0032613	regulation of interleukin-10 production	35	5	1.30	0.0089
Mammalia	Group 1 + Group 2	GO:0043666	regulation of phosphoprotein phosphatase activity	46	5	1.70	0.0269

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1 + Group 2	GO:0032706	negative regulation of tumor necrosis factor production	47	5	1.74	0.0292
Mammalia	Group 1 + Group 2	GO:0035306	positive regulation of dephosphorylation	47	5	1.74	0.0292
Mammalia	Group 1 + Group 2	GO:1903556	negative regulation of tumor necrosis factor superfamily cytokine production	49	5	1.81	0.0342
Mammalia	Group 1 + Group 2	GO:0072683	regulation of lymphocyte apoptosis process	50	5	1.85	0.0369
Mammalia	Group 1 + Group 2	GO:0072683	T cell extravasation	13	4	0.48	0.0010
Mammalia	Group 1 + Group 2	GO:0027110	negative regulation of T cell mediated immunity	15	4	0.56	0.0018
Mammalia	Group 1 + Group 2	GO:0072540	T-helper 17 lineage commitment	17	4	0.63	0.0030
Mammalia	Group 1 + Group 2	GO:0045624	positive regulation of T-helper cell differentiation	17	4	0.63	0.0030
Mammalia	Group 1 + Group 2	GO:1903659	regulation of complement-dependent cytotoxicity	5	4	0.19	0.0040
Mammalia	Group 1 + Group 2	GO:0008228	opsonization	19	4	0.70	0.0046
Mammalia	Group 1 + Group 2	GO:0043302	positive regulation of leukocyte degranulation	20	4	0.74	0.0056
Mammalia	Group 1 + Group 2	GO:0001916	positive regulation of T cell mediated cytotoxicity	20	4	0.74	0.0056
Mammalia	Group 1 + Group 2	GO:0001562	response to protozoan infection	21	4	0.78	0.0067
Mammalia	Group 1 + Group 2	GO:0027117	positive regulation of natural killer cell mediated immunity	21	4	0.78	0.0067
Mammalia	Group 1 + Group 2	GO:1903901	negative regulation of viral life cycle	21	4	0.78	0.0067
Mammalia	Group 1 + Group 2	GO:0035821	regulation of process of another organism	21	4	0.78	0.0067
Mammalia	Group 1 + Group 2	GO:0003304	regulation of mast cell degranulation	23	4	0.85	0.0094
Mammalia	Group 1 + Group 2	GO:0032740	positive regulation of interleukin-17 production	23	4	0.85	0.0094
Mammalia	Group 1 + Group 2	GO:0043032	positive regulation of macrophage activation	23	4	0.85	0.0094
Mammalia	Group 1 + Group 2	GO:0070498	interleukin-1-mediated signaling pathway	24	4	0.89	0.0110
Mammalia	Group 1 + Group 2	GO:0010922	positive regulation of phosphatase activity	26	4	0.96	0.0145
Mammalia	Group 1 + Group 2	GO:0019884	antigen processing and presentation of exogenous antigen	28	4	1.04	0.0188
Mammalia	Group 1 + Group 2	GO:0002446	neutrophil mediated immunity	29	4	1.07	0.0212
Mammalia	Group 1 + Group 2	GO:0032743	positive regulation of interleukin-2 production	30	4	1.11	0.0238
Mammalia	Group 1 + Group 2	GO:2003052	negative regulation of endothelial cell apoptotic process	30	4	1.11	0.0238
Mammalia	Group 1 + Group 2	GO:1902255	negative regulation of NLRP3 inflammasome complex assembly	30	4	1.11	0.0238
Mammalia	Group 1 + Group 2	GO:0043330	response to exogenous dsRNA	31	4	1.15	0.0265
Mammalia	Group 1 + Group 2	GO:0032733	positive regulation of interleukin-10 production	32	4	1.18	0.0294
Mammalia	Group 1 + Group 2	GO:0046006	regulation of T cell proliferation	32	4	1.18	0.0294
Mammalia	Group 1 + Group 2	GO:0046006	regulation of natural killer cell activation	32	4	1.18	0.0294
Mammalia	Group 1 + Group 2	GO:0032814	NLRP3 inflammasome complex assembly	33	4	1.22	0.0325
Mammalia	Group 1 + Group 2	GO:0044546	activated T cell proliferation	34	4	1.26	0.0358
Mammalia	Group 1 + Group 2	GO:0050798	negative regulation of alpha-beta T cell activation	34	4	1.30	0.0393
Mammalia	Group 1 + Group 2	GO:0043331	response to dsRNA	35	4	1.33	0.0430
Mammalia	Group 1 + Group 2	GO:0052307	positive regulation of protein dephosphorylation	36	4	1.33	0.0430
Mammalia	Group 1 + Group 2	GO:0030890	regulation of natural killer cell activation	36	4	1.33	0.0430
Mammalia	Group 1 + Group 2	GO:0044546	FC-epsilon receptor signaling pathway	37	4	1.37	0.0468
Mammalia	Group 1 + Group 2	GO:0050798	negative regulation of complement activation, classical pathway	5	3	0.19	0.0005
Mammalia	Group 1 + Group 2	GO:0038156	interleukin-3-mediated signaling pathway	7	3	0.26	0.0016
Mammalia	Group 1 + Group 2	GO:0046636	cytosis by host of symbionts	8	3	0.30	0.0025
Mammalia	Group 1 + Group 2	GO:0005727	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	8	3	0.30	0.0025
Mammalia	Group 1 + Group 2	GO:0038095	necrotic signaling pathway	8	3	0.37	0.0050
Mammalia	Group 1 + Group 2	GO:0045959	activation-induced cell death of T cells	10	3	0.37	0.0050
Mammalia	Group 1 + Group 2	GO:0006624	regulation of neutrophil activation	11	3	0.41	0.0067
Mammalia	Group 1 + Group 2	GO:1902563	interleukin-15-mediated signaling pathway	11	3	0.41	0.0067
Mammalia	Group 1 + Group 2	GO:0010526	retrotransposon silencing	10	3	0.37	0.0050
Mammalia	Group 1 + Group 2	GO:0051838	negative regulation of neuroinflammatory response	12	3	0.44	0.0087
Mammalia	Group 1 + Group 2	GO:0006977	regulation of acrosome reaction	13	3	0.48	0.0109
Mammalia	Group 1 + Group 2	GO:0032754	natural killer cell activation involved in immune response	13	3	0.48	0.0109
Mammalia	Group 1 + Group 2	GO:2003051	regulation of T helper 2 cell cytokine production	13	3	0.48	0.0109
Mammalia	Group 1 + Group 2	GO:0006923	positive regulation of platelet aggregation	13	3	0.48	0.0109
Mammalia	Group 1 + Group 2	GO:0032736	complement activation, alternative pathway	14	3	0.52	0.0135
Mammalia	Group 1 + Group 2	GO:0043306	positive regulation of interleukin-13 production	14	3	0.52	0.0135
Mammalia	Group 1 + Group 2	GO:1903027	regulation of mast cell degranulation	14	3	0.52	0.0135
Mammalia	Group 1 + Group 2	GO:0023035	CD40 signaling pathway	14	3	0.52	0.0135
Mammalia	Group 1 + Group 2	GO:0033008	positive regulation of mast cell activation involved in immune response	14	3	0.52	0.0135
Mammalia	Group 1 + Group 2	GO:0045064	T-helper 2 cell differentiation	15	3	0.56	0.0165
Mammalia	Group 1 + Group 2	GO:0043031	negative regulation of macrophage activation	15	3	0.56	0.0165
Mammalia	Group 1 + Group 2	GO:0023730	regulation of dendritic cell cytokine production	16	3	0.59	0.0197
Mammalia	Group 1 + Group 2	GO:0032516	positive regulation of phosphoprotein phosphatase activity	16	3	0.59	0.0197
Mammalia	Group 1 + Group 2	GO:0046597	negative regulation of viral entry into host cell	16	3	0.59	0.0197
Mammalia	Group 1 + Group 2	GO:0043045	epigenetic programming of gene expression	16	3	0.59	0.0197
Mammalia	Group 1 + Group 2	GO:002371	dendritic cell cytokine production	16	3	0.59	0.0197
Mammalia	Group 1 + Group 2	GO:1901538	changes to DNA methylation involved in embryo development	16	3	0.59	0.0197

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1 + Group 2	GO:0002888	positive regulation of myeloid leukocyte mediated immunity	17	3	0.63	0.0233
Mammalia	Group 1 + Group 2	GO:0002544	chronic inflammatory response	17	3	0.63	0.0233
Mammalia	Group 1 + Group 2	GO:0002693	positive regulation of cellular extravasation	17	3	0.63	0.0233
Mammalia	Group 1 + Group 2	GO:0002827	positive regulation of T-helper 1 type immune response	18	3	0.67	0.0272
Mammalia	Group 1 + Group 2	GO:0045346	regulation of MHC class II biosynthetic process	18	3	0.67	0.0272
Mammalia	Group 1 + Group 2	GO:0002832	defense response to protozoan	19	3	0.70	0.0315
Mammalia	Group 1 + Group 2	GO:0045342	MHC class II biosynthetic process	19	3	0.70	0.0315
Mammalia	Group 1 + Group 2	GO:0040019	positive regulation of embryonic development	19	3	0.70	0.0315
Mammalia	Group 1 + Group 2	GO:0002640	positive regulation of type I interferon-mediated signaling pathway	19	3	0.70	0.0315
Mammalia	Group 1 + Group 2	CO:0060333	type II interferon-mediated signaling pathway	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0032753	positive regulation of interleukin-4 production	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0045063	T-helper 1 cell differentiation	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0042104	positive regulation of activated T cell proliferation	21	3	0.78	0.0409
Mammalia	Group 1 + Group 2	GO:0002726	positive regulation of T cell cytokine production	22	3	0.81	0.0461
Mammalia	Group 1 + Group 2	GO:1903660	negative regulation of complement-dependent cytotoxicity	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0057528	execution phase of necrosis	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0030887	positive regulation of myeloid dendritic cell activation	2	2	0.07	0.0014
Mammalia	Group 1 + Group 2	GO:0006069	response to symbiotic bacterium	2	2	0.15	0.0078
Mammalia	Group 1 + Group 2	GO:0048006	antigen processing and presentation, endogenous lipid antigen via MHC class Ib	4	2	0.15	0.0078
Mammalia	Group 1 + Group 2	GO:0050859	antigen processing and presentation, exogenous lipid antigen via MHC class Ib	4	2	0.15	0.0078
Mammalia	Group 1 + Group 2	GO:0002726	negative regulation of B cell receptor signaling pathway	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	CO:002765	immune response-inhibiting signal transduction	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:0036015	response to interleukin-3	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:0036016	cellular response to interleukin-3	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:20040416	regulation of eosinophil migration	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:0036364	positive regulation of cell-cell adhesion mediated by integrin	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:0045630	positive regulation of T helper 2 cell differentiation	5	2	0.19	0.0127
Mammalia	Group 1 + Group 2	GO:0043313	regulation of neutrophil degranulation	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:003341	positive regulation of peptidyl-serine phosphorylation of STAT protein	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0028286	negative regulation of T type I immune response	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0166300	protein-DNA covalent cross-linking repair	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0045341	MHC class I biosynthetic process	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0043307	regulation of MHC class I biosynthetic process	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0040077	antigen processing and presentation, exogenous lipid antigen via MHC class Ib	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0006332	positive regulation of type II interferon	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0060235	positive regulation of type II interferon-mediated signaling pathway	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	CO:2000559	regulation of interleukin-1-mediated signaling pathway	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0060456	positive regulation of necrotic process	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0043307	eosinophil activation	6	2	0.22	0.0186
Mammalia	Group 1 + Group 2	GO:0002733	regulation of myeloid dendritic cell cytokine production	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0048007	cellular response to type III interferon	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0002735	positive regulation of myeloid dendritic cell cytokine production	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:2000391	positive regulation of neutrophil extravasation	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:00010940	myeloid dendritic cell cytokine production	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0060467	positive regulation of necrotic cell death	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0060468	negative regulation of necrotic process	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0002733	prevention of polyspermy	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0071358	regulation of myeloid dendritic cell cytokine production	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0002735	positive regulation of neutrophil extravasation	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:2000391	myeloid dendritic cell cytokine production	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0062100	positive regulation of programmed necrotic cell death	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0017181	neutrophil apoptosis process	7	2	0.26	0.0254
Mammalia	Group 1 + Group 2	GO:0043312	neutrophil degranulation	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0042796	snRNA transcription by RNA polymerase III	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0043432	response to type III interferon	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:1903896	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:2000389	regulation of neutrophil extravasation	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:002674	negative regulation of acute inflammatory response	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0026538	negative regulation of immunoglobulin production	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0032650	regulation of interleukin-1 alpha production	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0032610	interleukin-4-mediated signaling pathway	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0057711	negative regulation of single stranded viral RNA replication via double stranded DNA intermediate	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0045869	cellular response to interleukin-2	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0071352	positive regulation of T-helper 2 cell cytokine production	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	CO:2000553	interleukin-2-mediated signaling pathway	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0038110	negative regulation of mast cell activation	8	2	0.30	0.0330
Mammalia	Group 1 + Group 2	GO:0032004	response to interleukin-2	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0070669	microglial cell proliferation	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0001518	regulation of peptidyl-serine phosphorylation of STAT protein	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0033139	adhesion of symbiont to host	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0044406					

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1 + Group 2	GO:0010918	positive regulation of mitochondrial membrane potential	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0036322	regulation of cell-cell adhesion mediated by integrin	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0046255	regulation of T-1-helper 1 cell differentiation	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0046258	regulation of T1-helper 1 cell differentiation	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:1961625	cellular response to ergosterol	9	2	0.33	0.0415
Mammalia	Group 1 + Group 2	GO:0003497	regulation of protein heterodimerization activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:19012569	negative regulation of Janus kinase activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0071660	positive regulation of IP-10 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0044555	clearance of foreign intracellular DNA	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0071652	regulation of chemoattractant (C-C motif) ligand 1 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0071654	positive regulation of chemoattractant (C-C motif) ligand 1 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0002736	regulation of plasmacytoid dendritic cell cytokine production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0022327	negative regulation of plasmacytoid dendritic cell cytokine production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0071610	chemokine (C-C motif) ligand 1 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0051073	disruption of plasma membrane integrity in another organism	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0150073	regulation of protein-glutamine gamma-glutamyltransferase activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0150074	positive regulation of protein-glutamine gamma-glutamyltransferase activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032759	positive regulation of TRAIL production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1961511	negative regulation of voltage-gated sodium channel activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0160006	Fc receptor-mediated immune complex endocytosis	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032723	positive regulation of connective tissue growth factor production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1905154	negative regulation of membrane invagination	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032745	positive regulation of interleukin-21 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0016068	type I hypersensitivity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1904784	NLRP1 inflammasome complex assembly	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0055397	helper T cell enhancement of adaptive immune response	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1905223	epicardium morphogenesis	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0050092	leukocyte adhesive activation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0002518	lymphocyte chemotaxis across high endothelial venules	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0034156	negative regulation of toll-like receptor 7 signaling pathway	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1904848	negative regulation of fibroblast growth factor	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0120042	negative regulation of macrophage proliferation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0034125	negative regulation of MyD88-dependent toll-like receptor signaling pathway	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0000937	cytoskeletal rearrangement involved in phagocytosis, engulfment	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:2000422	regulation of eosinophil chemotaxis	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:2000424	positive regulation of eosinophil chemotaxis	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0051973	positive regulation of protein homodimerization activity	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0051977	lysophospholipid transport	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032665	regulation of interleukin-21 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032679	regulation of TRAIL production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0032839	interleukin-21 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0001971	negative regulation of activation of membrane attack complex	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0006496	regulation of translational initiation by eIF2 alpha dephosphorylation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0036497	eIF2alpha dephosphorylation in response to endoplasmic reticulum stress	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:2000229	regulation of pancreatic stellate cell proliferation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0051041	positive regulation of calcium-independent cell adhesion	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:2000231	positive regulation of pancreatic stellate cell proliferation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1900450	negative regulation of glutamatergic receptor signaling pathway	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0018003	peptidyl-lysine N6-acetylation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1903916	regulation of endoplasmic reticulum stress-induced eIF2 alpha dephosphorylation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1903917	positive regulation of endoplasmic reticulum stress-induced eIF2 alpha dephosphorylation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:005046	clearance of foreign intracellular nucleic acids	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0150140	regulation of CD86 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0150142	positive regulation of CD86 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0150143	regulation of CD80 production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0036153	triacyceride acyl-chain remodeling	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0036155	acylglycerol acyl-chain remodeling	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1905675	negative regulation of adaptive immune memory response	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:000320	regulation of chylomicron remnant clearance	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:000321	positive regulation of chylomicron remnant clearance	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0140121	Lewy body formation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0140122	regulation of Lewy body formation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0140123	negative regulation of Lewy body formation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1902310	positive regulation of pepstatin-serine dephosphorylation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:2005045	negative regulation of endothelial cell chemotaxis to fibroblast growth factor	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0060101	negative regulation of phagocytosis, engulfment	1	1	0.04	0.0370

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1 + Group 2	GO:0070246	natural killer cell apoptotic process	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0070247	regulation of natural killer cell apoptotic process	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1901251	positive regulation of lung goblet cell differentiation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1901247	negative regulation of lung ciliated cell differentiation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:1901249	regulation of lung goblet cell differentiation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0030186	melatonin metabolic process	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0030187	melatonin biosynthetic process	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0072343	pancreatic stellate cell proliferation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0010185	regulation of cellular defense response	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0010186	positive regulation of cellular defense response	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0023844	hepatocyte immune response	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:19013496	response to 11-deoxycorticosterone	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0002373	plasmacytoid dendritic cell cytokine production	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:008784	biofilm matrix organization	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:008786	biofilm matrix disassembly	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0002450	B cell antigen processing and presentation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0046603	negative regulation of mitotic centrosome separation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:002470	plasmacytoid dendritic cell antigen processing and presentation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0072139	glomerular parietal epithelial cell differentiation	1	1	0.04	0.0370
Mammalia	Group 1 + Group 2	GO:0002090	regulation of primary metabolic process	673	47	23.83	0.0231
Mammalia	Group 1 + Group 2	GO:0019219	regulation of nucleobase-containing compound metabolic process	541	36	19.16	0.0270
Mammalia	Group 1 + Group 2	GO:0081252	regulation of nitrogen compound metabolic process	517	32	18.31	0.0342
Mammalia	Group 1 + Group 2	GO:000950	cellular response to stress	370	29	13.10	0.0270
Mammalia	Group 1 + Group 2	GO:003554	cellular response to stress	275	21	9.74	0.0419
Bivalvia	Group 1	GO:0051172	negative regulation of nitrogen compound metabolic process	117	15	4.14	0.0075
Bivalvia	Group 1	GO:0031326	positive regulation of cellular metabolic process	125	15	4.43	0.0100
Bivalvia	Group 1	GO:0066310	DNA recombination	66	13	2.34	0.0010
Bivalvia	Group 1	GO:0051173	positive regulation of nitrogen compound metabolic process	137	13	4.85	0.0336
Bivalvia	Group 1	GO:0010629	negative regulation of gene expression	78	11	2.76	0.0034
Bivalvia	Group 1	GO:0048513	animal organ development	83	11	2.94	0.0172
Bivalvia	Group 1	GO:0045334	negative regulation of nucleobase-containing compound metabolic process	64	11	2.27	0.0186
Bivalvia	Group 1	GO:006915	apoptotic processes	95	9	3.36	0.0062
Bivalvia	Group 1	GO:0045892	negative regulation of DNA-templated transcription	59	9	2.09	0.0151
Bivalvia	Group 1	GO:1902679	negative regulation of RNA biosynthetic process	59	9	2.09	0.0151
Bivalvia	Group 1	GO:0001253	negative regulation of RNA metabolic process	61	9	2.16	0.0192
Bivalvia	Group 1	GO:0006417	regulation of translation	52	8	1.84	0.0004
Bivalvia	Group 1	GO:005126	regulation of cell cycle	75	8	2.66	0.039
Bivalvia	Group 1	GO:0065009	regulation of molecular function	114	8	4.04	0.0485
Bivalvia	Group 1	GO:0045893	positive regulation of DNA-templated transcription	67	7	2.37	0.0091
Bivalvia	Group 1	GO:0042981	regulation of apoptotic process	70	7	2.48	0.0115
Bivalvia	Group 1	GO:0043067	regulation of programmed cell death	72	7	2.55	0.0133
Bivalvia	Group 1	GO:000122	negative regulation of transcription by RNA polymerase II	31	5	1.10	0.0043
Bivalvia	Group 1	GO:0066402	mRNA catabolic process	35	5	1.24	0.0073
Bivalvia	Group 1	GO:0045944	positive regulation of transcription by RNA polymerase II	38	5	1.35	0.0103
Bivalvia	Group 1	GO:0071310	cellular response to organic substance	52	5	1.84	0.0359
Bivalvia	Group 1	GO:0069628	response to abiotic stimulus	53	5	1.88	0.0385
Bivalvia	Group 1	GO:0051248	negative regulation of protein metabolic process	55	5	1.95	0.0441
Bivalvia	Group 1	GO:0030155	regulation of cell adhesion	11	4	0.39	0.0004
Bivalvia	Group 1	GO:0048568	embryonic organ development	12	4	0.42	0.0006
Bivalvia	Group 1	GO:0001607	defense response to virus	13	4	0.46	0.0009
Bivalvia	Group 1	GO:0010869	regulation of double-strand break repair via homologous recombination	5	4	0.18	0.0036
Bivalvia	Group 1	GO:0000902	cell morphogenesis	31	4	1.10	0.0227
Bivalvia	Group 1	GO:0002080	nuclear division	38	4	1.35	0.0441
Bivalvia	Group 1	GO:0050769	positive regulation of neurogenesis	5	3	0.18	0.0044
Bivalvia	Group 1	GO:0001819	positive regulation of cytokine production	11	3	0.25	0.0014
Bivalvia	Group 1	GO:0007192	chromosome organization involved in meiotic cell cycle	11	3	0.39	0.0059
Bivalvia	Group 1	GO:0001501	skelatal system development	11	3	0.39	0.0059
Bivalvia	Group 1	GO:0067368	muscle organ development	13	3	0.46	0.0096
Bivalvia	Group 1	GO:0010894	determination of left/right symmetry	7	3	0.25	0.0014
Bivalvia	Group 1	GO:0002037	tissue homeostasis	10	3	0.35	0.0044
Bivalvia	Group 1	GO:008285	heart morphogenesis	10	3	0.35	0.0119
Bivalvia	Group 1	GO:0051093	negative regulation of cell population proliferation	14	3	0.50	0.0119
Bivalvia	Group 1	GO:0032981	NADH dehydrogenase complex assembly	14	3	0.60	0.0206
Bivalvia	Group 1	GO:0051345	positive regulation of hydrolase activity	17	3	0.60	0.0206

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1	GO:0140013	meiotic nuclear division	17	3	0.60	
Bivalvia	Group 1	GO:0065976	polysaccharide metabolic process	18	3	0.64	0.0241
Bivalvia	Group 1	GO:0048729	tissue morphogenesis	19	3	0.67	0.0279
Bivalvia	Group 1	GO:0035295	tube development	20	3	0.71	0.0320
Bivalvia	Group 1	GO:0042274	ribosomal small subunit biogenesis	21	3	0.74	0.0364
Bivalvia	Group 1	GO:002603	regulation of anatomical structure morphogenesis	23	3	0.81	0.0460
Bivalvia	Group 1	GO:0140053	mitochondrial gene expression	23	3	0.81	0.0460
Bivalvia	Group 1	GO:1965168	positive regulation of double-strand break repair via homologous recombination	2	2	0.07	0.0013
Bivalvia	Group 1	GO:0032488	Cdc42 protein signal transduction	2	2	0.07	0.0013
Bivalvia	Group 1	GO:002600	digestive system process	2	2	0.07	0.0013
Bivalvia	Group 1	GO:003232	negative regulation of actin filament bundle assembly	2	2	0.07	0.0013
Bivalvia	Group 1	GO:003617	mitochondrial cytochrome c oxidase assembly	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0065910	negative regulation of DNA recombination	3	2	0.11	0.0037
Bivalvia	Group 1	GO:2000179	positive regulation of neural precursor cell proliferation	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0061383	trabecula morphogenesis	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0051701	biological process involved in interaction with host	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0010833	telomere maintenance via telomere lengthening	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0070959	mitotic G2 DNA damage checkpoint signaling	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0006959	mitochondrial RNA metabolic process	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0026064	epithelial cell development	3	2	0.11	0.0037
Bivalvia	Group 1	GO:0010001	glial cell differentiation	4	2	0.14	0.0071
Bivalvia	Group 1	GO:0046620	regulation of organ growth	4	2	0.14	0.0071
Bivalvia	Group 1	GO:0003190	atrioventricular valve formation	4	2	0.14	0.0071
Bivalvia	Group 1	GO:008781	ncRNA transcription	4	2	0.14	0.0071
Bivalvia	Group 1	GO:0030514	negative regulation of BMP signaling pathway	4	2	0.14	0.0071
Bivalvia	Group 1	GO:0035023	Rho protein signal transduction	4	2	0.14	0.0071
Bivalvia	Group 1	GO:007416	synapse assembly	4	2	0.14	0.0071
Bivalvia	Group 1	GO:0061947	heart looping	5	2	0.18	0.0116
Bivalvia	Group 1	GO:0042026	protein refolding	5	2	0.18	0.0116
Bivalvia	Group 1	GO:003143	embryonic heart tube morphogenesis	5	2	0.18	0.0116
Bivalvia	Group 1	GO:0061371	determination of heart left/right asymmetry	5	2	0.18	0.0116
Bivalvia	Group 1	GO:0035050	embryonic heart tube development	5	2	0.18	0.0116
Bivalvia	Group 1	GO:007162	negative regulation of cell adhesion	5	2	0.18	0.0116
Bivalvia	Group 1	GO:0007129	homologous chromosome pairing at meiosis	5	2	0.18	0.0116
Bivalvia	Group 1	GO:0066360	transcription by RNA polymerase I	6	2	0.21	0.0170
Bivalvia	Group 1	GO:0001889	liver development	6	2	0.21	0.0170
Bivalvia	Group 1	GO:0045143	homologous chromosome segregation	6	2	0.21	0.0170
Bivalvia	Group 1	GO:0061008	hepatobiliary system development	7	2	0.25	0.0233
Bivalvia	Group 1	GO:0010212	response to ionizing radiation	7	2	0.25	0.0233
Bivalvia	Group 1	GO:0030490	maturation of SSU-RNA	7	2	0.25	0.0233
Bivalvia	Group 1	GO:0061448	connective tissue development	7	2	0.25	0.0233
Bivalvia	Group 1	GO:0048732	gland development	8	2	0.28	0.0304
Bivalvia	Group 1	GO:001832	kidney development	8	2	0.28	0.0304
Bivalvia	Group 1	GO:0045596	negative regulation of cell differentiation	8	2	0.28	0.0304
Bivalvia	Group 1	GO:000271	polysaccharide biosynthetic process	8	2	0.28	
Bivalvia	Group 1	GO:0015053	osification	8	2	0.28	
Bivalvia	Group 1	GO:0060490	meiotic chromosome segregation	8	2	0.28	
Bivalvia	Group 1	GO:005132	snRNA metabolic process	8	2	0.28	
Bivalvia	Group 1	GO:0016073	renal system development	9	2	0.32	
Bivalvia	Group 1	GO:0072001	epithelial tube morphogenesis	9	2	0.32	
Bivalvia	Group 1	GO:0060562	embryonic organ morphogenesis	9	2	0.32	
Bivalvia	Group 1	GO:0098662	negative regulation of phosphorylation	9	2	0.32	0.0381
Bivalvia	Group 1	GO:0042326	positive regulation of protein phosphorylation	10	2	0.35	0.0466
Bivalvia	Group 1	GO:0001934	regulation of cell morphogenesis	10	2	0.35	0.0466
Bivalvia	Group 1	GO:002604	Arp2/3 complex-mediated actin nucleation	10	2	0.35	0.0466
Bivalvia	Group 1	GO:0043134	epithelial structure maintenance	11	1	0.04	0.0381
Bivalvia	Group 1	GO:0010669	negative regulation of organ growth	11	1	0.04	0.0381
Bivalvia	Group 1	GO:0046621	mitochondrial mRNA polyadenylation	11	1	0.04	0.0381
Bivalvia	Group 1	GO:007222	regulation of transcription by RNA polymerase I	11	1	0.04	0.0354
Bivalvia	Group 1	GO:0061668	regulation of protein K63-linked ubiquitination	11	1	0.04	0.0354
Bivalvia	Group 1	GO:0048799	semaphorin-plexin signaling pathway involved in neuron projection guidance	11	1	0.04	0.0354
Bivalvia	Group 1	GO:1900446	adhesion of symbiont to host	11	1	0.04	0.0354
Bivalvia	Group 1	GO:003247	telomere maintenance in response to DNA damage	11	1	0.04	0.0354
Bivalvia	Group 1	GO:0061009	common bile duct development	11	1	0.04	0.0354
Bivalvia	Group 1	GO:0063556	regulation of transcription by RNA polymerase I	11	1	0.04	0.0354
Bivalvia	Group 1	GO:1900044	semaphorin-plexin signaling pathway involved in neuron projection guidance	11	1	0.04	0.0354
Bivalvia	Group 1	GO:1902285	negative regulation of protein K63-linked ubiquitination	11	1	0.04	0.0354
Bivalvia	Group 1	GO:1900045	semaphorin-plexin signaling pathway involved in axon guidance	11	1	0.04	0.0354
Bivalvia	Group 1	GO:1902287	semaphorin-plexin signaling pathway involved in axon guidance	11	1	0.04	0.0354

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1	GO:0080772	positive regulation of axonogenesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0070977	bone maturation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0031120	snRNA pseudouridine synthesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0031126	sno(s)RNA 3'-end processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0044650	adhesion of symbiont to host cell	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0060546	mitochondrial tRNA processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0016074	sno(s)RNA metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0032978	protein insertion into membrane from inner side	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0032979	protein insertion into mitochondrial inner membrane from matrix	1	1	0.04	0.0354
Bivalvia	Group 1	CO:1904862	inhibitory synapse assembly	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0072695	regulation of DNA recombination at telomere	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0022330	regulation of defense response to virus by host	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0068653	lipopolysaccharide metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0045943	positive regulation of transcription by RNA polymerase I	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0052572	response to host immune response	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0051274	beta-D-glucan biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0051278	fungal-type cell wall polysaccharide biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1902200	semaphorin-peptid signaling pathway involved in bone trabecula morphogenesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0086265	extrinsic apoptotic signaling pathway via death domain receptors	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0069301	snRNA transcription	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0015743	malate transport	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0032065	maintenance of protein location in cell cortex	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0032196	transposition	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0032197	retroposition	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0042149	cellular response to glucose starvation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0052173	response to defenses of other organism	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0060674	(1->3)-beta-D-glucan metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0060675	(1->3)-beta-D-glucan biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0071536	response to host	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0090534	presynapse assembly	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0054138	protein localization to synapse	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0093031	cisification involved in bone maturation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0071966	fungal-type cell wall polysaccharide metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0099172	presynapse organization	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0033979	box H/ACA sno(s)RNA metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	CO:1902914	regulation of protein polyubiquitination	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1902915	negative regulation of protein polyubiquitination	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0071947	protein deubiquitination involved in ubiquitin-dependent protein catabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0052012	regulation of defense response to virus by host	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0060691	protein localization to cell junction	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0064933	lipid A metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0017001	antibiotic catabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0071423	malate transmembrane transport	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1901269	lipooligosaccharide metabolic processes	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0070212	protein poly-ADP-ribosylation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0070213	protein localization to cell junction	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0001100	negative regulation of exit from mitosis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0048144	fibroblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0048145	regulation of fibroblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0048147	negative regulation of fibroblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1901271	lipopolysaccharide biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0035622	intrahepatic bile duct development	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0030533	beta-lactam antibiotic metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0030655	beta-lactam antibiotic catabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0052200	response to host defenses	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1900481	mRNA pseudouridine synthesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0069103	lipopolysaccharide biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0068634	regulation of muscle organ development	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0010526	retrotansposon silencing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0046402	O antigen metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0042783	evasion of host immune response	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0031314	sno(s)RNA processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0031848	protection from non-homologous end joining at telomere	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0061430	bone trabecula morphogenesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0020033	antigenic variation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:000963	mitochondrial RNA processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:007168	receptor guanyl cyclase signaling pathway	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0043628	regulatory ncRNA 3'-end processing	1	1	0.04	0.0354

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Bivalvia	Group 1	GO:0016999	antibiotic metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0072340	lactam catabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:1902041	regulation of extrinsic apoptotic signaling pathway via death domain receptors	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0009272	fungal-type cell wall biogenesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0072338	lactam metabolic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0052779	miRNA-mediated gene silencing by mRNA destabilization	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0034964	box H/ACA sno(s)RNA processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:000495	box H/ACA sno(s)RNA 3'-end processing	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0067140	male meiotic nuclear division	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0003382	epithelial cell morphogenesis	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0069243	O antigen biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0069245	lipid A biosynthetic process	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0033687	osteoblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0036868	regulation of osteoblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0036869	negative regulation of osteoblast proliferation	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0050277	maintenance of gastrointestinal epithelium	1	1	0.04	0.0354
Bivalvia	Group 1	GO:0068339	negative regulation of DNA recombination at telomere	1	1	0.04	0.0354
Drosophila	Group 1	GO:0051276	chromosome organization	331	18	11.07	0.0293
Drosophila	Group 1	GO:0045132	meiotic chromosome segregation	64	7	2.14	0.0053
Drosophila	Group 1	GO:0067131	reciprocal meiotic recombination	37	6	1.24	0.0013
Drosophila	Group 1	GO:0031146	SCF-d ⁻ dependent proteasomal ubiquitin-dependent protein catabolic process	57	5	1.91	0.0413
Drosophila	Group 1	GO:007141	male meiosis I	13	4	0.43	0.0007
Drosophila	Group 1	GO:1902275	regulation of chromatin organization	35	4	1.17	0.0285
Drosophila	Group 1	GO:0042078	germ-line stem cell division	36	4	1.20	0.0312
Drosophila	Group 1	GO:0010526	retrotransposon silencing	8	3	0.27	0.0018
Drosophila	Group 1	GO:0071218	cellular response to misfolded protein	12	3	0.40	0.0065
Drosophila	Group 1	GO:0034508	centromere complex assembly	19	3	0.64	0.0242
Drosophila	Group 1	GO:0067080	mitotic metaphase chromosome alignment	23	3	0.77	0.0400
Drosophila	Group 1	GO:0070720	microtubule nucleation	23	3	0.77	0.0400
Drosophila	Group 1	GO:0031445	regulation of heterochromatin formation	25	3	0.84	0.0495
Drosophila	Group 1	GO:0120261	regulation of heterochromatin organization	25	3	0.84	0.0495
Drosophila	Group 1	GO:0098136	male germ-line cyst formation	2	2	0.07	0.0011
Drosophila	Group 1	GO:0061964	negative regulation of entry into reproductive diapause	5	2	0.17	0.0104
Drosophila	Group 1	GO:0051382	kinetochore assembly	5	2	0.17	0.0104
Drosophila	Group 1	GO:0055116	entry into reproductive diapause	6	2	0.20	0.0153
Drosophila	Group 1	GO:0071712	ER-associated misfolded protein catabolic process	6	2	0.20	0.0153
Drosophila	Group 1	GO:0061963	regulation of entry into reproductive diapause	6	2	0.20	0.0153
Drosophila	Group 1	GO:0043984	histone H4-K16 acetylation	6	2	0.20	0.0153
Drosophila	Group 1	GO:0055115	entry into diapause	7	2	0.23	0.0210
Drosophila	Group 1	GO:1900834	response to odorant	7	2	0.23	0.0210
Drosophila	Group 1	GO:0042795	snRNA transcription by RNA polymerase II	8	2	0.27	0.0273
Drosophila	Group 1	GO:0022611	dormancy process	8	2	0.27	0.0273
Drosophila	Group 1	GO:0069301	snRNA transcription	8	2	0.27	0.0273
Drosophila	Group 1	GO:0071786	endoplasmic reticulum tubular network organization	9	2	0.30	0.0344
Drosophila	Group 1	GO:0030007	intracellular potassium ion homeostasis	9	2	0.30	0.0344
Drosophila	Group 1	GO:0051383	kinetochore organization	9	2	0.30	0.0344
Drosophila	Group 1	GO:0000220	regulation of meiotic nuclear division	9	2	0.30	0.0344
Drosophila	Group 1	GO:0043967	histone H4 acetylation	9	2	0.30	0.0344
Drosophila	Group 1	GO:0063376	sodium ion export across plasma membrane	9	2	0.30	0.0344
Drosophila	Group 1	GO:0066833	intracellular sodium ion homeostasis	9	2	0.30	0.0344
Drosophila	Group 1	GO:0001915	snRNA ⁺ transcription by RNA polymerase II	1	1	0.03	0.0335
Drosophila	Group 1	GO:0010778	meiotic DNA repair synthesis involved in reciprocal meiotic recombination	1	1	0.03	0.0335
Drosophila	Group 1	GO:0069302	sno(s)RNA transcription	1	1	0.03	0.0335
Drosophila	Group 1	GO:00151308	male meiosis chromosome separation	1	1	0.03	0.0335
Drosophila	Group 1	GO:0051415	microtubule nucleation by interphase microtubule organizing center	1	1	0.03	0.0335
Mammalia	Group 1	GO:0069355	immune response	1,297	120	38.75	0.0027
Mammalia	Group 1	GO:0045087	innate immune response	647	69	19.33	0.0000
Mammalia	Group 1	GO:0050778	positive regulation of immune response	419	47	12.52	0.0068
Mammalia	Group 1	GO:0002250	adaptive immune response	342	39	10.22	0.0000
Mammalia	Group 1	GO:0026944	regulation of leukocyte activation	496	39	13.62	0.0052
Mammalia	Group 1	GO:0061819	positive regulation of cytokine production	402	37	12.01	0.0025
Mammalia	Group 1	GO:0019221	cytokine-mediated signaling pathway	382	35	11.41	0.0000
Mammalia	Group 1	GO:0042142	defense response to bacterium	233	32	6.96	0.0004
Mammalia	Group 1	GO:0042110	T cell activation	432	31	12.91	0.0133
Mammalia	Group 1	GO:0062697	regulation of immune effector process	308	31	9.20	0.0188
Mammalia	Group 1	GO:0051607	defense response to virus	257	28	7.68	0.0000
Mammalia	Group 1	GO:0022744	myeloid leukocyte activation	195	26	5.83	0.0294
Mammalia	Group 1	GO:0002703	regulation of leukocyte mediated immunity	186	25	5.56	0.0136

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1	GO:0070661	leukocyte proliferation	273	24	8.16	0.0271
Mammalia	Group 1	GO:0051349	positive regulation of defense response	240	23	7.17	0.0057
Mammalia	Group 1	GO:1903131	mononuclear cell differentiation	392	23	11.71	0.0482
Mammalia	Group 1	GO:0050727	regulation of inflammatory response	300	21	8.96	0.0025
Mammalia	Group 1	GO:0069539	humoral immune response	177	21	5.29	0.0490
Mammalia	Group 1	GO:0002768	immune response-regulating cell surface receptor signaling pathway	177	20	5.29	0.0018
Mammalia	Group 1	GO:0001906	cell killing	153	20	4.57	0.0449
Mammalia	Group 1	GO:0026999	positive regulation of immune effector process	208	20	6.21	0.0477
Mammalia	Group 1	GO:0032017	T cell differentiation	255	18	7.62	0.0441
Mammalia	Group 1	GO:0050889	positive regulation of innate immune response	113	16	3.38	0.0139
Mammalia	Group 1	GO:0071222	cellular response to lipopolysaccharide	164	15	4.90	0.0001
Mammalia	Group 1	GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	137	15	4.09	0.0045
Mammalia	Group 1	GO:0010466	negative regulation of peptidase activity	163	14	4.87	0.0004
Mammalia	Group 1	GO:0002429	immune response-activating cell surface receptor signaling pathway	104	14	4.90	0.0052
Mammalia	Group 1	GO:0032650	regulation of tumor necrosis factor production	133	14	3.97	0.0142
Mammalia	Group 1	GO:0032640	tumor necrosis factor production	133	14	3.97	0.0142
Mammalia	Group 1	GO:0002705	positive regulation of leukocyte mediated immunity	113	14	3.38	0.0395
Mammalia	Group 1	GO:0050829	defense response to Gram-negative bacterium	66	13	1.97	0.0000
Mammalia	Group 1	GO:0002444	myeloid leukocyte mediated immunity	90	13	2.69	0.0001
Mammalia	Group 1	GO:0051091	positive regulation of DNA-binding transcription factor activity	227	13	6.78	0.0191
Mammalia	Group 1	GO:0002456	T cell mediated immunity	82	13	2.45	0.0395
Mammalia	Group 1	GO:0031341	regulation of cell killing	71	12	2.12	0.0042
Mammalia	Group 1	GO:0002695	negative regulation of leukocyte activation	148	12	4.42	0.0178
Mammalia	Group 1	GO:0002286	T cell activation involved in immune response	94	12	2.81	0.0194
Mammalia	Group 1	GO:008066	negative regulation of cell activation	165	12	4.93	0.0351
Mammalia	Group 1	GO:0001818	negative regulation of cytokine production	225	12	6.72	0.0381
Mammalia	Group 1	GO:0050830	defense response to Gram-positive bacterium	87	11	2.60	0.0001
Mammalia	Group 1	GO:0002275	myeloid cell activation involved in immune response	83	11	2.48	0.0010
Mammalia	Group 1	GO:0050909	sensory perception of taste	52	11	1.55	0.0289
Mammalia	Group 1	GO:0042102	positive regulation of T cell proliferation	85	10	2.54	0.0002
Mammalia	Group 1	GO:0043299	leukocyte degranulation	62	10	1.85	0.0005
Mammalia	Group 1	GO:0030101	natural killer cell activation	66	10	1.97	0.0009
Mammalia	Group 1	GO:007696	cell surface receptor signaling pathway via STAT	139	10	4.15	0.0089
Mammalia	Group 1	GO:0016064	immunoglobulin mediated immune response	107	10	3.20	0.0321
Mammalia	Group 1	GO:0019724	B cell mediated immunity	108	10	3.23	0.0336
Mammalia	Group 1	GO:0032760	positive regulation of tumor necrosis factor production	84	9	2.51	0.0009
Mammalia	Group 1	GO:0069536	complement activation	47	9	1.40	0.0100
Mammalia	Group 1	GO:0050729	positive regulation of inflammatory response	122	9	3.64	0.0109
Mammalia	Group 1	GO:0051250	negative regulation of lymphocyte activation	124	9	3.70	0.0121
Mammalia	Group 1	GO:0050832	defense response to fungi	48	9	1.43	0.0126
Mammalia	Group 1	GO:0002823	negative regulation of adaptive immune response [...]	39	9	1.17	0.0174
Mammalia	Group 1	GO:0007259	cell surface receptor signaling pathway via JAK-STAT	134	9	4.00	0.0192
Mammalia	Group 1	GO:0051092	positive regulation of NF-kappaB transcription factor activity	136	9	4.06	0.0209
Mammalia	Group 1	GO:0002820	negative regulation of adaptive immune response	42	9	1.25	0.0259
Mammalia	Group 1	GO:0042100	response to fungus	76	9	2.27	0.0261
Mammalia	Group 1	GO:009620	sperm egg recognition	58	9	1.73	0.0323
Mammalia	Group 1	GO:0050336	leukocyte apoptotic process	46	9	1.37	0.0400
Mammalia	Group 1	GO:0071887	fusion of sperm to egg plasma membrane involved in single fertilization	98	9	2.93	0.0468
Mammalia	Group 1	GO:0007342	activation of innate immune response	25	8	0.75	0.0000
Mammalia	Group 1	GO:0001580	regulation of type II interferon production	45	8	1.34	0.0000
Mammalia	Group 1	GO:0002218	detection of chemical stimulus involved in sensory perception of bitter taste	88	8	2.63	0.0048
Mammalia	Group 1	GO:0032649	activation of innate immune response	54	8	1.61	0.0063
Mammalia	Group 1	GO:0045576	regulation of type II interferon production	98	8	2.93	0.0090
Mammalia	Group 1	GO:0071346	masl cell activation	114	8	3.41	0.0209
Mammalia	Group 1	GO:0001959	cellular response to type II interferon	119	8	3.55	0.0263
Mammalia	Group 1	GO:0030317	regulation of cytokine-mediated signaling pathway	119	8	3.55	0.0263
Mammalia	Group 1	GO:0057722	flagellated sperm motility	120	8	3.58	0.0275
Mammalia	Group 1	GO:0068584	male gonad development	121	8	3.61	0.0287
Mammalia	Group 1	GO:0060759	development of primary male sexual characteristics	124	8	3.70	0.0325
Mammalia	Group 1	GO:006294	regulation of response to cytokine stimulus	128	8	3.82	0.0382
Mammalia	Group 1	GO:007339	cilium movement involved in cell motility	35	7	1.05	0.0001
Mammalia	Group 1	GO:0019731	binding of sperm to zona pellucida	40	7	1.19	0.0002
Mammalia	Group 1	GO:0032757	antibacterial humoral response	54	7	1.61	0.0011
Mammalia	Group 1	GO:0032222	positive regulation of interleukin-8 production	60	7	1.79	0.0020
Mammalia	Group 1	GO:0002251	organ or tissue specific immune response	22	7	0.66	0.0036
Mammalia	Group 1	GO:0085856	cellular response to virus	78	7	2.33	0.0086
Mammalia	Group 1	GO:0030593	neutrophil chemotaxis	80	7	2.39	0.0099

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1	GO:0043303	mast cell degranulation	43	7	1.28	0.0101
Mammalia	Group 1	GO:0062279	mast cell activation involved in immune response	44	7	1.31	0.0114
Mammalia	Group 1	GO:1904892	regulation of receptor signalling pathway via STAT	83	7	2.48	0.0120
Mammalia	Group 1	GO:0028866	regulation of myeloid leukocyte mediated immunity	45	7	1.34	0.0128
Mammalia	Group 1	GO:002448	mast cell mediated immunity	46	7	1.37	0.0143
Mammalia	Group 1	GO:0050691	regulation of defense response to virus by host	37	7	1.11	0.0336
Mammalia	Group 1	GO:005303	regulation of dephosphorylation	103	7	3.08	0.0346
Mammalia	Group 1	GO:002825	regulation of T-helper 1 type immune response	29	6	0.87	0.0002
Mammalia	Group 1	GO:0042119	neutrophil activation	32	6	0.96	0.0003
Mammalia	Group 1	GO:005071	negative regulation of viral genome replication	39	6	1.17	0.0010
Mammalia	Group 1	GO:0027111	positive regulation of T cell mediated immunity	44	6	1.31	0.0019
Mammalia	Group 1	GO:0026339	positive regulation of immunoglobulin production	44	6	1.31	0.0019
Mammalia	Group 1	GO:0042267	natural killer cell mediated cytotoxicity	44	6	1.31	0.0019
Mammalia	Group 1	GO:002731	positive regulation of interleukin-1 beta production	46	6	1.37	0.0023
Mammalia	Group 1	GO:006337	type I interferon-mediated signaling pathway	52	6	1.55	0.0044
Mammalia	Group 1	GO:0030888	regulation of B cell proliferation	54	6	1.61	0.0053
Mammalia	Group 1	GO:0051293	establishment of spindle localization	54	6	1.61	0.0053
Mammalia	Group 1	GO:0032729	positive regulation of type II interferon production	62	6	1.85	0.0103
Mammalia	Group 1	GO:0010921	regulation of phosphatase activity	65	6	1.94	0.0129
Mammalia	Group 1	GO:0062270	positive regulation of cytokine production involved in immune response	66	6	0.0138	
Mammalia	Group 1	GO:002385	mucosal immune response	19	6	0.57	0.0218
Mammalia	Group 1	GO:200106	regulation of leukocyte apoptotic process	25	5	0.0245	
Mammalia	Group 1	GO:0046425	regulation of receptor signalling pathway via JAK-STAT	78	6	2.33	
Mammalia	Group 1	GO:0032255	positive regulation of interleukin-6 production	83	6	2.48	0.0377
Mammalia	Group 1	GO:0061760	antifungal innate immune response	17	5	0.51	0.0001
Mammalia	Group 1	GO:0035458	cellular response to interferon-beta	21	5	0.63	
Mammalia	Group 1	GO:0048240	sperm capacitation	25	5	0.75	0.0008
Mammalia	Group 1	GO:0001914	regulation of T cell mediated cytotoxicity	25	5	0.75	
Mammalia	Group 1	GO:0002230	antiviral innate immune response	27	5	1.22	0.00245
Mammalia	Group 1	GO:0028330	positive regulation of defense response to virus by host	83	6	0.81	0.0011
Mammalia	Group 1	GO:0046596	positive regulation of type II immune response	14	5	0.42	0.0019
Mammalia	Group 1	GO:0061774	regulation of viral entry into host cell response	31	5	0.93	
Mammalia	Group 1	GO:0014632	microglial cell activation	35	5	1.05	0.0036
Mammalia	Group 1	GO:006953	canonical inflammasome complex assembly	35	5	1.05	0.0036
Mammalia	Group 1	GO:0140374	acute-phase response	38	5	1.14	0.0052
Mammalia	Group 1	GO:0001912	antiviral innate immune response	41	5	1.22	0.0072
Mammalia	Group 1	GO:0032720	positive regulation of leukocyte mediated cytotoxicity	42	5	1.25	0.0080
Mammalia	Group 1	GO:0041113	negative regulation of tumor necrosis factor production	47	5	1.40	0.0127
Mammalia	Group 1	GO:1903556	heterotypic cell-cell adhesion	47	5	1.40	0.0127
Mammalia	Group 1	GO:0070228	negative regulation of tumor necrosis factor superfamily cytokine production	49	5	1.46	0.0151
Mammalia	Group 1	GO:0019651	regulation of lymphocyte apoptotic process	50	5	1.49	0.0164
Mammalia	Group 1	GO:0060760	positive regulation of cytokine-mediated signaling pathway	50	5	1.49	0.0164
Mammalia	Group 1	GO:0038034	positive regulation of response to cytokine stimulus	57	5	1.70	0.0273
Mammalia	Group 1	GO:0051792	signal transduction in absence of ligand	59	5	1.76	0.0311
Mammalia	Group 1	GO:0071260	extrinsic apoptotic signaling pathway in absence of ligand	59	5	1.76	0.0311
Mammalia	Group 1	GO:0031640	cellular response to mechanical stimulus	60	5	1.79	
Mammalia	Group 1	GO:0061844	killing of cells of another organism	64	5	1.91	0.0421
Mammalia	Group 1	GO:0002227	antimicrobial humoral immune response mediated by antimicrobial peptide	64	5	1.91	0.0421
Mammalia	Group 1	GO:0056155	innate immune response in mucosa	11	4	0.33	
Mammalia	Group 1	GO:0048245	pyroptotic inflammatory response	14	4	0.42	0.0006
Mammalia	Group 1	GO:0031295	T cell costimulation	26	4	0.78	
Mammalia	Group 1	GO:0019884	T-helper 17 cell lineage commitment	17	4	0.51	0.0014
Mammalia	Group 1	GO:0062177	positive regulation of natural killer cell mediated immunity	21	4	0.63	0.0031
Mammalia	Group 1	GO:0032740	neutrophil mediated immunity	29	4	0.87	0.0103
Mammalia	Group 1	GO:0030352	negative regulation of endothelial cell apoptotic process	30	4	0.90	0.0116
Mammalia	Group 1	GO:0032743	positive regulation of interleukin-2 production	23	4	0.69	0.0044
Mammalia	Group 1	GO:1900225	regulation of macrophage activation	30	4	0.90	0.0116
Mammalia	Group 1	GO:0043330	interleukin-1-mediated signaling pathway	24	4	0.72	0.0052
Mammalia	Group 1	GO:0046006	response to exogenous dsRNA	31	4	0.93	0.0131
Mammalia	Group 1	GO:0032814	regulation of activated T cell proliferation	32	4	0.96	0.0146
Mammalia	Group 1	GO:0044546	regulation of natural killer cell activation	32	4	0.96	0.0146
Mammalia	Group 1	GO:0050798	NLRP3 inflammasome complex assembly	33	4	0.99	0.0162
Mammalia	Group 1	GO:0066968	activated T cell proliferation	34	4	1.02	0.0180
Mammalia	Group 1		cellular defense response	34	4	1.02	0.0180

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1	GO:0002701	negative regulation of production of molecular mediator of immune response	35	4	1.05	
Mammalia	Group 1	GO:0043331	response to dsRNA	36	4	1.08	0.0198
Mammalia	Group 1	GO:0026951	regulation of cellular extravasation	38	4	1.14	0.0218
Mammalia	Group 1	GO:2000351	regulation of endothelial cell apoptotic process	43	4	1.28	0.0260
Mammalia	Group 1	GO:0032653	regulation of interleukin-10 production	45	4	1.34	0.0388
Mammalia	Group 1	GO:0032613	interleukin-10 production	45	4	1.34	0.0447
Mammalia	Group 1	GO:0046244	positive regulation of T-helper cell differentiation	17	4	0.51	0.0479
Mammalia	Group 1	GO:0043666	regulation of phosphoprotein phosphatase activity	46	4	1.37	0.0479
Mammalia	Group 1	GO:1904694	positive regulation of receptor signaling pathway via STAT	46	4	1.37	0.0479
Mammalia	Group 1	GO:0038095	Fc-epsilon receptor signalling pathway	10	3	0.30	0.0027
Mammalia	Group 1	GO:0045959	negative regulation of complement activation, classical pathway	10	3	0.30	0.0027
Mammalia	Group 1	GO:0051838	cytosis by host of symbiont cells	5	3	0.15	0.0003
Mammalia	Group 1	GO:0067527	necrotic signalling pathway	8	3	0.24	0.0013
Mammalia	Group 1	GO:006924	activation-induced cell death of T cells	8	3	0.24	0.0013
Mammalia	Group 1	GO:1901731	positive regulation of platelet aggregation	10	3	0.30	0.0027
Mammalia	Group 1	GO:2000551	regulation of T-helper 2 cell cytokine production	10	3	0.30	0.0027
Mammalia	Group 1	GO:0007543	egg activation	11	3	0.33	0.0037
Mammalia	Group 1	GO:00515079	interleukin-15-mediated signalling pathway	11	3	0.33	0.0037
Mammalia	Group 1	GO:0150079	negative regulation of neuroinflammatory response	12	3	0.36	0.0048
Mammalia	Group 1	GO:0002233	natural killer cell activation involved in immune response	13	3	0.39	0.0061
Mammalia	Group 1	GO:0032197	retrotransposition	13	3	0.39	0.0061
Mammalia	Group 1	GO:0060046	regulation of acrosome reaction	13	3	0.39	0.0061
Mammalia	Group 1	GO:0072683	T cell extravasation	13	3	0.39	0.0061
Mammalia	Group 1	GO:0043306	positive regulation of mast cell degranulation	14	3	0.42	0.0075
Mammalia	Group 1	GO:0060557	complement activation, alternative pathway	14	3	0.42	0.0075
Mammalia	Group 1	GO:0050835	CD40 signalling pathway	14	3	0.42	0.0075
Mammalia	Group 1	GO:0023035	regulation of epoxidation	14	3	0.42	0.0075
Mammalia	Group 1	GO:1903027	negative regulation of T cell mediated immunity	14	3	0.42	0.0075
Mammalia	Group 1	GO:0002710	negative regulation of macrophage activation	15	3	0.45	0.0092
Mammalia	Group 1	GO:0043031	negative regulation of viral entry into host cell	15	3	0.45	0.0092
Mammalia	Group 1	GO:0046597	positive regulation of myeloid leukocyte mediated immunity	16	3	0.48	0.0111
Mammalia	Group 1	GO:0028988	positive regulation of T-helper 1 type immune response	17	3	0.51	0.0132
Mammalia	Group 1	GO:0002827	regulation of MHC class II biosynthetic process	18	3	0.54	0.0155
Mammalia	Group 1	GO:0045346	defense response to protozoan	18	3	0.54	0.0155
Mammalia	Group 1	GO:0042832	positive regulation of embryonic development	19	3	0.57	0.0180
Mammalia	Group 1	GO:0000019	positive regulation of type I interferon-mediated signalling pathway	19	3	0.57	0.0180
Mammalia	Group 1	GO:0006340	type II interferon-mediated signalling pathway	19	3	0.57	0.0180
Mammalia	Group 1	GO:0006333	MHC class II biosynthetic process	19	3	0.57	0.0180
Mammalia	Group 1	GO:0045342	positive regulation of interleukin-4 production	20	3	0.60	0.0207
Mammalia	Group 1	GO:0032753	positive regulation of T cell mediated cytotoxicity	20	3	0.60	0.0207
Mammalia	Group 1	GO:0001916	positive regulation of interleukin-4 production	21	3	0.63	0.0236
Mammalia	Group 1	GO:0001562	positive regulation of viral life cycle	21	3	0.63	0.0236
Mammalia	Group 1	GO:0006332	modulation of process of another organism	21	3	0.63	0.0236
Mammalia	Group 1	GO:0022726	positive regulation of T cell cytolytic production	22	3	0.66	0.0268
Mammalia	Group 1	GO:0042104	positive regulation of activated T cell proliferation	22	3	0.66	0.0268
Mammalia	Group 1	GO:0010743	regulation of macrophage derived foam cell differentiation	25	3	0.75	0.0374
Mammalia	Group 1	GO:0032673	regulation of natural killer cell mediated cytotoxicity	25	3	0.75	0.0374
Mammalia	Group 1	GO:0032633	execution phase of necrosis	25	3	0.75	0.0374
Mammalia	Group 1	GO:0071354	cellular response to interleukin-6	26	3	0.78	0.0414
Mammalia	Group 1	GO:0010922	positive regulation of phosphatase activity	26	3	0.78	0.0414
Mammalia	Group 1	GO:1902019	regulation of cilium-dependent cell motility	28	3	0.84	0.0500
Mammalia	Group 1	GO:0070741	response to interleukin-6	4	2	0.12	0.0051
Mammalia	Group 1	GO:0007451	antigen processing and presentation, endogenous lipid antigen via MHC class Ib	4	2	0.12	0.0051
Mammalia	Group 1	GO:0062765	immune response-inhibiting signal transduction	5	2	0.15	0.0084
Mammalia	Group 1	GO:0045630	regulation of natural killer cell mediated cytotoxicity	5	2	0.15	0.0084
Mammalia	Group 1	GO:0043313	regulation of neutrophil degranulation	6	2	0.18	0.0123
Mammalia	Group 1	GO:0106300	protein-DNA covalent cross-linking repair	6	2	0.18	0.0123
Mammalia	Group 1	GO:0060545	positive regulation of necrototic process	6	2	0.18	0.0123
Mammalia	Group 1	GO:0028226	negative regulation of T-helper 1 type immune response	6	2	0.18	0.0123
Mammalia	Group 1	GO:003141	positive regulation of peptidyl-serine phosphorylation of STAT protein	6	2	0.18	0.0123
Mammalia	Group 1	GO:0060332	positive regulation of response to type II interferon	6	2	0.18	0.0123

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
	Mammalia	Group 1	GO:0060335 positive regulation of type II interferon-mediated signaling pathway	6	2	0.18	0.0123
	Mammalia	Group 1	GO:0048007 antigen processing and presentation, exogenous lipid antigen via MHC class Ib	6	2	0.18	0.0123
	Mammalia	Group 1	GO:2006559 regulation of interleukin-1-mediated signaling pathway	6	2	0.18	0.0123
	Mammalia	Group 1	GO:0045341 MHC class I biosynthetic process	6	2	0.18	0.0123
	Mammalia	Group 1	GO:0045243 regulation of MHC class I biosynthetic process	6	2	0.18	0.0123
	Mammalia	Group 1	GO:0001781 neutrophil apoptotic process	6	2	0.18	0.0169
	Mammalia	Group 1	GO:0002372 myeloid dendritic cell cytokine production	7	2	0.21	0.0169
	Mammalia	Group 1	GO:002733 regulation of myeloid dendritic cell cytokine production	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0062735 positive regulation of myeloid dendritic cell cytokine production	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0038196 type III interferon-mediated signaling pathway	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0060467 negative regulation of fertilization	7	2	0.21	0.0169
	Mammalia	Group 1	GO:000468 prevention of polysemes	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0062100 positive regulation of programmed, necrotic cell death	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0011358 cellular response to type III interferon	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0010940 positive regulation of necrotic cell death	7	2	0.21	0.0169
	Mammalia	Group 1	GO:0042796 snRNA transcription by RNA polymerase III	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0064312 neutrophil degranulation	8	2	0.24	0.0221
	Mammalia	Group 1	GO:1905669 positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0069977 DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0038110 interleukin-2-mediated signaling pathway	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0026388 negative regulation of immunoglobulin production	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0032650 regulation of interleukin-1 alpha production	8	2	0.24	0.0221
	Mammalia	Group 1	GO:005771 interleukin-4-mediated signaling pathway	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0026110 interleukin-1 alpha production	8	2	0.24	0.0221
	Mammalia	Group 1	GO:001352 cellular response to interleukin-2	8	2	0.24	0.0221
	Mammalia	Group 1	GO:2006533 positive regulation of T-helper 2 cell cytokine production	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0034342 response to type III interferon	8	2	0.24	0.0221
	Mammalia	Group 1	GO:0044006 adhesion of symbiont to host	9	2	0.27	0.0279
	Mammalia	Group 1	GO:005625 regulation of interleukin-1 alpha production	9	2	0.27	0.0279
	Mammalia	Group 1	GO:0032319 regulation of peptidyl-serine phosphorylation of STAT protein	9	2	0.27	0.0279
	Mammalia	Group 1	GO:0061518 microglial cell proliferation	9	2	0.27	0.0279
	Mammalia	Group 1	GO:0070669 response to interleukin-2	9	2	0.27	0.0279
	Mammalia	Group 1	GO:0010918 positive regulation of mitochondrial membrane potential	9	2	0.27	0.0279
	Mammalia	Group 1	GO:1902563 regulation of neutrophil activation	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0062430 adhesion of symbiont to host	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0064625 complement receptor mediated signaling pathway	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0032754 regulation of interleukin-5 production	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0010526 retrotransposon silencing	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0010528 regulation of transposition	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0010529 negative regulation of transposition	10	2	0.30	0.0342
	Mammalia	Group 1	GO:1902563 regulation of neutrophil activation	10	2	0.30	0.0342
	Mammalia	Group 1	GO:0180078 positive regulation of neuroinflammatory response	11	2	0.33	0.0410
	Mammalia	Group 1	GO:0007011 cellular response to granulocyte macrophage colony-stimulating factor stimulus	11	2	0.33	0.0410
	Mammalia	Group 1	GO:0067012 response to granulocyte macrophage colony-stimulating factor	11	2	0.33	0.0410
	Mammalia	Group 1	GO:002732 positive regulation of dendritic cell cytokine production	11	2	0.33	0.0410
	Mammalia	Group 1	GO:0032376 negative regulation of interleukin-13 production	11	2	0.33	0.0410
	Mammalia	Group 1	GO:0018387 positive regulation of cellular respiration	11	2	0.33	0.0410
	Mammalia	Group 1	GO:0061517 macrophage proliferation	12	2	0.36	0.0482
	Mammalia	Group 1	GO:1904226 negative regulation of NLRP3 inflammasome complex assembly	12	2	0.36	0.0482
	Mammalia	Group 1	GO:0051770 positive regulation of nitric-oxide synthase biosynthetic process	12	2	0.36	0.0482
	Mammalia	Group 1	GO:0023036 regulation of prostaglandin secretion	12	2	0.36	0.0482
	Mammalia	Group 1	GO:0032308 positive regulation of prostaglandin secretion	12	2	0.36	0.0482
	Mammalia	Group 1	GO:004250 B cell antigen processing and presentation	1	1	0.03	0.0299
	Mammalia	Group 1	GO:002450 plasmacytoid dendritic cell proliferation	1	1	0.03	0.0299
	Mammalia	Group 1	GO:0072343 pancreatic stellate cell proliferation	1	1	0.03	0.0299
	Mammalia	Group 1	GO:00185 regulation of cellular defense response	1	1	0.03	0.0299
	Mammalia	Group 1	GO:0010186 positive regulation of cellular defense response	1	1	0.03	0.0299
	Mammalia	Group 1	GO:008784 beta1 integrin matrix disassembly	1	1	0.03	0.0299
	Mammalia	Group 1	GO:008786 type I hyperactivity	1	1	0.03	0.0299
	Mammalia	Group 1	GO:0055397 hepatocyte enhancer of adaptive immune response	1	1	0.03	0.0299
	Mammalia	Group 1	GO:0072139 glomerular parietal epithelial cell differentiation	1	1	0.03	0.0299
	Mammalia	Group 1	GO:0140121 Lewy body formation	1	1	0.03	0.0299

Tab. S10 continued from previous page

Dataset	Group of genes	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Mammalia	Group 1	GO:0140122	regulation of Lewy body formation	1	1	0.03	
Mammalia	Group 1	GO:0140123	negative regulation of Lewy body formation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0030186	melatonin metabolic process	1	1	0.03	0.0299
Mammalia	Group 1	GO:0030187	melatonin biosynthetic process	1	1	0.03	0.0299
Mammalia	Group 1	GO:0071660	positive regulation of IP-10 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0060997	cytoskeletal rearrangement involved in phagocytosis, engulfment	1	1	0.03	0.0299
Mammalia	Group 1	GO:0051041	positive regulation of calcium-independent cell-cell adhesion	1	1	0.03	0.0299
Mammalia	Group 1	GO:2000422	regulation of eosinophil chemotaxis	1	1	0.03	0.0299
Mammalia	Group 1	GO:2000424	positive regulation of eosinophil chemotaxis	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032759	positive regulation of TRAIL production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032273	positive regulation of connective tissue growth factor production	1	1	0.03	0.0299
Mammalia	Group 1	GO:190251	positive regulation of lung goblet cell differentiation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032245	positive regulation of interleukin-21 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0051977	lysophospholipid transport	1	1	0.03	0.0299
Mammalia	Group 1	GO:1901247	negative regulation of lung ciliated cell differentiation	1	1	0.03	0.0299
Mammalia	Group 1	GO:1901249	regulation of lung goblet cell differentiation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0070246	natural killer cell apoptotic process	1	1	0.03	0.0299
Mammalia	Group 1	GO:0070247	regulation of natural killer cell apoptotic process	1	1	0.03	0.0299
Mammalia	Group 1	GO:1902310	positive regulation of peptidyl-serine dephosphorylation	1	1	0.03	0.0299
Mammalia	Group 1	GO:2000229	regulation of pancreatic stellate cell proliferation	1	1	0.03	0.0299
Mammalia	Group 1	GO:1904784	NLRP1 inflammasome complex assembly	1	1	0.03	0.0299
Mammalia	Group 1	GO:2000231	positive regulation of pancreatic stellate cell proliferation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0150140	regulation of CD86 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0150142	positive regulation of CD86 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0150143	regulation of CD80 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032665	positive regulation of CD80 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032679	regulation of interleukin-21 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032625	regulation of TRAIL production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0032639	interleukin-21 production	1	1	0.03	0.0299
Mammalia	Group 1	GO:0050320	regulation of chylomicron remnant clearance	1	1	0.03	0.0299
Mammalia	Group 1	GO:0060321	positive regulation of chylomicron remnant clearance	1	1	0.03	0.0299
Mammalia	Group 1	GO:1900450	negative regulation of glutamate receptor signaling pathway	1	1	0.03	0.0299
Mammalia	Group 1	GO:0120042	negative regulation of macrophage proliferation	1	1	0.03	0.0299
Mammalia	Group 1	GO:1903916	regulation of endoplasmic reticulum stress-induced eIF2 alpha dephosphorylation	1	1	0.03	0.0299
Mammalia	Group 1	GO:1903917	positive regulation of endoplasmic reticulum stress-induced eIF2 alpha dephosphorylation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0036496	regulation of translational initiation by eIF2 alpha dephosphorylation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0036497	eIF2alpha dephosphorylation in response to endoplasmic reticulum stress	1	1	0.03	0.0299
Mammalia	Group 1	GO:1901625	cellular response to ergosterol	1	1	0.03	0.0299
Mammalia	Group 1	GO:0100006	Fc receptor-mediated immune complex endocytosis	1	1	0.03	0.0299
Mammalia	Group 1	GO:0036153	triglyceride acyl-chain remodeling	1	1	0.03	0.0299
Mammalia	Group 1	GO:0036155	acylglycerol acyl-chain remodeling	1	1	0.03	0.0299
Mammalia	Group 1	GO:0046603	negative regulation of mitotic centrosome separation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0050902	leukocyte adhesive activation	1	1	0.03	0.0299
Mammalia	Group 1	GO:0034156	negative regulation of toll-like receptor 7 signaling pathway	1	1	0.03	0.0299
Mammalia	Group 1	GO:1905151	negative regulation of voltage-gated sodium channel activity	1	1	0.03	0.0299

Supplementary Table S11 – Accession numbers and general statistics of RNA-sequencing libraries of *Mytilus galloprovincialis* developmental stages.

SRA acc. no.	Time point (hpf)	Developmental stage	Biological replicate	No. of trimmed reads	No. of uniquely mapped reads	% of uniquely mapping reads	No. of multi-mapped reads	% of multi-mapped reads	No. of unmapped reads	% of unmapped reads
SRR25.387458	0	Unfertilized Egg	R2	31,614,834	22,677,680	71.73	4,525,340	14.31	4,411,814	13.96
SRR25.387459	0	Unfertilized Egg	R1	31,229,282	22,930,909	73.43	3,631,800	11.63	4,666,573	14.94
SRR25.387436	4	Zygote	R2	31,573,942	22,238,537	70.43	4,747,330	15.04	4,588,075	14.53
SRR25.387447	4	Zygote	R1	31,215,045	22,657,884	72.59	3,980,706	12.75	4,576,485	14.66
SRR25.387434	8	Embryo	R2	31,679,197	22,361,698	70.59	5,224,621	16.49	4,092,878	12.92
SRR25.387435	8	Embryo	R1	31,137,192	22,231,464	71.40	4,558,377	14.63	4,347,351	13.96
SRR25.387432	12	Gastrula 1	R2	30,684,472	21,477,819	70.00	5,351,506	17.44	3,855,147	12.57
SRR25.387433	12	Gastrula 1	R1	31,006,745	22,158,048	71.46	4,808,473	15.51	4,040,224	13.03
SRR25.387430	16	Gastrula 2	R2	31,129,558	21,172,217	68.01	5,943,113	19.10	4,014,228	12.89
SRR25.387431	16	Gastrula 2	R1	31,108,790	22,366,338	71.90	4,797,011	15.42	3,945,441	12.68
SRR25.387456	20	Trochophore 1	R2	31,313,006	21,923,726	70.01	5,234,475	16.71	4,154,805	13.27
SRR25.387457	20	Trochophore 1	R1	31,153,366	22,818,864	73.25	4,258,642	13.67	4,075,860	13.08
SRR25.387454	24	Trochophore 2	R2	31,237,678	21,993,262	70.41	5,095,540	16.32	4,148,876	13.29
SRR25.387455	24	Trochophore 2	R1	30,568,998	22,102,616	72.30	4,376,490	14.31	4,089,892	13.38
SRR25.387452	28	Trochophore 3	R2	31,251,736	22,218,911	71.10	4,737,140	15.16	4,295,685	13.74
SRR25.387453	28	Trochophore 3	R1	30,151,120	21,916,806	72.69	4,155,633	13.78	4,078,681	13.53
SRR25.387450	32	Advanced Trochophore 1	R2	31,450,472	22,501,615	71.55	4,808,549	15.29	4,140,308	13.16
SRR25.387451	32	Advanced Trochophore 1	R1	30,344,616	22,168,913	73.06	4,088,100	13.47	4,087,603	13.47
SRR25.387448	36	Advanced Trochophore 2	R2	31,281,625	22,138,801	70.77	5,099,606	16.30	4,043,218	12.92
SRR25.387449	36	Advanced Trochophore 2	R1	30,446,461	22,282,330	73.18	4,164,126	13.68	4,000,205	13.13
SRR25.387445	40	Advanced Trochophore 3	R2	31,286,823	22,253,466	71.13	4,808,187	15.37	4,225,170	13.50
SRR25.387446	40	Advanced Trochophore 3	R1	30,427,358	22,021,178	72.37	4,139,754	13.61	4,266,426	14.02
SRR25.387443	44	D-veliger 1	R2	30,469,689	21,902,055	71.88	4,591,610	15.06	3,976,024	13.05
SRR25.387444	44	D-veliger 1	R1	30,486,877	22,346,398	73.30	4,187,438	13.74	3,953,041	12.96
SRR25.387441	48	D-veliger 2	R2	30,209,384	21,152,507	70.02	5,122,804	16.95	3,934,073	13.03
SRR25.387442	48	D-veliger 2	R1	30,492,559	22,356,764	73.32	4,132,649	13.55	4,003,146	13.12
SRR25.387439	52	D-veliger 3	R2	30,581,805	22,287,023	72.88	4,273,235	13.98	4,021,547	13.15
SRR25.387440	52	D-veliger 3	R1	30,340,062	22,508,657	73.70	3,938,157	12.90	4,093,248	13.40
SRR25.387437	72	Late D-veliger	R2	31,144,538	22,754,899	73.06	4,155,579	13.35	4,234,060	13.60
SRR25.387438	72	Late D-veliger	R1	31,504,212	23,356,674	74.14	3,830,231	12.16	4,317,307	13.70

Supplementary Table S12 – Set of HCR probes generated with the ‘insitu_probe_generator’ script from the Ozpolat Lab (Kuehn et al., 2022).

Pool name	Sequence
B1_Mgal_10B017427_vasa_33_Dla0	GAGGGAGGGCAGCAAACGaaAAATGCTGTAAAAACAAAACGTATA TTAAAATGTTGAAAATGATTTAATAtaGAAGAGTCTTCCTTTACG GAGGAGGGCACCAAAACGaaAGAGTTCATCACGATTAAAGAAAGA ACTTTAAAACATTACATCATACAGAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGCACATGATTTAAGGTAGCACTACT AAACAAATAAGAACCTGGATTTGATTtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGACGAACCTTCTCCAGACAGGCCG AAAAAAATATGCCGCTTCCAATCAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaTTCAGCTGTTGGTTCACTTGGTAAA CCAATAGTGATTACCTAACTGAACAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGCAAGACATGGCTGGTTATAGTT ATTATGCCCTAATGGTTTAGTCGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaACATTCCCCTGTAAGTCGTTAGTGA GTGACAACGTGTTGATTATATGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCATTGAAGAATTATCTGCTGAGAGT TGCAACCCACTTTAGTCGAAACAAAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGCATTCAATTAATGAATGGAGT ACAATTCCCTGCTTCCTCGCAGGTTAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaAGAAATGTGATTAATCCAGGATTCC ATTTACAATGGTTGAAATAACATGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaTTGTCCTTGGTTGGCTTTCTGAT CTCCGGAATCTCTGCTGTTGCTGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCTGAATCATCTGTTGAGAAACTGGT GTAACAAAGCTTGGCAATTCCACCTaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaACATGTTAACAGGAAATGCTA ATAGATTGTTGAAAGCTATAATTAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGCAAGGAATGTCAGGATTTCTTCT GGAATTCAGACTGAGACAAGTAAGGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaATGCACATACTGCAACATCGGTA TTGACGTTCTGATCACGGGGAACTtGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaTTGCACTGAACATAAGAGTTGCT CAAGTCTCTGGATTCTCTGGGAAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCTTCCCTTACCAATGACATCAATT TAAATATTTCAGCTTCTAAACTGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCTAGGTTAACATGTGCTGAGTCT GAGGAGGGCAGCAAACGaaACATGTTAACAGGAAATGCTA ACTAAAGCTGTGCTCTGAACAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGACTTTACACATTAAAGAAGG TGTACAGGTGTTGCTATCATAGGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaTGCCTGTTGCTGTTAAATTTC TTTTCTATATTGTCAGGTTATGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGGGCCATATGCTCTTCTTAC AAAGACCTCTGTTAGGACAATCTaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCACTTCAATTGAGCAGTT CATTAGGACACTCTCTGCTGCTGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCCTTGCCTCCACCTGATCTC CCACTTCCACACATTGAGCAGTTGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaCCTCCACCTTATTCCTCAAATC CCCCCTTGCTGCCAATCCACCTtGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaAGATATGCTCATGCTTTAC AGGCTTCCCTCAAAGCAGGGGTTaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaTCCTTGAGCACGGCCTACTGGT TCTTCCAAACCCACACGCCCCAAAtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaATTGTAATGTTAGGTGTTGTT ATAACAATTCTGCTCCAGGACTtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaACTTGAAATTCTGTAACACACGT AACTAACTAACATGCAAATTCTGtaGAAGAGTCTTCCTTTACG GAGGAGGGCAGCAAACGaaGTGAAAACAAACGAATGTTAAATT AAAAATATGGTCACATGAACCTCAAtaGAAGAGTCTTCCTTTACG

Table S12 continued from previous page

Pool name	Sequence
B1_Mgal_10B017427_vasa_33_Dla0	GAGGAGGGCAGCAAACGGaaTTGAAACTTTAGAACATATTATA ACAGTGTAAACATTTCTGCATGCCtaGAAGAGTCCTCCCTTACG GAGGAGGGCAGCAAACGGaaGTCAGTTATCTCTTTCAACAG CTCTACTGTGTCATGTGAATAttaGAAGAGTCCTCCCTTACG GAGGAGGGCAGCAAACGGaaCAGACAGTCCCTTGCCCTTCTC TGTGTAGTCCCTTCAACCTTTCgtGAAGAGTCCTCCCTTACG
B2_Mgal_10B093608_dmrt1l_32_Dla0	CCTCGTAAATCCTCATCaaaaaaaaggggggtatttttaaa AAAAAATATGTCTGATTTCAattaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaACAAAGCTCAATATGATATATC taaccCTTCAGATATACTTGAGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAAAGTAACACAGctaaaaattaaga CAGTTGAACTTTATTCACCTTaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaattatacaattatacatTGACTG ttataccTCCCTCATCTGTATCCaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAAATCACCCCTGGTGGACAAA ACACATATTCAACCTTATTTTGGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaATGAATGTTGATAATCAAAGTACT CTATGATTCAATCAAGTATCATTAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaatacatgatatatgggATATGAAGT TCAGACAAAGCATTGTATCttgaaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTAACTTGTGCAAACATCCCTT TGACCTGGACTTTGCAAACCAAATAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaacattttttttttttttttttttt ACGGGCACCAACATGTCATTTcaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaATTTCAGACTCACATTCTATGT attt CCTCGTAAATCCTCATCaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAAATTCAGACTCACATTCTATGT ttt TTCAATGATCAAAAGTTCTATAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaCATGCAGACAGACACTCTAAACAT tggatgtAAAAAATCAGACTTCaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTCATTATTATCTCAAAGGTGTGC TCACCTCACCCCAACATAAGCAAGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTTTGAATTGTTGCTGtatttagtaat CTATAATTATTTACCCCTGAAATTGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTTTCTCTCTTATGACAACATATCTT aatcattttttagaaCATCAAAATAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaactttttttttttttttttttttt TCTTGATTTAACATTATcatacaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAAATTTCTACAAATTtgcacatgtat actt CCTCGTAAATCCTCATCaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAAATGAAATATGCACTATCaaataa ttttaattggtcTTAAGATAATTGGaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTTCTCAAATATGGTGCCTTAAATCC tagtaatgttaATTGTAGTTGtaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaATGCATGACTACAGAGATTTAACT GAAAGTTCTAACAGACTCAATTAAAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaAGCCACGACTACAAAATAAGTTTC AAAAACGTTGCTTCTATCTTGtaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaGTACAGTTGAATGTGGAAATATCATG CTTGTGCTAGCTGGAGTCCATaaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaATTGTACGGCATCTTGTGTCGCA TGGTCCTGCATATCTGGGGTaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTGTAAACAGTTGGTGGATAAAATCCA TGATGGTGGGTTATGGTGTGGATAAGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaATTATTCTACACTCTGTGACCAAC TGGCTACACACAGAGGACTGACAATAaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaCTGGCACTAGATTTTAACACCATG CTCTGTATGGACAATCTCTGTGaaATCATCCAGTAAACCGCC CCTCGTAAATCCTCATCaaaTCTGTGAGATAATTGGTTTGTGT TTGACATCTACGGCAATATACAGCTaaATCATCCAGTAAACCGCC

Table S12 continued from previous page

Pool name	Sequence
B2_Mgal_10B093608_dmrt1l_32_Dla0	CCTCGTAATCCTCATCAaaacaaattttttacattttccattTC TTCCCTTCCATTTCGCTgttgtaaATCATCCAGTAAACGCC CCTCGTAATCCTCATCAaaCTGGCGTTAAACTAACAAACCTGA TTTTGAGCAGTAACATCTTGTaaATCATCCAGTAAACGCC CCTCGTAATCCTCATCAaaATCTAACATGGACACCATGAGCAAG AAAACATCCAAGCTTGACCTTaaATCATCCAGTAAACGCC CCTCGTAATCCTCATCAaaTGGCGGAATTGACGTTACTTCA aaattacAACTTTCTCTCTTaaATCATCCAGTAAACGCC
B3_Mgal_10B014180_soxh_27_Dla0	GTCCTGCCTCTATATCTttCTTTCTTGAGACATCCACATT ATTTAGTTATCGCTCTCCCTTTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttCTTGAGACATCCACATTAGTTAT TTATCGTCTCTCCCTTTCTTTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttCATCCACATTAGTTATCGTCTT TTCTCCCTTTCTCTCTGAtttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttATTTAGTTATCGTCTCTCCCTT CTTTCTCTCTCTGAGACATCCtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttATATGCTCTACTAAAGACCTAGAGA TTTCTCTACAGAAGTAGGTACAAtttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttTTCTCTCTGAGACATCCAC CAACATTAGTTATCGTCTCTCCtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttTTCTCTGAGACATCCACATTAGT TAGTTATGCTCTCCCTTCCtCCACTCAACTTAAACCG GTCCTGCCTCTATCTtGAGACATCCACATTAGTTATCGT TCGCTCTCCCTTCTCTCTGAtttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttACATGTCGCCAGTGGAGATTCTG CAATGGTATCATCTTGGACCTCtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGTCTCTGATGTCCTCTCAGC TGCTGGTGGGAACGGGAACGGCAAtttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGCAAGCTCTGAATCATCGTAG GTCGTACTTCCGGAAAGAGTGGTAGttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttCTGCTTCTTGATTTCTTGAAGA TGCTGTGTCTAAAGTGTGCTATGtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttCTTCTAGCTTTATCCTCTACTGG TTTTCTTGTCTCTGCTTCTTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGGGTGCTTGTATCTTCTGACAT GTCAGTATTGTCATAAGCAGACGCTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGCATTACAGGCTTGTCTTCAATG TCATTTGTCTTCTTCTGTTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGACCACAGGGCTTACCCCTCGGA TGCTGGAACAGCTTAAAGTCTCTCtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGTATTATTGTGACATTGTACGG GGTATTGAAACAGGAACTGGTTGGtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttATCTGGAAACGTCTTGAAGAAG TGAGTCAGGTATTCTGCTGTTGtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGAATGGCTGTCCTGACATGGGATAT ATTTCatTCAGAAATTCTTGtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGGACAGACAAATTGGTACCTTGTAG ATTGGTGAACGATTCTGAGTCGTTtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttCATTACGGGTAATTGGTCTGTT GCATCGTTACAGGTTGAATAGGCAtttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttATCCCAGTACCTGGCATGAGAGTAG GAACCTTGACAGGTTCTATACGtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGTCTTCTGACATTCTGAGGGTT tCCATCTGCTGACATAAGAAAATtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttGCATTGCAAGTATGGCATTTGT GCTTGAATGGTGTGCAACATTCAACTCAACTTAAACCG GTCCTGCCTCTATATCTttTTTAACCTGTTCAACTCCGTC CAGGTTGGGTGAATGACATACAAActtCCACTCAACTTAAACCG GTCCTGCCTCTATATCTtttttattatgcaaAGTCCATTTC TTCTGTGTTAAATTAGccgaatttttCCACTCAACTTAAACCG GTCCTGCCTCTATATCTttTTATCACATGAAACATGTTGTGAAA CCCGCGATAAAAGTTTGCTGATTtCCACTCAACTTAAACCG

Table S12 continued from previous page

Pool name	Sequence
B4_Mgal_10B094018_foxI2_28_Dla0	CCTCAACCTACCTCCAA Caaagaaaaaaaataacaataaatataat TTGCATGGTAAGAAATTGccttaaatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTAATTTCTCTGTAGGCTCCATAA ttaaaatataatgaaaaattctCGATatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCATTAGGTATGATAATATCTCA TTTTCTGGCACAGAATCGTGACACatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTCCAGTAGGTATAATGACCGGCTC ACTGTATGCCAATTATTACCTCTCatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaACTGGACCCATCGAACACTGTAGTCG TTGTTGCTGTAAGGAAAGGAAAAtTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTTCATGGCATGGCTGATTGCGATTG AACTGGGAAATGTTGGTCAGGAtatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCCGGTGCGGAACACGGGACTGGTT TATAACTACAGGCCGCTAAACTGGGAtTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTGGAGATGCAGAATTGGTGCAGA ACTGtataactgttatttgtaCAtTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTACGGTGGCGTGA AAAATACGGTT CACGGCGAATATTGTAACTGAGAtTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaACTGTCCGTCACAAATAATGGTT TTAAAGAAAATTGTTGTATGGAGAtatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaACGCAATTCTCGTGCAGCTGGTAG GAGAGATATCGATGCTGGTATGGAtatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaGGGTCTAAAGTCCAATAATTCCCTT CCTTCTCAAACATGTCCTCAAATGatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTTAACAAACATTCACTAGACTTAG GTTCTCCCCACCTCTCTGGGACatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCTTATTTCATAGTAGGGAAT GTGAGGAATACTATTGGCATCTTatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaACTGTTATCTTATCATCGATT TTCATGATATATTGATAAATTCCACatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaAAGATAATGGTGTTGACATCGGG TAATTGCCATAGCAAATAGAGCAACatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTTGTTCTCTTCTGGTGTGTTT TTTTCACCGGTTATTCACAGATTatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCATTCTTGTACTATTCTCCAT TCTTATCTCAAAGTAGTTATTatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTTAACATGTTAGGAATTGAGGT TGACGGATTCCACATTTCGCTAatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCTGAATTGGGATTCGAGAGGGT CACATTTCGAGTCATAAAGTTTatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCCGTCAACTGTcatctaaatttg GCTCTAAATTAAATCCAAATCTGAatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaataatattagAATTTCACAAACAGC tatttcatgtttccatTTGtaatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCATGATTAACTTTtttgtgaa TCATAATGAAGTTATAatTGTCTTAtTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTAACCTCTTCAATTGTTGCAATT CTCAGGTTTAGTGAATTCAAATCatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTAGTCACCAAATGTTATTCATGG TTTCTCTCTTTAATTTCGTCatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCACTTCTCAGGATCTTAAACCTT GTCTTTCTTACTAAAGGTGACTTatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaTCAAACCTATTGGCTGATCTGTAT TGTCTGATACACCAGACAACTCTAatTCTCACCATATTGCGTTC CCTCAACCTACCTCCAA CaaCGTTGCTTGTCTTCTTGTGG CGTTGTTGGTGTATTTCGATTatTCTCACCATATTGCGTTC
B4_Mgal_10B094018_foxI2_28_Dla0	

Supplementary Table S13 – Set of Vasa/Ddx4 and Ddx3 sequences used in the phylogenetic analysis. For each sequence, the species, the gene accession number (Gene ID), the orthology group and the gene names for reference species are shown. Reference species are marked with an asterisk.

Species	Gene ID	Orthology group	Gene name
<i>Phreagena okutanii</i>	DN44424.c0.g1.i2.p1	Vasa/Ddx4	-
<i>Calyptogena (Archivesica) marissinica</i>	Ama38729	Vasa/Ddx4	-
<i>Calyptogena (Archivesica) marissinica</i>	Ama38727	Vasa/Ddx4	-
<i>Corbicula fluminea</i>	DN125059.c0.g1.i19.p1	Vasa/Ddx4	-
<i>Mactra chinensis</i>	DN48157.c2.g2.i2.p1	Vasa/Ddx4	-
<i>Ruditapes decussatus</i>	DN22317.c4.g2.i1.p1	Vasa/Ddx4	-
<i>Ruditapes philippinarum</i>	XP_060562671.1	Vasa/Ddx4	-
<i>Mercenaria mercenaria</i>	XP_053394752.1	Vasa/Ddx4	-
<i>Cyclina sinensis</i>	Hic.asm.11.970.2	Vasa/Ddx4	-
<i>Pisidium coreanum</i>	DN31082.c0.g1.i7.p1	Vasa/Ddx4	-
<i>Dreissena polymorpha</i>	XP_052283635.1	Vasa/Ddx4	-
<i>Mya arenaria</i>	XP_052775885.1	Vasa/Ddx4	-
<i>Sinonovacula constricta</i>	Chr8.1697	Vasa/Ddx4	-
<i>Solen grandis</i>	DN2375.c0.g1.i12.p1	Vasa/Ddx4	-
<i>Tridacna squamosa</i>	DN1975.c0.g1.i5.p1	Vasa/Ddx4	-
<i>Panopea generosa</i>	DN2386.c0.g1.i4.p1	Vasa/Ddx4	-
<i>Cristaria plicata</i>	DN71694.c8.g1.i1.p1	Vasa/Ddx4	-
<i>Hyriopsis bialata (Unio delphinus)</i>	M00000006703	Vasa/Ddx4	-
<i>Potamilus streckersoni</i>	KAK3601505.1	Vasa/Ddx4	-
<i>Margaritifera margaritifera</i>	MMAM00000026330	Vasa/Ddx4	-
<i>Perna viridis</i>	s01977g89	Vasa/Ddx4	-
<i>Perumytilus purpuratus</i>	DN96437.c0.g1.i2.p1	Vasa/Ddx4	-
<i>Mytilus galloprovincialis</i>	VDI03911.1	Vasa/Ddx4	-
<i>Mytilus galloprovincialis</i>	VDI03912.1	Vasa/Ddx4	-
<i>Modiolus modiolus</i>	DN179.c0.g1.i2.p1	Vasa/Ddx4	-
<i>Argopecten irradians concentricus</i>	Contig172.33	Vasa/Ddx4	-
<i>Pecten maximus</i>	XP_033738807.1	Vasa/Ddx4	-
<i>Patinopecten yessoensis</i>	XP_021370692.1	Vasa/Ddx4	-
<i>Ostrea edulis</i>	XP_056020028.1	Vasa/Ddx4	-
<i>Magallana (Crassostrea) gigas</i>	XP_011437246.2	Vasa/Ddx4	-
<i>Crassostrea virginica</i>	XP_022316564.1	Vasa/Ddx4	-
<i>Saccostrea glomerata</i>	Sgl001349	Vasa/Ddx4	-
<i>Pinctada margaritifera</i>	DN36893.c1.g3.i1.p1	Vasa/Ddx4	-
<i>Atrina pectinata</i>	DN813.c0.g1.i1.p1	Vasa/Ddx4	-
<i>Tegillarca granosa</i>	KAJ8305640.1	Vasa/Ddx4	-
<i>Anadara (Scapharca) broughtonii</i>	EVM0008860.1	Vasa/Ddx4	-
<i>Drosophila melanogaster*</i>	NP_001260458.1	Vasa/Ddx4	vasa
<i>Homo sapiens*</i>	NP_077726.1	Vasa/Ddx4	DDX4
<i>Mus musculus*</i>	NP_001139357.1	Vasa/Ddx4	Ddx4
<i>Danio rerio*</i>	NP_571132.1	Vasa/Ddx4	ddx4
<i>Caenorhabditis elegans*</i>	NP_491876.1	Vasa/Ddx4	glh-2
<i>Caenorhabditis elegans*</i>	NP_491963.1	Vasa/Ddx4	glh-1

Table S13 continued from previous page

Species	Gene ID	Orthology group	Gene name
<i>Caenorhabditis elegans</i> *	NP_491681.1	Vasa/Ddx4	<i>glh-3</i>
<i>Caenorhabditis elegans</i> *	NP_491207.3	Vasa/Ddx4	<i>glh-4</i>
<i>Magallana (Crassostrea) gigas</i>	XP_011446924.2	Ddx3	-
<i>Magallana (Crassostrea) gigas</i>	XP_034330003.1	Ddx3	-
<i>Crassostra virginica</i>	XP_022337075.1	Ddx3	-
<i>Ostrea edulis</i>	XP_056006193.1	Ddx3	-
<i>Saccostrea glomerata</i>	Sgl003232	Ddx3	-
<i>Atrina pectinata</i>	DN371.c0.g4.i2.p1	Ddx3	-
<i>Pinctada margaritifera</i>	DN39745.c0.g1.i3.p1	Ddx3	-
<i>Perumytilus purpuratus</i>	DN34627.c0.g1.i16.p1	Ddx3	-
<i>Perna viridis</i>	s00037g119	Ddx3	-
<i>Mytilus galloprovincialis</i>	VDI00208.1	Ddx3	-
<i>Modiolus modiolus</i>	DN49076.c0.g1.i10.p1	Ddx3	-
<i>Tegillarca granosa</i>	KAJ8310842.1	Ddx3	-
<i>Argopecten irradians concentricus</i>	Contig829.57.3	Ddx3	-
<i>Pecten maximus</i>	XP_033759680.1	Ddx3	-
<i>Patinopecten yessoensis</i>	XP_021341010.1	Ddx3	-
<i>Hyriopsis bialata (Unio delphinus)</i>	M00000003015	Ddx3	-
<i>Cristaria plicata</i>	DN67742.c10.g2.i2.p1	Ddx3	-
<i>Megalonaia nervosa</i>	g136014.t1	Ddx3	-
<i>Potamilus streckersoni</i>	KAK3605786.1	Ddx3	-
<i>Margaritifera margaritifera</i>	MMAM00000009046	Ddx3	-
<i>Ruditapes decussatus</i>	DN22481.c1.g4.i1.p1	Ddx3	-
<i>Ruditapes philippinarum</i>	XP_060588962.1	Ddx3	-
<i>Cyclina sinensis</i>	Hic.asm.6.43.1	Ddx3	-
<i>Calyptogena (Archivesica) marissinica</i>	Ama32770	Ddx3	-
<i>Phreagena okutanii</i>	DN58569.c0.g1.i6.p1	Ddx3	-
<i>Mactra chinensis</i>	DN49476.c41.g3.i1.p1	Ddx3	-
<i>Corbicula fluminea</i>	DN126815.c0.g1.i7.p1	Ddx3	-
<i>Dreissena polymorpha</i>	XP_052217061.1	Ddx3	-
<i>Mya arenaria</i>	XP_052782518.1	Ddx3	-
<i>Pisidium coreanum</i>	DN3392.c0.g2.i6.p1	Ddx3	-
<i>Sinonovacula constricta</i>	Chr9.1187	Ddx3	-
<i>Sinonovacula constricta</i>	Chr9.1230	Ddx3	-
<i>Solen grandis</i>	DN51.c4.g1.i12.p1	Ddx3	-
<i>Panopea generosa</i>	DN13909.c0.g3.i2.p1	Ddx3	-
<i>Tridacna squamosa</i>	DN33643.c0.g2.i2.p1	Ddx3	-
<i>Danio rerio</i> *	NP_001119895.1	Ddx3	<i>ddx3xa</i>
<i>Danio rerio</i> *	NP_571016.2	Ddx3	<i>ddx3xb</i>
<i>Mus musculus</i> *	NP_149068.1	Ddx3	<i>Pl10</i>
<i>Homo sapiens</i> *	NP_001180346.1	Ddx3	<i>DDX3X</i>
<i>Drosophila melanogaster</i> *	NP_001262379.1	Ddx3	<i>bel</i>
<i>Pisidium coreanum</i>	DN29220.c0.g1.i2.p1	Ddx3	-
<i>Caenorhabditis elegans</i> *	NP_001021793.1	Ddx3	<i>vbh-1</i>
<i>Caenorhabditis elegans</i> *	NP_001254859.1	Ddx3	<i>laf-1</i>

Table S13 continued from previous page

Species	Gene ID	Orthology group	Gene name
<i>Pisidium coreanum</i>	DN19615.c0.g1.i2.p1	Ddx3	-

Data availability

All the supplementary materials, as well as high-resolution figures and a parsable version of tables, are accessible online at the following GitHub repository:

https://github.com/filonico/phd_thesis_tex

Activity report

This is the report of the activities carried out during my 3-year PhD course (2021–2024).

Research activity

Here are the research activities not directly related to the main topic of the PhD thesis.

- Manual curation of long interspersed nuclear element (LINE) libraries of several bivalve species;
- comparative genomics analysis of Hox and ParaHox genes in brachiopod crustaceans;
- comparative genomics analysis of brachiopod crustaceans to investigate the molecular underpinnings of morphological stasis and genome size variations;
- molecular phylogenetics and Bayesian dating of brachiopod crustaceans;
- preparation of mRNA sequences of genes involved in body segmentation in *Triops cancriformis* (Pancrustacea, Branchiopoda), to be used to generate probes for mRNA *in-situ* HCR on larvae (in collaboration with the Patel Lab; Marine Biology Lab, Woods Hole, MA, USA);
- collection, fixation, and storing of juvenile stages of several stick insect (Insecta, Phasmida) species, to be used for mRNA *in-situ* HCR to investigate the temporal and spatial transcription of genes involved in wing morphogenesis (in collaboration with the Patel Lab; Marine Biology Lab, Woods Hole, MA, USA);
- preparation of a review on the evolutionary causes and consequences of trait loss reversals;
- preparation of mitotic chromosome plates in the red wood ant *Formica paralugubris* from cerebral ganglia of pre-pupae.

Visiting scholar

- Nuzhdin Lab (University of Southern California, Los Angeles, CA, UA; Aug 20th, 2023–Feb 20th, 2024), to accomplish the abroad period of my PhD;
- Juan Pasantes' lab (University of Vigo, Vigo, Spain; Jan 12th–22nd, 2023), for a specific training on chromosome mitotic plate preparation in bivalve species.

Teaching activity

- Practical class “CAFE: estimating gene family turnover across a phylogenetic tree” (Apr 23, 2024) for first-year students of the course “Molecular phylogenetics” pursuing a Master degree in “Bioinformatics” at the University of Bologna (Italy);
- Practical invertebrate zoology class (Sep, 2022–Jan, 2023) for first-year students pursuing a Bachelor degree in “Biological Sciences” at the University of Bologna (Italy).

Co-supervised thesis

- *Evaluation of different calibration methods on Branchiopoda (Crustacea) phylogeny.* Niccolò Righetti. Master degree in “Biodiversità ed evoluzione”, University of Bologna, Bologna (Italy). Supervisor: Andrea Luchetti. Co-supervisor: Filippo Nicolini. AA 2022–2023;
- *Filogenesi molecolare di alcune famiglie dell'ordine Phasmatodea con enfasi sulla famiglia Heteropterygidae (Bacilloidea).* Giacomo Orsini. Bachelor degree in “Scienze biologiche”, University of Bologna, Bologna (Italy). Supervisor: Andrea Luchetti. Co-supervisor: Simona Corneti, Filippo Nicolini. AA 2021–2022;
- *Filogenesi molecolare di specie appartenenti alle famiglie Heteropterygidae e Anisacanthidae (Phasmatodea, Bacilloidea).* Alessandro Siragusa Camacho. Bachelor degree in “Scienze biologiche”, University of Bologna, Bologna (Italy). Supervisor: Andrea Luchetti. Co-supervisor: Simona Corneti, Filippo Nicolini. AA 2021/2022;
- *Filogenesi molecolare di specie della famiglia Pseudophasmatidae.* Giovanni Amedeo Paselli. Bachelor degree in “Scienze biologiche”, University of Bologna, Bologna (Italy).

Supervisor: Barbara Mantovani. Co-supervisors: Simona Corneti, Filippo Nicolini. AA 2020–2021.

Courses and workshops

- *Establishing state-of-the-art mollusc genomics*. EMBO Workshop. Namur, Belgium. May 28th–31st, 2024;
- *Art (Science) Attack*. Physalia Courses. Online. May 20th–23rd, 2024;
- *Introduction to Python for biologists*. Physalia Courses. Online. Sep 2nd–5–2nd–8, 2023;
- *ITA *PHY phylogenetics workshop*. Trento, Italy. Jan 6th–9th, 2023;
- *Sex chromosome evolution*. Physalia Courses. Online. Jan 23rd–27th, 2023.

Awards and scholarships

- Travel grant to attend the “Evoluzione2024” congress in Naples (Italy). Stazione Zoologica Anton Dohrn. Sep 8th–11th, 2024;
- Travel grant to attend the EMBO workshop *Establishing state-of-the-art mollusc genomics* in Namur (Belgium). EMBO. May 28th–31st, 2024;
- Laura Bassi scholarship for editorial assistance to postgraduates and junior academics. Editing Press. Apr 13th, 2023.

Presentations at congresses

Oral presentations

- Nicolini F, Iannello M, Piccinini G, Ghiselli F, Luchetti A, Milani L. (2024). Advancing the study of bivalve sex determination in the light of comparative genomics. Establishing state-of-the-art mollusc genomics (EMBO workshop). Namur (Belgium). May 27-30, 2024;
- Nicolini F, Ghiselli F, Milani L, Luchetti A. (2023). Contrasting patterns of amino acid

evolution and shared ancestry between putative sex-determining genes in bivalve molluscs. EVOLMAR 2023. Online. Nov 14–17, 2023;

- Nicolini F, Ghiselli F, Milani L, Luchetti A. (2023). Sex-determination related genes in bivalves: novel acquisitions and high rates of sequence evolution. Evolution 2023 (Ernst Mayr Award symposium). Online. Jun 2–3, 2023.

Poster presentations

- Nicolini F, Iannello M, Piccinini G, ghiselli F, Nuzhdin S, Luchetti A, Milani L. (2024). How to detect sex-determining genes through molecular evolution: bivalves a a case study. Evoluzione 2024. Naples, Italy. Sep 8–11, 2024;
- Nicolini F, Ghiselli F, Milani L, Luchetti A. (2022). Clues of accelerated molecular evolution in gene families associated wit sex determination in bivalves. SMBE 2023. Ferrara, Italy. Jul 24–27, 2023;
- Nicolini F, Ghiselli F, Milani L, Luchetti A. (2022). Clues of accelerated molecular evolution in gene families associated wit sex determination in bivalves. SIBE/ISEB 2022. Ancona, Italy. Sep 4–7, 2022;
- Nicolini F, Martelossi J, Forni G, Mantovani B, Luchetti A. (2021) First insights and comparative genomics of Hox and ParaHox genes in tadpole shrimps. EuroEvoDevo 2022. Naples, Italy. May 31–Jun 3, 2022.

Invited talks

- *From comparative genomics to fluorescence imaging: a multi-disciplinary approach to study bivalve sex determination.* Auer Lab. University of Fribourg, Fribourg. Jul 26, 2024.

Outreach activity

- Editor and web writer for BioPills – the Italian community of life sciences (biopills.net/). Jul 2017–ongoing;
- Presenter at the European Researchers' Night 2024, University of Bologna, Bologna

(Italy). Sep 27, 2024;

- Presenter at the BiGeA Day 2023, University of Bologna, Bologna (Italy). May 27, 2023;
- Opening Days, University of Bologna, Bologna (Italy). Nov 18, 2022.

Scientific publications

* equal contribution

- Righetti N*, Nicolini F*, Forni G, & Luchetti A. (2024). Towards a time-tree solution for Branchiopoda diversification: a jackknife assessment of fossil age priors. *Submitted for peer-review*
- Nicolini F, Ghiselli F, Luchetti A, & Milani L. (2023). Bivalves as emerging model systems to study the mechanisms and evolution of sex determination: a genomic point of view. *Genome Biology and Evolution*, 15(10), evad181. doi: [10.1093/gbe/evad181](https://doi.org/10.1093/gbe/evad181)
- Martelossi J, Nicolini F, Subacchi S, Pasquale D, Ghiselli F, & Luchetti A. (2023). Multiple and diversified transposon lineages contribute to early and recent bivalve genome evolution. *BMC Biology*, 21(1), 1–23. doi: [10.1186/s12915-023-01632-z](https://doi.org/10.1186/s12915-023-01632-z)
- Nicolini F, Martelossi J, Forni G, Savojardo C, Mantovani B, & Luchetti A. (2023). Comparative genomics of Hox and ParaHox genes among major lineages of Branchiopoda with emphasis on tadpole shrimps. *Frontiers in Ecology and Evolution*, 11, 23. doi: [10.3389/fevo.2023.1046960](https://doi.org/10.3389/fevo.2023.1046960)
- Forni G, Cussigh A, Brock PD, Jones BR, Nicolini F, Martelossi J, Luchetti A, & Mantovani B. (2023). Taxonomic revision of the Australian stick insect genus *Candovia* (Phasmida: Necrosciinae): insight from molecular systematics and species-delimitation approaches. *Zoological Journal of the Linnean Society*, 197(1), 189–210. doi: [10.1093/zoolinnean/zlac074](https://doi.org/10.1093/zoolinnean/zlac074)