



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

**A comparative and evolutionary approach to
study bivalve sex determination from a
broad-phylogenetic perspective**

Candidate: Filippo Nicolini

PhD Coordinator:
prof. Stefano Goffredo

Supervisor:
prof. Andrea Luchetti

Final exam 2025

Contents

Table of contents	i
List of abbreviations	1
Chapter 1: Introduction	4
1.1 The diversity of sexual processes	4
1.2 Genetic sex determination and the evolution of sex-determining genes	5
1.3 Sex determination in bivalves: a long-standing enigma	6
Chapter 2: Bivalves as emerging model systems to study the mechanisms and evolution of sex determination: a genomic point of view.	9
2.1 Open yet inspiring topics in bivalve sex determination	11
2.1.1 Transitions between environmental and genetic sex determination	11
2.1.2 Evolution of sex chromosomes	12
2.1.3 Mito-nuclear interactions	13
2.1.4 Evolution of sex-determination related genes	15
2.2 The case of the Dmrt gene family in bivalves	16
2.3 Conclusions: bivalves as new models in the study of sex determination	19
2.4 Acknowledgments	26
2.5 Data Availability	26
Chapter 3: Identification of putative sex-determination related genes in bivalves through comparative molecular evolutionary analyses	27
3.1 Introduction	28
3.2 Materials and Methods	31
3.2.1 Dataset of bivalve annotated genomes and transcriptomes	31
3.2.2 Identification and classification of Dmrt, Sox and Fox genes in bivalves .	31
3.2.3 Sequence diversity of bivalve single-copy orthogroups	33
3.2.4 Mammals and <i>Drosophila</i> spp. as test datasets	34

3.2.5	GO-term enrichment	35
3.3	Results	36
3.3.1	Genomic and transcriptomic datasets	36
3.3.2	The Dmrt, Sox, and Fox complements in bivalves	36
3.3.3	Amino acid sequence divergence of Dmrt, Sox, and Fox genes in bivalves	38
3.3.4	Dmrt, Sox, and Fox genes, and amino acid sequence divergence in the test datasets	46
3.4	Discussion	48
3.4.1	A new manually-curated and phylogenetic-based reference dataset of Dmrt, Sox, and Fox genes in bivalves	48
3.4.2	High amino acid sequence divergence identifies putative sex-determining genes	52
3.5	Conclusions.	57
3.6	Supplementary Materials	58
3.6.1	Supplementary Figures	58
3.6.2	Supplementary Tables	72
Chapter 4:	Expression patterns of three sex-related genes and the germline marker <i>Vasa</i> in early developmental stages of <i>Mytilus galloprovincialis</i> embryos	125
4.1	Introduction	126
4.2	Materials and Methods	126
4.2.1	Time-series gene expression	126
4.2.2	Sample collection, MitoTracker staining and fixation	127
4.2.3	mRNA <i>in-situ</i> Hybridization Chain Reaction (HCR)	128
4.2.4	Immunolocalization of Vasa	130
4.3	Results	130
4.4	Discussion	131
References.	149
Appendix	150

List of abbreviations

AASD	amino acid sequence divergence
BSA	bovine serum albumin
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
DM	<i>dsx</i> and <i>mab-3</i> [domain]
DMA	DM-associated [domain]
Dmrt	<i>doublesex</i> and <i>mab-3</i> related transcription factor
Dmrt-1L	<i>doublesex</i> and <i>mab-3</i> related transcription factor 1-like
Dm-W	<i>doublesex</i> and <i>mab-3</i> related gene W
Dmy	<i>doublesex</i> and <i>mab-3</i> related gene Y
DSFG	<i>doublesex</i> and <i>mab-3</i> related transcription factor (Dmrt), <i>Sry</i> -related HMG-box (Sox), and forkhead box (Fox) gene
dsx	<i>doublesex</i>
DUI	doubly uniparental inheritance
EDTA	ethylenediaminetetraacetic acid
ESD	environmental sex determination
FASW	filtered artificial sea water
FHA	forkhead-associated [domain]
Fox	forkhead box
GSD	genetic sex determination
GO	gene ontology

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
mRNA-ISH	mRNA <i>in-situ</i> hybridization
<i>mab-3</i>	<i>male abnormal-3</i>
MCL	Markov clustering algorithm
ML	maximum likelihood
Mya	million years ago
ORF	open reading frame
PBS	1× phosphate-buffered saline
PBS-Tw	1× PBS with 0.1% Tween 20
PFA	paraformaldehyde
PGC	primordial germ cell
PSD	primary sex differentiation
RNAi	RNA interference
RT	room temperature
SC	sex chromosome
SDS	sodium dodecyl sulfate
SCO	single-copy orthogroup
SD	sex determination
SDG	sex-determining gene
<i>Sxl</i>	<i>Sex-lethal</i>
Sox	<i>Sry</i> -related HMG-box
SRG	sex-determination related gene
<i>Sry</i>	<i>Sex-determining region of chromosome Y</i>
SSC-Tw	5× saline-sodium citrate with Tween 20 buffer
TBS	1× Tris-buffered saline

TBS-Tx 1× TBS with Triton X-100

tra *transformer*

Chapter 1

Introduction

1.1 The diversity of sexual processes

The process of sex determination (SD) has been traditionally associated with the very first step of gonad differentiation, where an initial trigger activates the molecular pathway that establishes organism sex. According to this view, two alternative types of SD have been recognized at first: the environmental sex determination (ESD) and the genetic sex determination (GSD), depending on whether the very first cues are of environmental or genetic origin. Conversely, all the downstream events of gonad development (i.e., after SD) have been appointed as primary sex differentiation (PSD), which consists of the entire set of morphogenetic, molecular, and physiological events leading to the full maturation of testes or ovaries (**Uller and Helanterä, 2011; Beukeboom and Perrin, 2014**). Lately, however, the dichotomous views of ESD/GSD and of SD/PSD have been questioned. On the one hand, a growing number of studies on non-model organisms proved that ESD and GSD represent a continuum of mixed conditions rather than two mutually exclusive phenomena. On the other, 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 PSD of unclear utility (**Beukeboom and Perrin, 2014**).

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”, without any actual distinction between environmental/genetic initial triggers or the downstream effectors. However, I argue that this definition should be expanded to encompass not only the embryonic stage of the animal life cycle

but also adulthood, since cases of sex reversals and sex changes (sequential hermaphroditism) legitimately express proper SD processes during post-embryonic life stages as well.

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

In its most intimate core, animal SD is the manifestation of complex gene regulatory networks where, in accordance with the Wilkins' theory (1995), the downstream actors appear to be nearly conserved both from functional and identity point of views, while the master top regulators (the commonly recognized sex determinants, such as the *Sex-determining region of chromosome Y (Sry)* in therians or the ratio between sex and autosome chromosomes in *Drosophila*) are often the most variable part (**Beukeboom and Perrin, 2014**). As a matter of fact, this evolutionary pattern of animal sex-determining cascades has been observed in major animal clades, including vertebrates (e.g., **Marshall Graves and Peichel, 2010**), insects (e.g., **Verhulst et al., 2010**), and nematodes (e.g., **Stothard and Pilgrim, 2003**).

Sex-determination related genes (SRGs) are of particular interest not only from a regulatory point of view but also because of their patterns of molecular evolution. In fact, transcriptionally sex-biased genes (including SRGs) 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 (reviewed in **Parsch and Ellegren, 2013; Grath and Parsch, 2016**), as it has been repeatedly observed in well-studied organisms such as fruit flies (e.g., **Meisel and Connallon, 2013**), nematodes (e.g., **Cutter and Ward, 2005**), mice (e.g., **Kousathanas et al., 2014**) and primates (e.g., **Khaitovich et al., 2005**), and in other emerging model systems, such as *Daphnia pulex* (**Eads et al., 2007**), aphids (**Purandare et al., 2014**), and two wasp species of the genus *Nasonia* (**Wang et al., 2015**). Growing evidence is however showing cases in which instead 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**).

The pattern of molecular evolution of sex-biased genes is particularly evident in organisms with sex chromosomes (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 for sex-chromosome

associated genes (**Vicoso and Charlesworth, 2006; Mank et al., 2007; Meisel and Connallon, 2013**). This high rate of sequence evolution in sex-biased genes and sex chromosomes (SCs) can be the result of both adaptative and non-adaptative processes, since the observed higher ratio between non-synonymous and synonymous mutations (dN/dS) can be caused by natural selection, sexual selection or sexual antagonism, as well as genetic drift (**Vicoso and Charlesworth, 2006; Meisel and Connallon, 2013; Parsch and Ellegren, 2013; Grath and Parsch, 2016**).

1.3 Sex determination in bivalves: a long-standing enigma

Bivalves are the second largest clade in molluscs, counting more than 18,000 species (Catalogue of Life) 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**). However, despite their socio-economic and scientific importance, the knowledge concerning the molecular basis of bivalve reproduction and SD is still quite limited (**Breton et al., 2018**). Clues from various works seem to suggest that both genetic and environmental factors (e.g., temperature, food availability, and steroids) are involved in SD, and that heteromorphic sex chromosomes (HeSCs) are absent (**Breton et al., 2018; Han et al., 2022**). However, the exact process by which sex is determined and gonad commitment is established is, currently, still unknown. Actually, bivalves represent a dazzling example of how the traditional dichotomies between ESD/GSD and SD/PSD can sometimes hamper scientific research, as many bivalve species exhibit various forms of hermaphroditism and because a master environmental or genetic sex determinant inducing PSD may just not exist.

In the attempt to identify SRGs, many differential gene expression analyses have been recently performed on a variety of species covering most of the phylogenetic diversity of bivalves (e.g., **Milani et al., 2013; Zhang et al., 2014; Chen et al., 2017; Capt et al., 2018; Ghiselli et al., 2018; Shi et al., 2018**). Some of the genes that were found to be differentially expressed between gonads of different sex were systematically retrieved across species, such as those belonging to the *doublesex* and *mab-3* related transcription factor (Dmrt), *Sry*-related HMG-box (Sox), and forkhead box (Fox) families, which act in concert in various animal

developmental processes including the SD cascade (Marshall Graves and Peichel, 2010; Beukeboom and Perrin, 2014). To this regard, Zhang et al., 2014 proposed a working model for the sex-determining pathway of the Pacific oyster *Crassostrea gigas* in which: *CgSoxH* promotes male gonad development by activating *CgDsx*, which belong to the Dmrt family, and inhibiting *CgFoxL2*; *CgFoxL2*, when not inhibited by the pair *CgSoxH/CgDsx*, promotes female gonad development. Moreover, Han et al., 2022 recently identified homomorphic sex chromosomes (HoSCs) in eight scallop species and appointed *FoxL2* as a putative SRG in *Patinopacten yessoensis* and *Chlamys farreri*. Though, much of the recent research effort on bivalve SRGs has been limited to their molecular cloning, differential transcription, and tissue localization (Liang et al., 2019; Sun et al., 2022). Furthermore, few works have directly investigated the biological functions of Dmrt, Sox, and Fox genes in bivalves so far, and most used post-transcriptional silencing of target mRNAs [RNA interference (RNAi)]. 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*, *FoxL2* might be related to the *Wnt/β-catenin* signaling pathway, which takes part in ovarian differentiation also in vertebrates. Sun et al., 2022 found instead that in *C. gigas*, *FoxL2* and *Dmrt1L* mRNA knockdown results in the size reduction of female and male mature gonads, respectively.

In this sense, bivalve molluscs represent a striking example of the difficulty to reconcile the traditional view of a single sex determinant with an apparent multifactorial model in which many genes and environmental cues act in concert to establish the sexual identity of the individual (Breton et al., 2018). Lately, much effort has been put in the characterisation of bivalve SD and a general framework is eventually taking shape. Functional assays with RNAi and CRISPR-Cas9 techniques (e.g., Wang et al., 2020; Sun et al., 2022; Wang et al., 2022), as well as with mRNA *in-situ* hybridization (mRNA-ISH) and immunohistochemistry (e.g., Perez-Garcia et al., 2011; Milani et al., 2013), are 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 class of organisms (Milani and Ghiselli, 2020). The high evolutionary rates and plasticity of SRGs make the situation even harder, since phylogenetic and orthology

inferences can lead to erroneous reconstructions in the presence of signal saturation and high sequence divergence (reviewed in **Natsidis et al., 2021**; **Lozano-Fernandez, 2022**).

Chapter 2

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.
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 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.

2.1 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 et al., 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.

2.1.1 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. 2.1**). 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**).

2.1.2 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. 2.1**). 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.

2.1.3 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. 2.1; 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

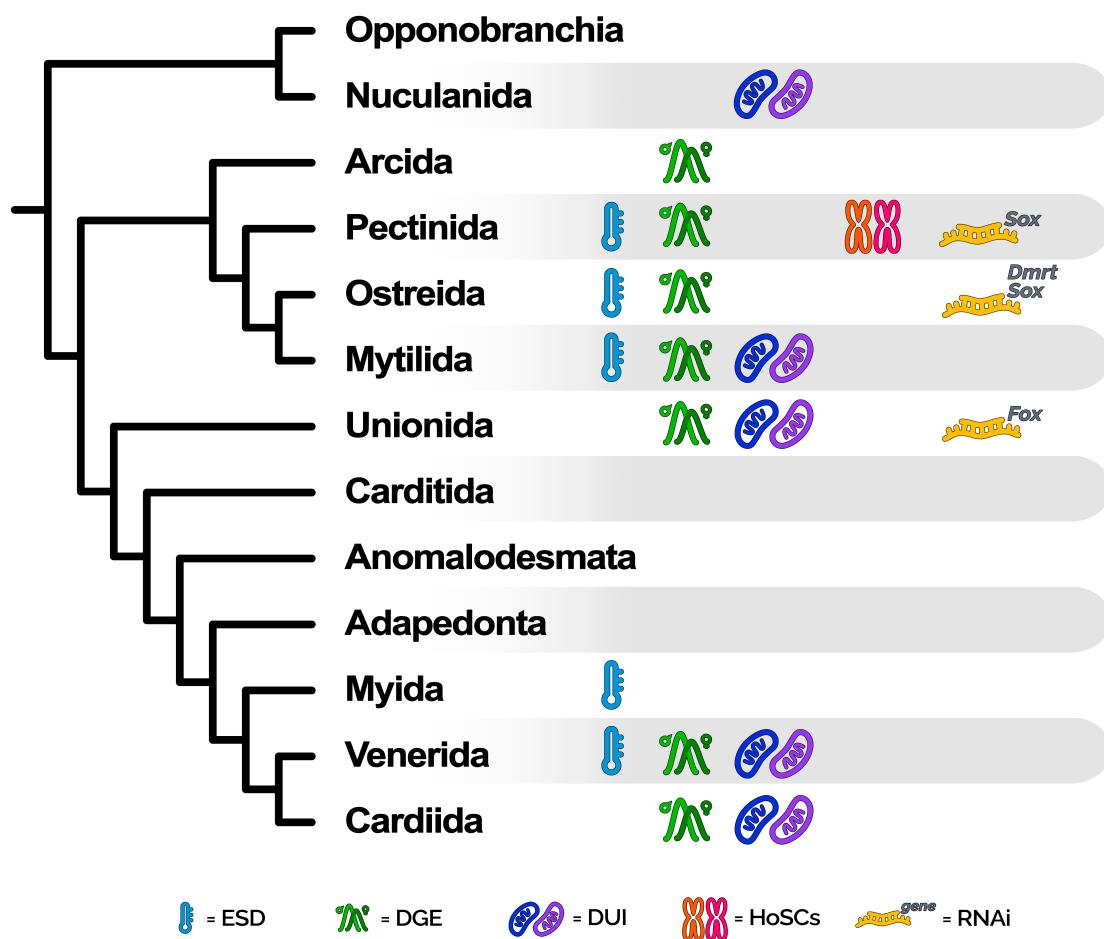


Figure 2.1. Graphical summary of the available knowledge and experiments concerning the genetic basis of SD in bivalves, at the level of major taxonomic orders (as reported in WoRMS; accessed before or on 14/03/2023). For each bivalve clade it is reported: (i) the availability of records of ESD (ii) 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 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 test.

of plasticity of such mitochondria-related SD systems? Are mitochondria-related SD systems more widespread in eukaryotes than currently thought?).

2.1.4 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. 2.1**). 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 *doublesex* 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* (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, Dmrt, Sox, and Fox genes 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. 2.1**; e.g., **Liang et al., 2019; Wang**

et al., 2020; Sun et al., 2022].

Overall, Dmrt, Sox, and Fox genes 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.

2.2 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 *doublesex and mab-3 related transcription factor 1-like* (*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 (Fig. 2.1; Fig. 2.2A), 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. 2.2B). 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. 2.1C), which may

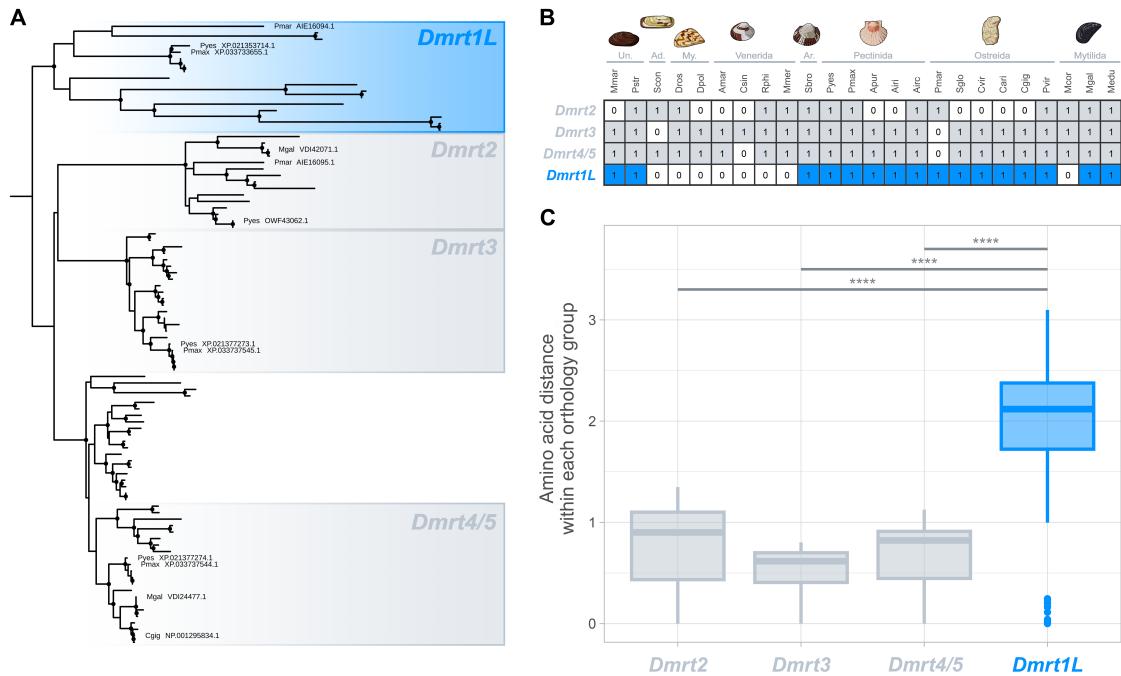


Figure 2.2. Phylogenetic tree (A) and taxonomic distribution (B) of Dmrt genes in bivalves, and comparison of amino acid pairwise distances within *Dmrt-1L* and the other Dmrts (C). (A) Dmrt orthologs from bivalve genome assemblies were obtained with HMMsearch (HMMER toolkit; Eddy, 2011) with the Pfam 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 trimAl (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. 2.1 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-value $\approx 2.2\text{e-}16$) and for group effect with the Kruskal-Wallis test (p-value $\approx 2.2\text{e-}16$), 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 $p \leq 2.2\text{e-}16$ (****) (Bonferroni correction for multiple test was applied). The list of genome assemblies used for these analyses and species identifiers can be found in Fig. 2.1. Un.: Unionida; Ad.: Adapedonta; My.: Myida; Ar.: Arcida.

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 et al., 2018; Evensen et al., 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 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. 2.2C**), 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 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. 2.1**). 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. 2.2A**), 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. In fact, *Fox-L2* has already been appointed as the female sex-determining gene 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.

2.3 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. 2.1**). 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 Dmrt, Sox and Fox 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 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 2.1. 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	NA
<i>Sinonovacula consticta</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

Tab. 2.1 continued from previous page

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<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_905397895.1
<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
<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%	Peñaloza et al., 2021	GCF_902806645.1

Tab. 2.1 continued from previous page

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<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	NA

Tab. 2.1 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>Mizuhoppecten (Patinopecten) yessoensis</i>	Pyes	Pectinida	Scaffold	C:98.6% [S:75.2%,D:23.4%] F:0.4% M:1.0%	Wang et al., 2017	GCF_002113885.1
<i>Margaritifera margaritifera</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>Potamilius streckeroni</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

Tab. 2.1 continued from previous page

Species	ID	Order	Assembly level	BUSCO score	Reference	NCBI Acc. No.
<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

2.4 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.

2.5 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 3

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.*

In preparation.

3.1 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 therians (Pamilo and O'Neill, 1997; Mawaribuchi et al., 2012), *doublesex and mab-3 related gene W (Dm-W)* of the African clawed frog *Xenopus laevis*, and *doublesex 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 *doublesex* 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 families. 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 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, this study aims to perform a thorough investigation 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 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 downregulation 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 (as *Sry* in therians; 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, while (ii) in the fruit fly dataset no gene among those working within the sex-determining cascade is evolving at a higher pace. 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.

3.2 Materials and Methods

3.2.1 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 (**Supp. Tab. S3.1**). 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 on the resulting gff annotation file.

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**. Briefly, raw reads were trimmed using Trimmomatic (**Bolger et al., 2014**) and assembled using Trinity (**Grabherr et al., 2011**) with default parameters. Isoforms were removed using the dedicated perl script from the Trinity utilities. Open reading frames were predicted through TransDecoder (**Haas, n.d.**), by also including diamond (**Buchfink et al., 2015**) and HMMER (v3.3.2; <http://hmmer.org/>) annotation of hits.

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**).

3.2.2 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; <http://hmmer.org/>). 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 1.0e–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 1.0e–5 were retained. Genes which were correctly annotated by both systems (on the basis of the PANTHER gene family and CDD domain identifiers; **Supp. Tab. S3.2**) were kept for subsequent analyses.

DSFGs from *Homo sapiens*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (**Supp. Tab. S3.3**; 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**), 1000 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 3.4), 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**).

3.2.3 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. 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) groups and built the distribution of their median amino acid sequence divergence (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 (**Supp. Tab. S3.1**), 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, JTTDCMut, 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**). 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 decomposed orthogroups (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 same pipeline was also employed to obtain pairwise amino acid distances for each DSFG single-copy orthologous group.

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’.

3.2.4 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 (**Supp. Tab. S3.4–S3.5**, 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 species or inside the boundaries of taxonomic genera (REFERENCE). 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 (**Figure 3.1**).

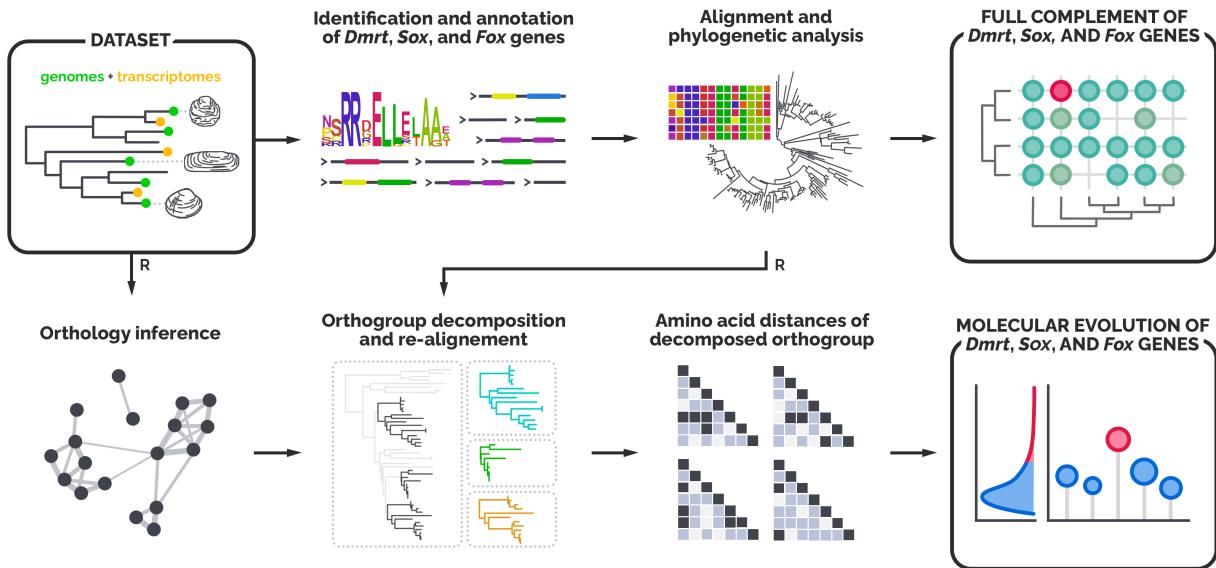


Figure 3.1. Workflow of the analyses for the bivalve dataset. Starting from a set of both genomes and transcriptomes covering a great portion of bivalve taxonomic diversity, we first characterized the entire complement of glsdsfg genes (upper row). In particular, we used sequence annotation and phylogenetic 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 gene single-copy orthologous groups, we calculated the amino acid distances within each orthogroup and then we 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.

3.2.5 GO-term enrichment

After having obtained the distributions of AASD in the three datasets (Bivalvia, Mammalia, and *Drosophila*) 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 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 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**).

3.3 Results

3.3.1 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.; **Supp. Tab. S3.1**). 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 (**Supp. Tab. S3.1**). 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. 3.1**) consists instead of 36 genomes and transcriptomes (**Supp. Tab. S3.1**), 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; **Supp. Tab. S3.4**), and covers 12 major orders, while the fruit fly dataset consists of 17 species and 1 outgroup (*Anopheles gambiae*, Culicidae; **Supp. Tab. S3.5**), 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 (**Supp. Tab. S3.4–S3.5**).

3.3.2 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 the paragraph **The Dmrt, Sox, and Fox complements and their amino acid divergence in the testing datasets COMPLETARE COMPLETARE**).

We retrieved four main orthology groups of Dmrt genes in bivalves (**Fig. 3.2**; **Supp.**

Fig. S3.1; Supp. Tab. S3.6), 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 *doublesex and mab-3 related transcription factor 1-like* (*Dmrt-1L*), as per **Li 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 (about 56%) and *Dmrt-4/5* the lowest (about 7%; **Supp. Tab. S3.7**). 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* (**Supp. Tab. S3.6**). Additionally, we retrieved six main orthology groups of Sox genes, none of which is restricted to molluscs or bivalves (**Fig. 3.2**; **Supp. Fig. S3.2**; **Supp. Tab. S3.6**). 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 (**Supp. Fig. S3.2**), 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 about 8%, with *Sox-H* having the highest (about 21%) and *Sox-B1/2* and *Sox-C* both having no missing genes (**Supp. Tab. S3.7**). The Sox N-terminal signature domain was annotated for *Sox-E* genes (**Supp. Tab. S3.6**). Concerning Fox genes, we retrieved 27 main orthology groups (**Fig. 3.2**; **Supp. Fig. S3.3**; **Supp. Tab. S3.6**), 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 **Materials and Methods** REFERENCE REFERENCE; **Supp. Tab. S3.6**). 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. 3.2; Supp. Fig. S3.3). The degree of missing data for Fox genes in bivalves is about 22%, with *Fox-H* having the highest (about 42%) and *Fox-J1* having no missing genes (Supp. Tab. S3.7). 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 (Supp. Tab. S3.6). 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 unguiculatus* (formerly *coruscus*), and *Pecten maximus* have no missing data, while *Loripes orbiculatus* has the highest proportion (about 64%; Supp. Tab. S3.7).

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

In the reduced bivalve dataset, OrthoFinder collectively analysed >1.2G 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, it includes both genomes and transcriptomes with highly different BUSCO scores (Supp. Tab. S3.1). In order to be able to analyse a greater number of genes, we decomposed OrthoFinder orthogroups using DISCO and eventually obtained 11k SCOs with at least 50% of the species. By running the same pipeline on DSFGs, we included in the AASD analysis 32 SCOs (Fig. 3.2) out of 33 initial Possvm-identified groups (*Fox-H* didn't meet the species occupancy threshold; Fig. 3.3).

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. 3.3), 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 ($R = 0.84, p < 2.2e-16$, calculated on 200 randomly-selected trees; Fig. 3.3C), while it showed no influence from the alignment length ($R = 0.11$) or the number of represented species ($R = -0.23$; Fig. 3.3D–E). Genes from Group 1 and Group 2 are strongly involved in cellular

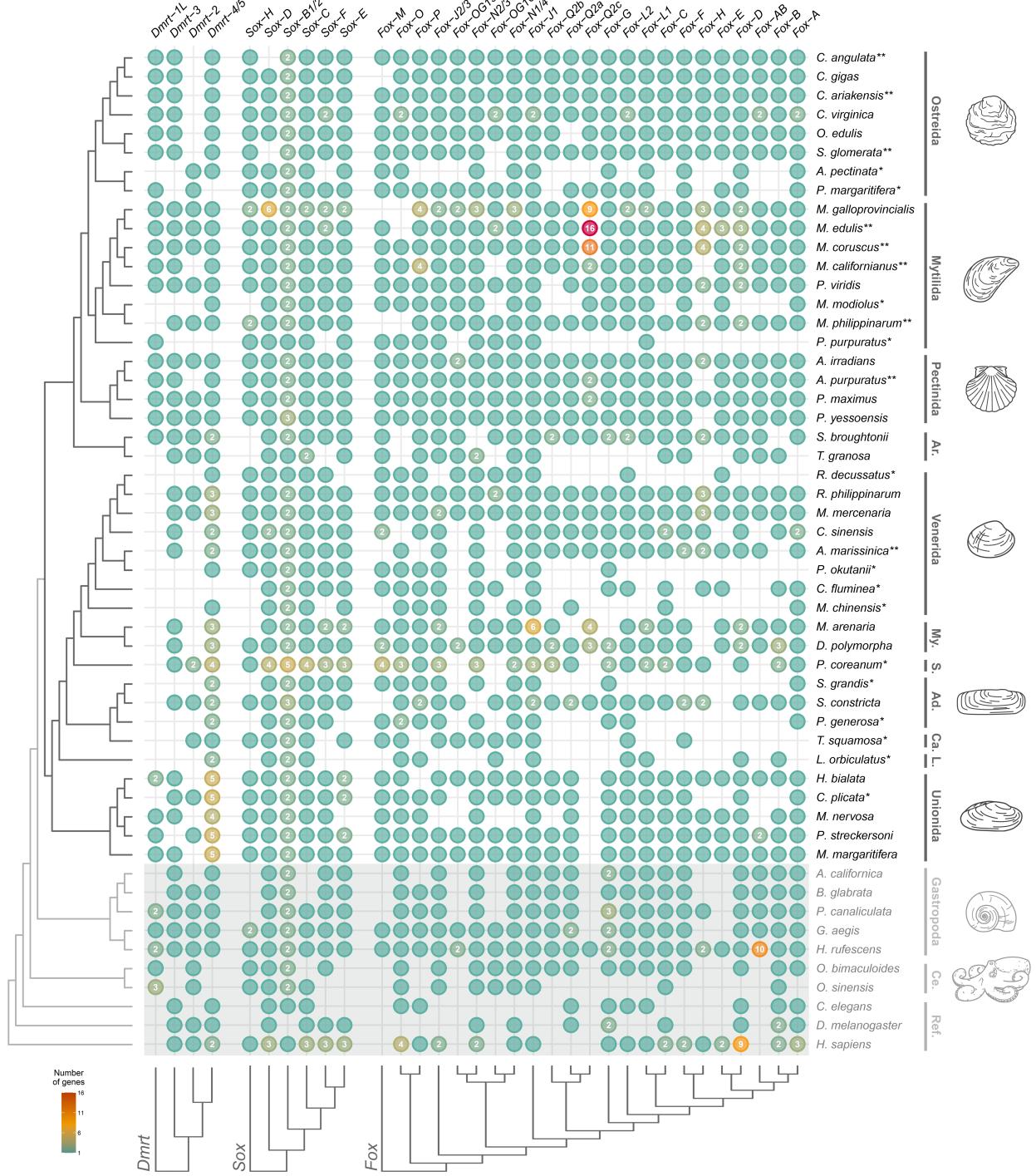


Figure 3.2. DSFG complement in bivalves and their outgroups. Presence/absence of genes in various species are indicated by filled circles. Numbers inside each circle specify genes with 2 or more copies. The shaded area highlights non-bivalve species, belonging either to other molluscs or to the references. 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. 3.1); note that the two categories do not overlap. DSFG trees are shown on the bottom (full trees can be found in Supp. Fig. S3.1–S3.3). Full species names, along with all assembly and taxonomic information, can be found in Supp. Tab. S3.1. Ad.: Adapedonta; Ar.: Arcida; Ca.: Cardiida; Ce.: Cephalopoda; L.: Lucinida; My.: Myida; Ref.: reference genes; S.: Sphaeriida.

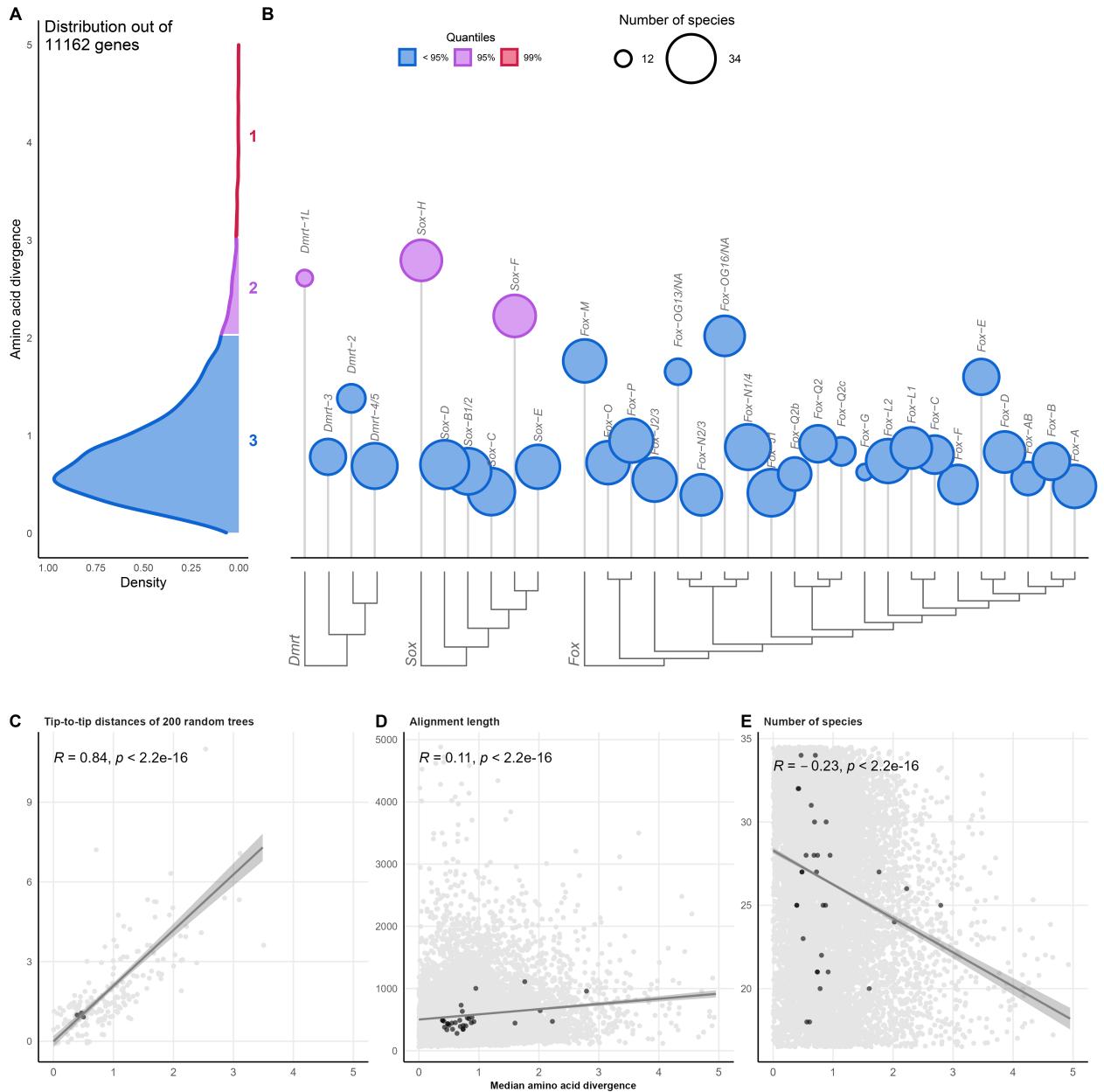


Figure 3.3. Distribution of AASD of single-copy orthogroups in bivalves (A), including DSFGs (B), and their correlations with tip-to-tip distances (C), alignment lengths (D), and number of species (E). The distribution of AASD has been computed on the median values of pairwise distances of >11k SCOs from the reduced bivalve dataset (see main text and Fig. 3.1). 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 Supp. Fig. S3.1–S3.3). 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.

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. 3.1**; **Supp. Tab. S10**).

Table 3.1. Enriched GO terms for Group 1 and Group 2 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 **Supp. Tab. S10**.

Dataset	GO.ID	Term	Annotated genes		Significant genes	Corrected p-value
			Annotated genes	Significant genes		
Bivalvia	GO:0060255	regulation of macromolecule metabolic process	737	59	0.04525	
	GO:0080090	regulation of primary metabolic process	673	53	0.01818	
	GO:0019219	regulation of nucleobase-containing compound metabolic process	541	41	0.02388	
	GO:0006351	DNA-templated transcription	571	39	0.03767	
	GO:0032774	RNA biosynthetic process	579	39	0.04490	
	GO:0051252	regulation of RNA metabolic process	517	37	0.02719	
	GO:0006355	regulation of DNA-templated transcription	490	35	0.03751	
	GO:2001141	regulation of RNA biosynthetic process	491	35	0.03844	
	GO:0006950	response to stress	370	33	0.01949	
	GO:0032502	developmental process	261	27	0.04445	
	GO:0006468	protein phosphorylation	345	23	0.02483	
	GO:0031325	positive regulation of cellular metabolic process	125	17	0.00801	
	GO:0010604	positive regulation of macromolecule metabolic process	151	17	0.04047	
	GO:0051172	negative regulation of nitrogen compound metabolic process	117	16	0.00814	
Mammals	GO:0051173	positive regulation of nitrogen compound metabolic process	137	15	0.02454	
	GO:0006310	DNA recombination	66	14	0.00087	
	GO:0048513	animal organ development	83	12	0.04088	
	GO:0010629	negative regulation of gene expression	78	11	0.00048	
	GO:0023051	regulation of signaling	133	11	0.02872	
	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	64	11	0.03637	
	GO:0009605	response to external stimulus	90	11	0.04544	

Tab. 3.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes		Significant genes	Corrected p-value
			63	11		
Bivalvia	GO:0044419	biological process involved in interspecies interaction between organisms	1297	145		0.04761
	GO:0006955	immune response	853	112		0.00061
	GO:0098542	defense response to other organism	647	82		0.02066
	GO:0045087	innate immune response	630	51		8.5e-10
	GO:0001817	regulation of cytokine production	233	45		0.04660
	GO:0042742	defense response to bacterium	642	45		1.7e-07
	GO:0006954	inflammatory response	382	44		0.01735
	GO:0019221	cytokine-mediated signaling pathway	342	44		3.9e-07
	GO:0002250	adaptive immune response	402	41		1.3e-05
	GO:0001819	positive regulation of cytokine production	308	37		0.02723
	GO:0002697	regulation of immune effector process	432	35		0.04426
	GO:0042110	T cell activation	257	34		1.9e-07
	GO:0051607	defense response to virus	491	32		0.02255
	GO:0048232	male gamete generation	478	31		0.02801
	GO:0007283	spermatogenesis	273	29		0.01285
	GO:0070661	leukocyte proliferation	221	29		0.04833
	GO:0002449	lymphocyte mediated immunity	212	25		0.01870
	GO:0070663	regulation of leukocyte proliferation	300	24		0.00235
	GO:0050727	regulation of inflammatory response	240	24		0.01239
	GO:0031349	positive regulation of defense response	177	22		0.00336
	GO:0002768	immune response-regulating cell surface receptor signaling pathway	66	17		1.7e-10
	GO:0050829	defense response to Gram-negative bacterium	164	17		0.00012
	GO:0071222	cellular response to lipopolysaccharide				

Tab. 3.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes		Significant genes	Corrected p-value
			Annotated genes	Significant genes		
Mammalia	GO:0010466	negative regulation of peptidase activity	163	16	0.00036	
	GO:0002429	immune response-activating cell surface receptor signaling pathway	164	16	0.00243	
	GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	137	16	0.01244	
	GO:0071706	tumor necrosis factor superfamily cytokine production	137	16	0.01244	
	GO:0070665	positive regulation of leukocyte proliferation	132	16	0.02765	
	GO:0045089	positive regulation of innate immune response	113	16	0.03224	
	GO:0071356	cellular response to tumor necrosis factor	175	15	0.00219	
	GO:0002695	negative regulation of leukocyte activation	148	15	0.01151	
	GO:0002456	T cell mediated immunity	82	15	0.01605	
	GO:0002705	positive regulation of leukocyte mediated immunity	113	15	0.01837	
<i>Drosophila</i>	GO:0032680	regulation of tumor necrosis factor production	133	15	0.03262	
	GO:0032640	tumor necrosis factor production	133	15	0.03262	
	GO:0050866	negative regulation of cell activation	165	15	0.04048	
	GO:0000819	sister chromatid segregation	140	11	0.02927	
	GO:0070192	chromosome organization involved in meiotic cell cycle	54	9	0.00849	
	GO:0007131	reciprocal meiotic recombination	37	7	0.00066	
	GO:0007143	female meiotic nuclear division	54	6	0.02270	
	GO:0035967	cellular response to topologically incorrect protein	44	5	0.03334	
	GO:0035966	response to topologically incorrect protein	47	5	0.04266	
	GO:0007141	male meiosis I	13	4	0.00150	
	GO:0140543	positive regulation of piRNA transcription	3	3	6.9e-05	
	GO:0010526	retrotransposon silencing	8	3	0.00331	
	GO:0007130	synaptonemal complex assembly	10	3	0.00666	

Tab. 3.1 continued from previous page

Dataset	GO.ID	Term	Annotated genes		Significant genes	Corrected p-value
			Annotated genes	Significant genes		
<i>Drosophila</i>	GO:0030719	P granule organization	11	3	0.00888	
	GO:0071218	cellular response to misfolded protein	12	3	0.01149	
	GO:0051788	response to misfolded protein	12	3	0.01149	
	GO:0007135	meiosis II	15	3	0.02169	
	GO:0034508	centromere complex assembly	19	3	0.04094	

3.3.4 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 (**Supp. Fig. S3.4A, S3.5–S3.7; Supp. 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 650M genes, and the number of SCOs used in the AASD analysis (thus resulting from the DISCO-based orthogroup decomposition pipeline) is >16k (**Fig. 3.4A**). 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. 3.4B**), while *Sry* and *Fox-D4* showed higher divergences, and have been accordingly placed in Group 1 and 2, respectively. Genes from Group 1 and Group 2 show a strong enrichment in immune-related functions (such as innate and adaptive immune response, defence response to bacteria and viruses, lymphocyte methabolism, etc.), but also in reproductive processes (such as spermatogenesis; **Tab. 3.1; Supp. Tab. 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 (**Supp. Fig. S3.4B, S3.8–S3.10; Supp. Tab. S9**). OrthoFinder analysed about 240M, and the distribution of median AASD was built after >12k SCOS (**Fig. 3.4C**). 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. 3.4D**). Genes of Group 1 and Group 2 show a GO-term enrichment in meiotic processes, such as chromosome/chromatid organisation, and retrotransposon silencing (**Tab. 3.1; Supp. Tab. S10**).

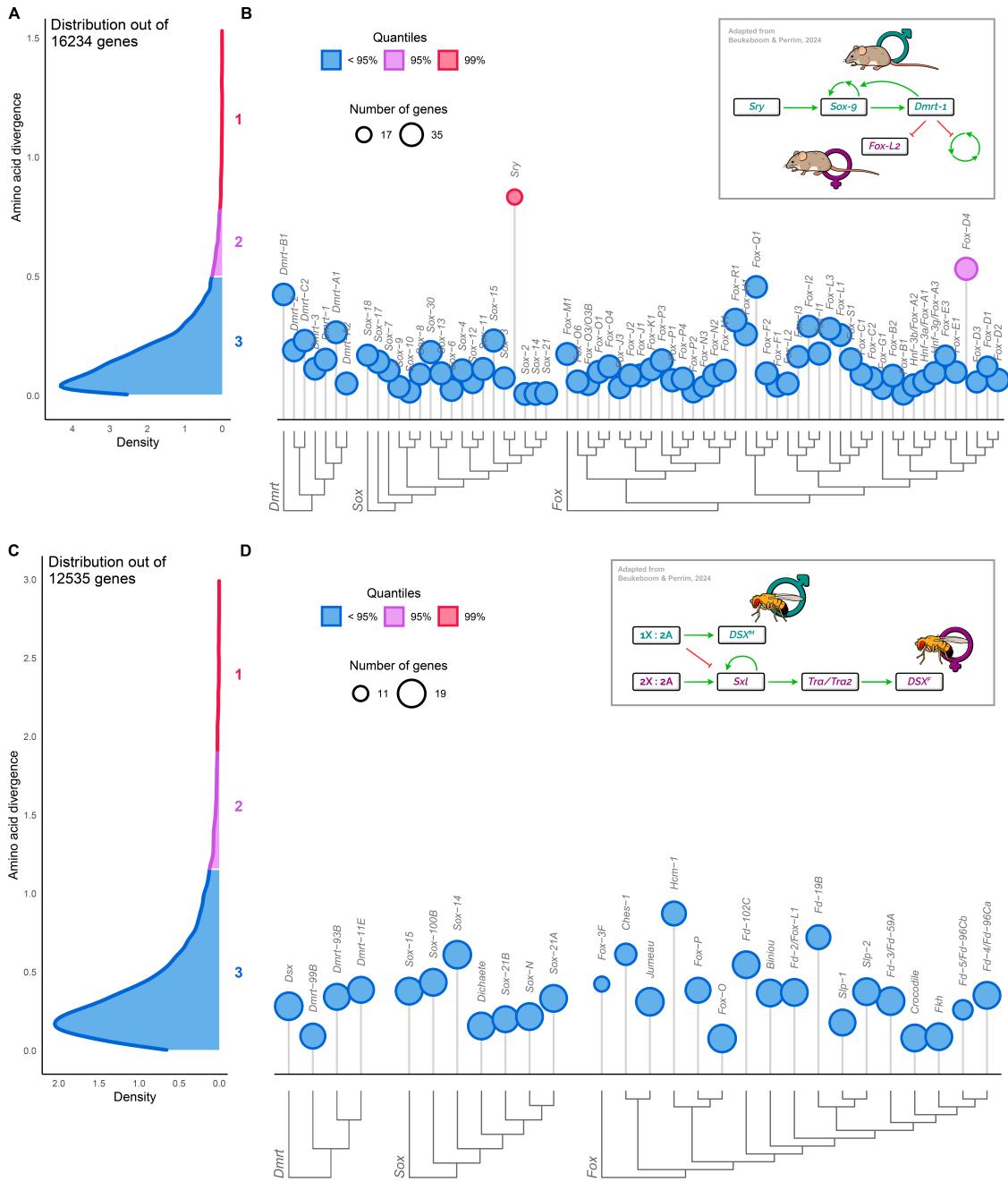


Figure 3.4. Distribution of amino acid divergence (AASD) of single-copy orthogroups in Mammalia (A) and *Drosophila* (C), including Dmrt, Sox, and Fox genes (DSFGs; B-D). The distributions of AASD in mammals and fruit flies have been computed on the median values of pairwise distances of over 16k and 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 Supp. Fig. S3.5–S3.7 for mammals and in Supp. Fig. S3.8–S3.10 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.

3.4 Discussion

3.4.1 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 annotation, 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*; **3.2; Supp. Fig. S3.1; Supp. Tab. S3.6**), 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. 3.2**). Furthermore, we confirm *Dmrt-1L* to be present in many bivalve species (mainly belonging to the Ostreida, Pectinida, Mytilida, and Unionida orders; **Fig. 3.2**), 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 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 (**Supp. Fig. S3.1**); though, it is recovered when analysing just genes from mollusc species (**Supp. Fig. S3.11**). 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 (**Supp. Tab. S3.3**), which are placed among the other bivalve genes (**Supp. Fig. S3.1**), but also because of the high AASD of *Dmrt-1L* genes (see the following section), 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, (i) the scallop-specific cluster of *Dmrt* genes retrieved by **Wang et al., 2023** rather belongs to the *Dmrt-1L* group, and (ii) 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 are *Dmrt-4/5*; *Dmrt-2* genes are *Dmrt-3*; *Dmrt-3* genes 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. 3.2; Supp. Fig. S3.2; Supp. Tab. S3.6**), 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 (**Supp. Fig. S3.2**). 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 (**Supp. Fig. S3.12**).

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. 3.2; Supp. Fig. S3.3; Supp. Tab. S3.6; 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* (**Supp. Fig. S3.3**), 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. 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 (**Supp. Fig. S3.3**). 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 (**Supp. Fig. S3.13**), which belong to the same Fox group. Regarding the *Fox-OG39/NA* group, it does not have any homolog in reference species (**Supp. Fig. S3.3**) 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**).

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 (Supp. Fig. S3.3), 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 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 (Supp. Fig. S3.14). Also, the forkhead domains of *Fox-OG13/NA* genes were annotated as ‘forkhead domain P’ (Supp. Tab S3.6). 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, even when analysing annelids (Seudre et al., 2022) and panarthropods (Schomburg et al., 2022) in a more focused effort. In this case, the forkhead domains were annotated as either a generic ‘forkhead domain’ or a ‘forkhead domain Q2’ (Supp. Tab. S3.6). All considered, we argue that bivalves possess two additional Fox groups (here *Fox-OG13/NA* and *Fox-OG16/NA*; Fig. 3.2; Supp. Fig. S3.3; Supp. Tab. S3.6) 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 (Supp. Fig. S3.1–S3.3; Supp. Tab. S3.6). 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; **Supp. Fig. S3.3**). 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. 3.2** 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 (**Supp. Fig. S3.5–S3.10; Supp. Tab. S8–S9**), with little or no manual curation. 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**).

3.4.2 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 *a-priori* 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, 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 more than 11k SCOs from bivalve genomes (**Fig. 3.3A**), 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. 3.3B**), i.e., within highly divergent genes (Group 1 and Group 2 genes of the distribution; see **Section 3.2**). 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. 3.3C**; $R = 0.84, p < 2.2\text{e-}6$).

Among DSFGs, three fell within the 5% upper quantile, namely *Dmrt-1L*, *Sox-H*, and *SoxF*. Interestingly, *Dmrt-1L* and *Sox-H* have been already proposed to be involved in the male SD pathway of *C. gigas* (*inset* in **Fig. 3.3B**; **Zhang et al., 2014**), on the basis of DGE analyses. Specifically, *Sox-H* would play a major role in *C. gigas* SD, by interacting with *Dmrt-1L* and determining the onset of the male phenotype development; at the same time, both *Sox-H* and *Dmrt-1L* would inhibit *Fox-L2*, 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 et al., 2018**; **Liang et al., 2019**; **Yue et al., 2021**) and RNA interference (RNAi) (**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. 3.4A–B**), the male sex-determining gene in eutherians (*inset* in **Fig. 3.4A–B**); (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. 3.4C–D**), including the downstream effector *dsx* (*inset* in **Fig. 3.4D**). Also *Sxl* and *tra*, both involved in the SD pathway of *Drosophila* (*inset* in **Fig. 3.4D**) 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 (**Fig. 3.4D**). 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 both *Drosophila* (**Mullon et al., 2012**) and vertebrates, despite in the latter case it 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. 3.1; Supp. Tab. 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 *c(2)M*, *c(3)G*, *corona*, and *corolla* proteins; **Tab. 3.1; Supp. Tab. S10**), which again are known to be rapidly evolving (**Hemmer and Blumenstiel, 2016**). In other words, the test datasets allow us to directly link the high AASD (as found 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. 3.1; Supp. Tab. 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 (i.e., molecular evolution can be used to look for putative primary SDGs in taxonomic-wide analyses). As a matter of fact, 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**, respectively] 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. 3.4D) have indeed 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 all congeneric and 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 (Supp. Fig. S3.15). 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 (Supp. Fig. S3.15).

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. 3.1; Supp. Tab. 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. 3.3B) and *Fox-D4* in mammals (Fig. 3.4B). 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 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.

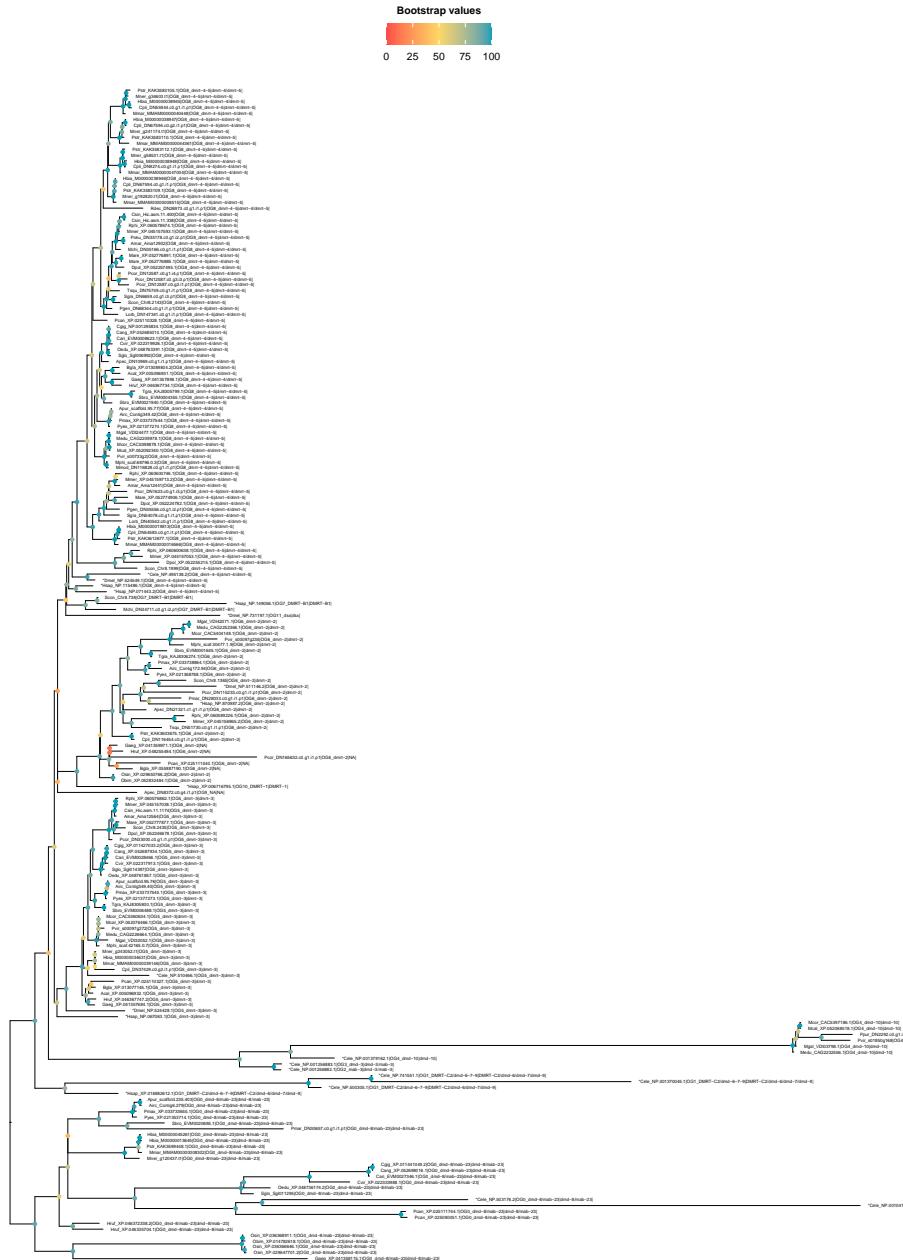
3.5 Conclusions.

In preparation.

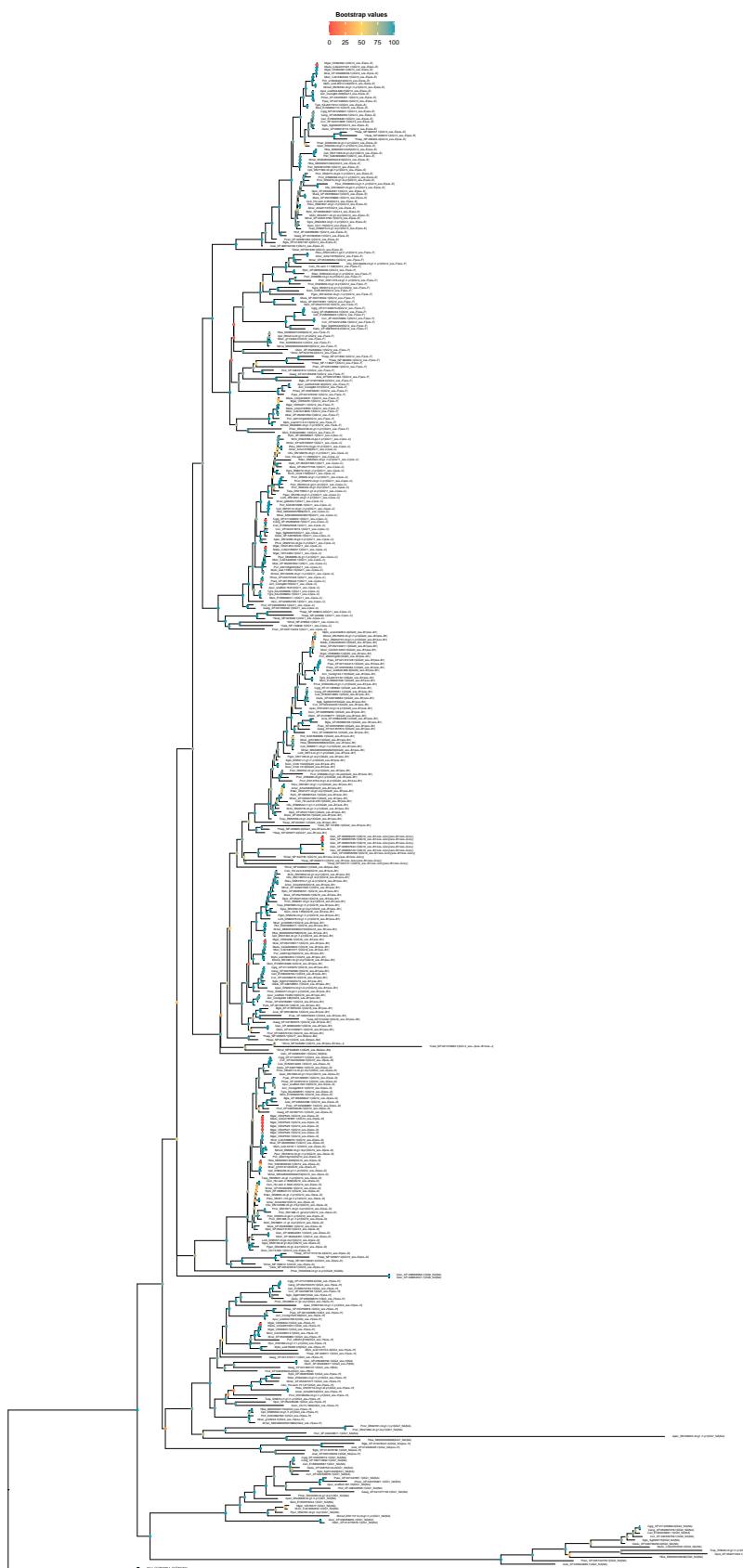
3.6 Supplementary Materials

3.6.1 Supplementary Figures

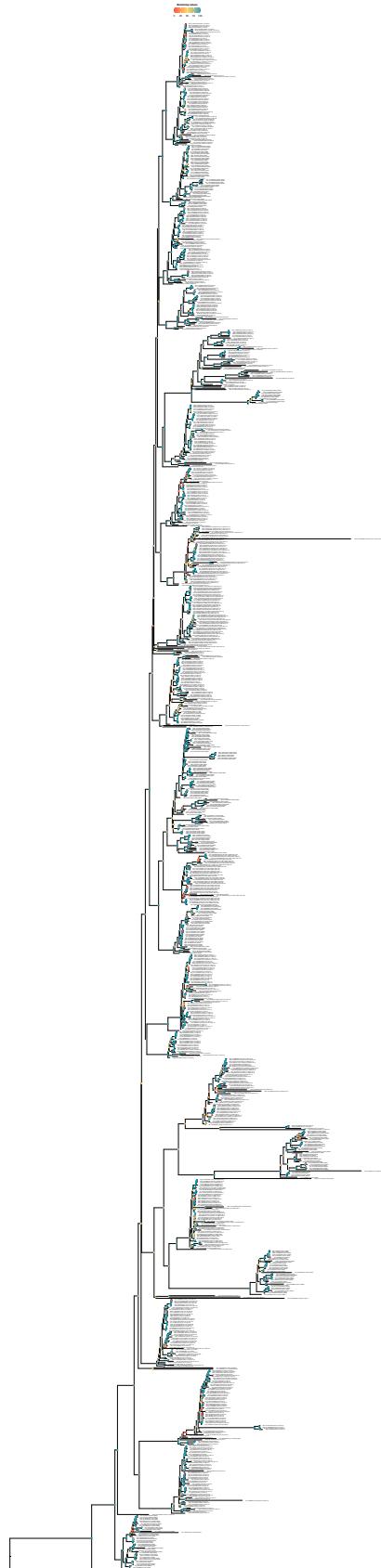
High-quality supplementary figures are available at the following GitHub repository: [LINK](#) [LINK](#) [LINK](#).



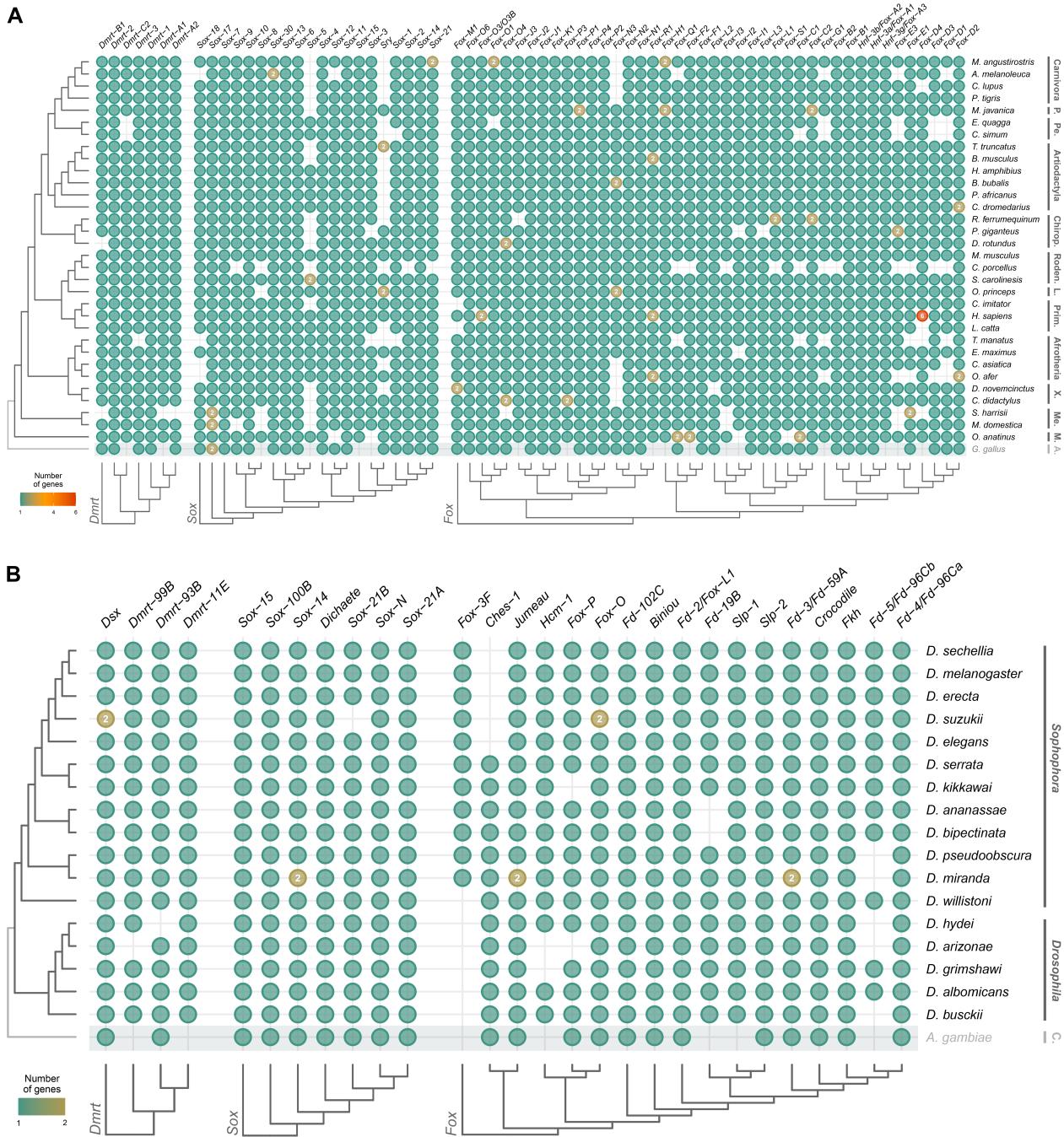
Supplementary Figure S3.1. ML phylogenetic tree of the Dmrt gene family in molluscs, including the possvm orthology inference. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in Supp. Tab. S3.1. The tree has been midpoint rooted. Bootstrap values are shown for each node.



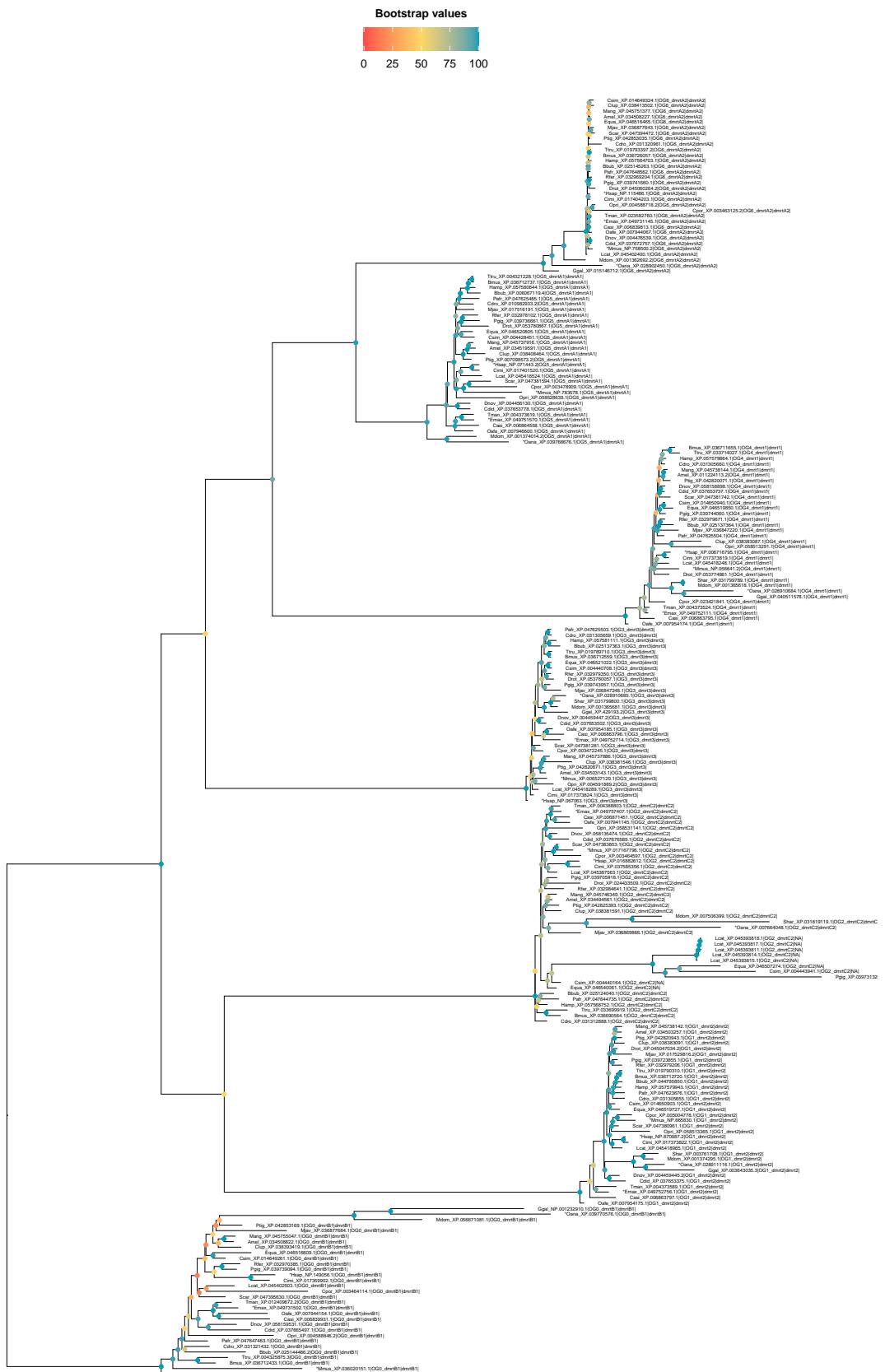
Supplementary Figure S3.2. ML phylogenetic tree of the Sox gene family in molluscs, including the possvm orthology inference. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in Supp. Tab. S3.1. Bootstrap values are shown for each node.



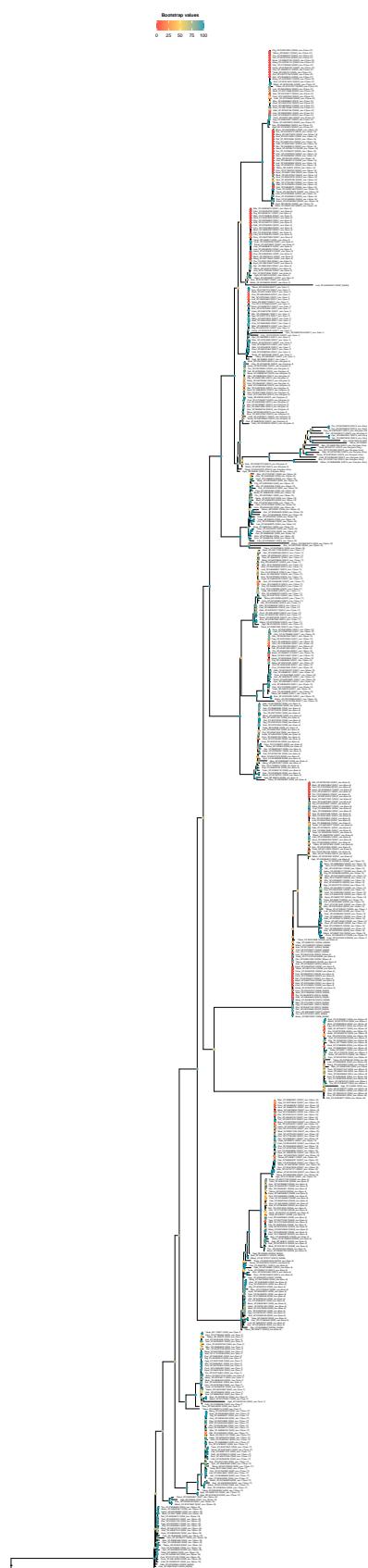
Supplementary Figure S3.3. ML phylogenetic tree of the Fox gene family in molluscs, including the possvm orthology inference. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.1**. Bootstrap values are shown for each node.



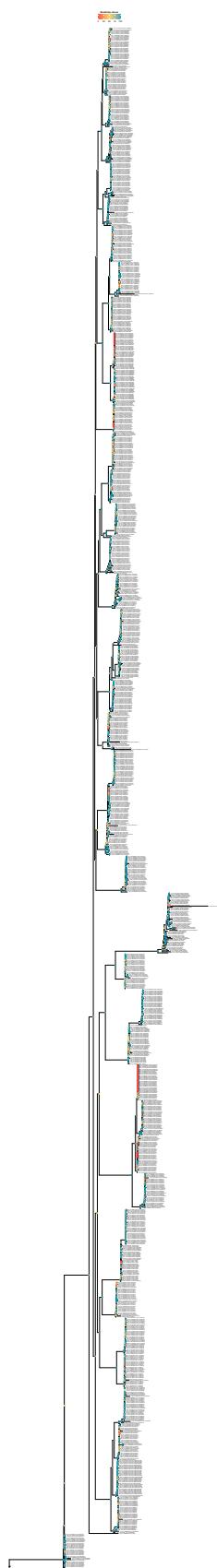
Supplementary Figure S3.4. The DSFG complement in Mammalia and *Drosophila* spp.
 Presence/absence of genes in various species are indicated by filled circles. Numbers inside each circle specify genes with 2 or more copies. The shaded area highlights outgroup species, *G. gallus* (Aves) for mammals and *A. 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. DSFG trees are shown on the bottom (full trees can be found in **Supp. Fig. S3.5–S3.7**). Full species names for both mammals and fruit flies, along with all assembly and taxonomic information, can be found in **Supp. Tab. S3.4** and **Supp. Tab. S3.5**, respectively. A.: Aves; Chiropt.: Chiroptera; L.: Lagomorpha; M.: Monotremata; Me.: Metatheria; P.: Pholidota; Pe.: Perissodactyla; Prim.: Primates; Roden.: Rodentia; X.: Xenarthra; C.: Culicidae.



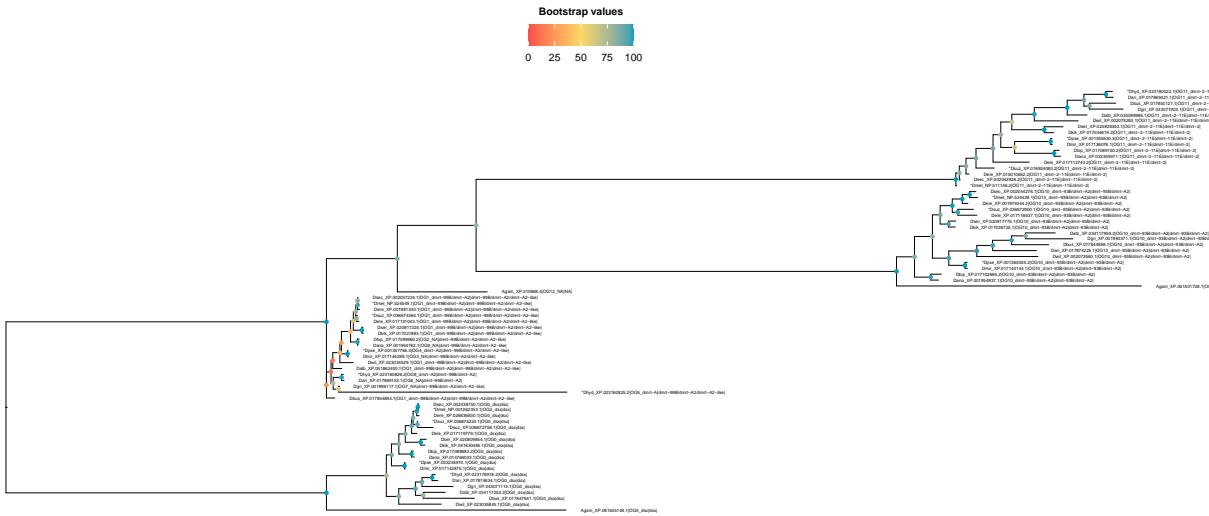
Supplementary Figure S3.5. ML phylogenetic tree of the Dmrt gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *Mus musculus*, *Elephas maximus indicus*, and *Ornithorhynchus anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.4**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



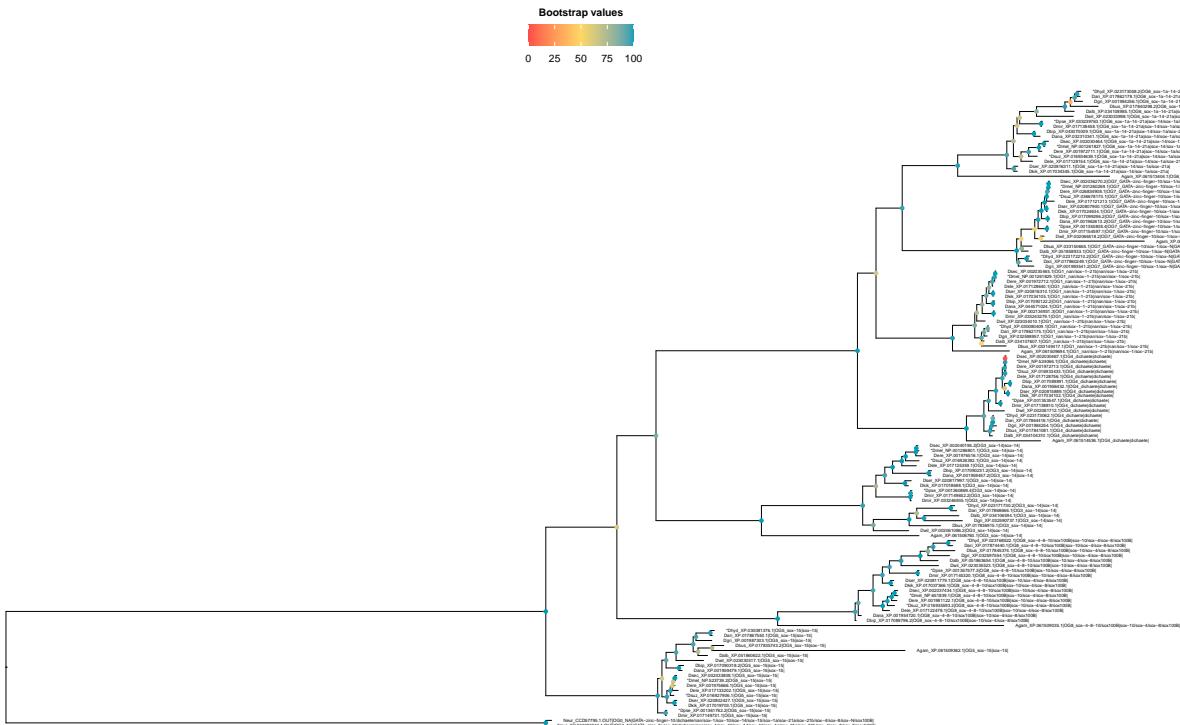
Supplementary Figure S3.6. ML phylogenetic tree of the Sox gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *M. musculus*, *E. maximus indicus*, and *O. anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.4**. Bootstrap values are shown for each node.



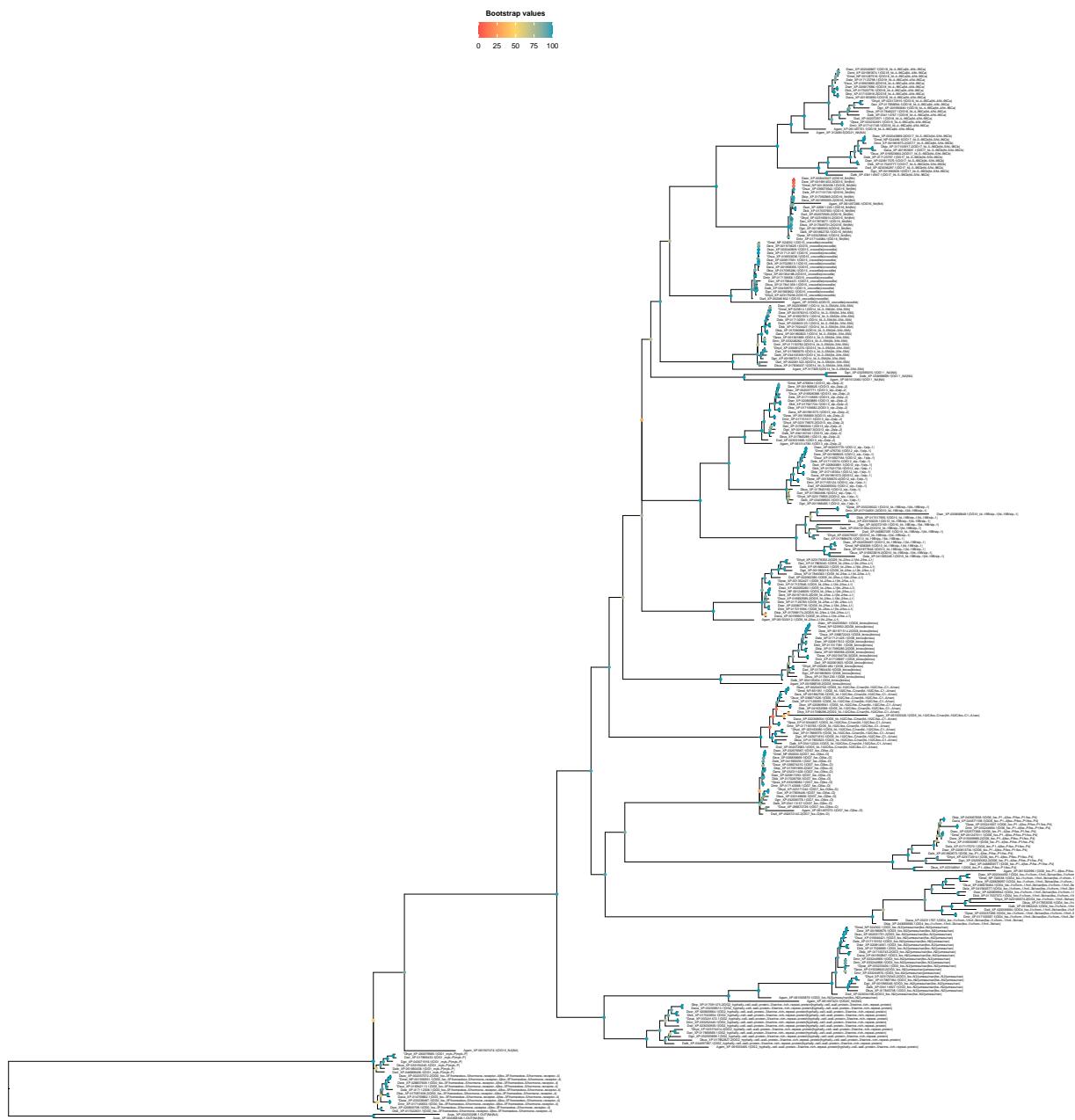
Supplementary Figure S3.7. ML phylogenetic tree of the Fox gene family in mammals, including the Possvm orthology inference. Reference genes from *H. sapiens*, *M. musculus*, *E. maximus*, *O. indicus*, and *O. anatinus* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.4**. Bootstrap values are shown for each node.



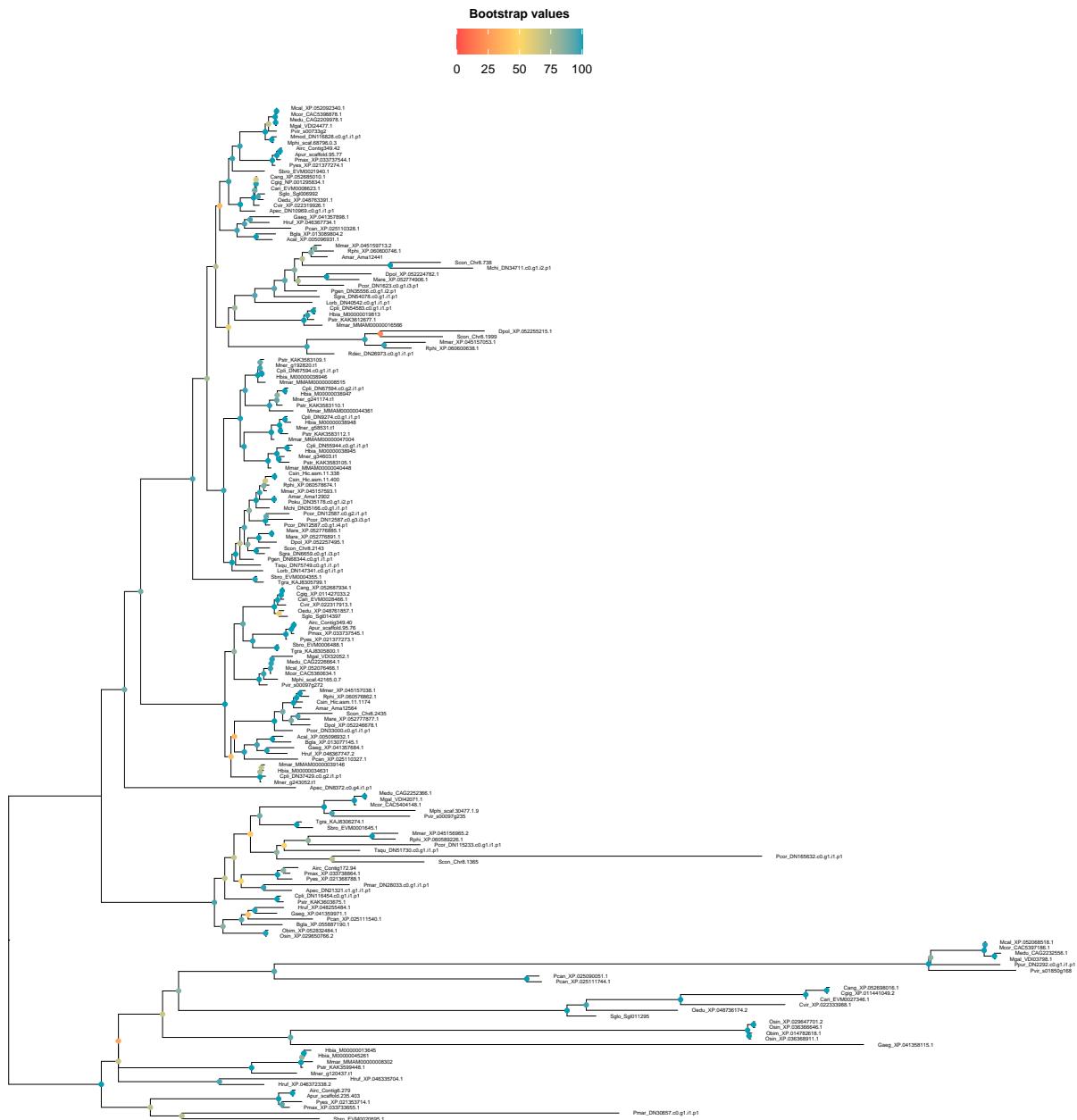
Supplementary Figure S3.8. ML phylogenetic tree of the Dmrt gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *D. hydei*, *D. pseudoobscura*, and *D. suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.5**. The tree has been midpoint rooted. Bootstrap values are shown for each node.



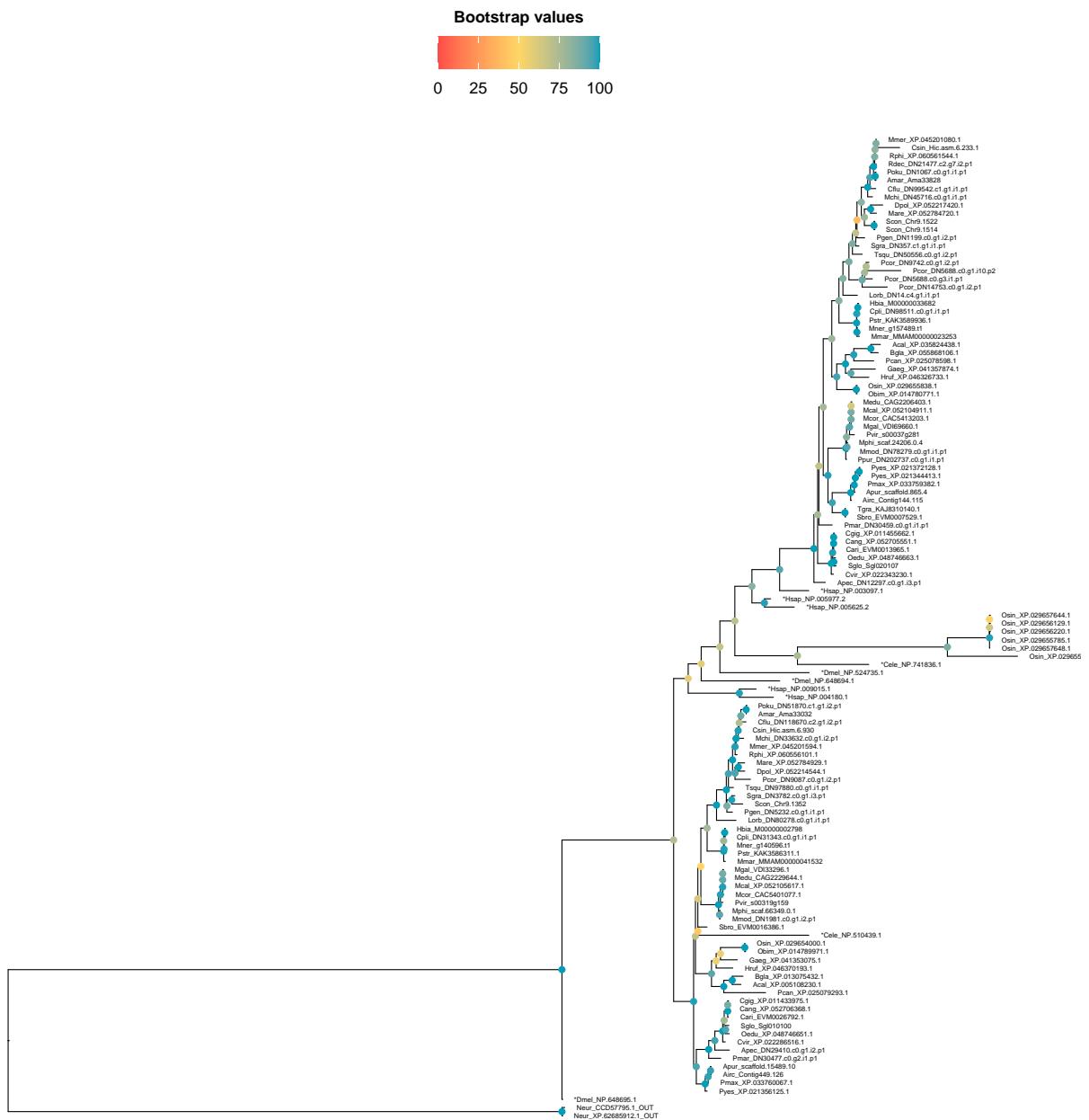
Supplementary Figure S3.9. ML phylogenetic tree of the Sox gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *D. hydei*, *D. pseudoobscura*, and *D. suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.5**. Bootstrap values are shown for each node.



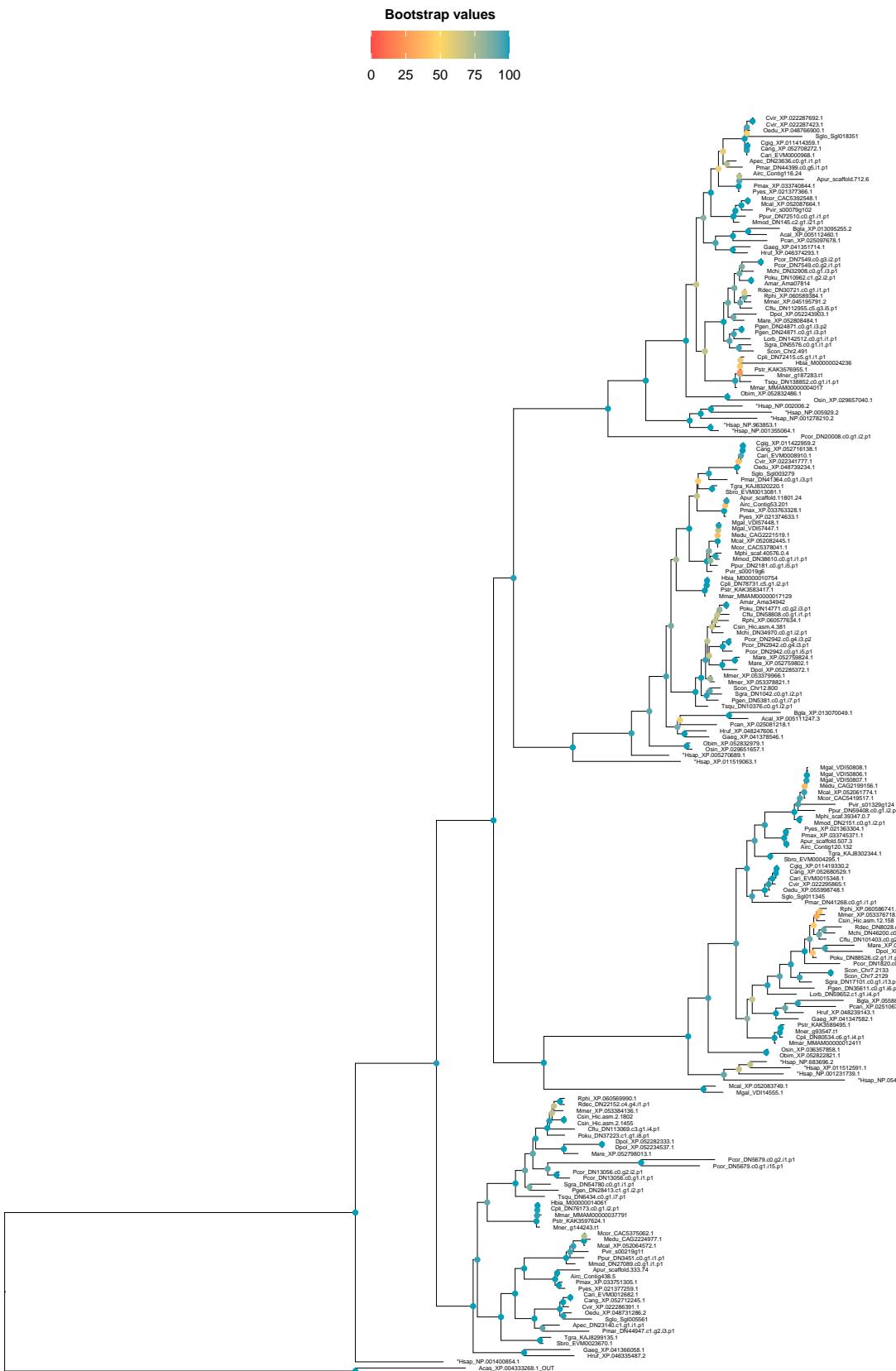
Supplementary Figure S3.10. ML phylogenetic tree of the Fox gene family in fruit flies, including the Possvm orthology inference. Reference genes from *D. melanogaster*, *D. hydei*, *D. pseudoobscura*, and *D. suzukii* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.5**. Bootstrap values are shown for each node.



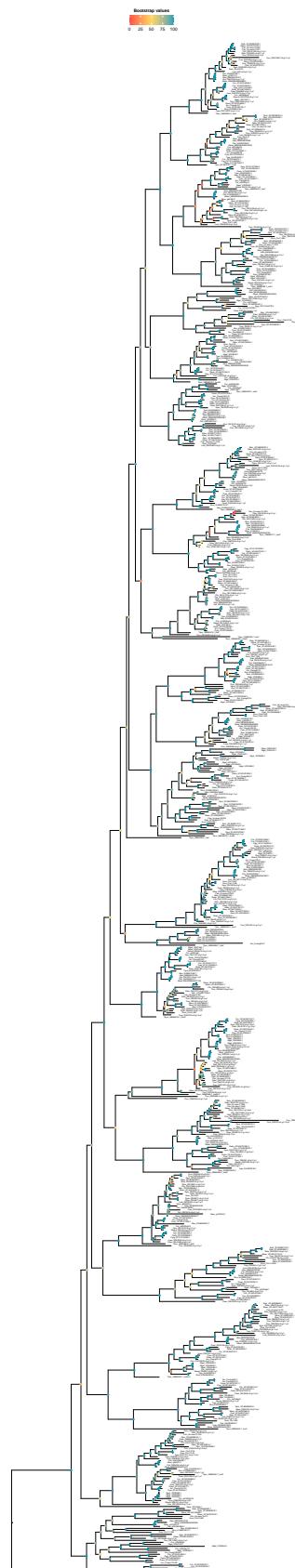
Supplementary Figure S3.11. ML phylogenetic tree of the Dmrt gene family in mollusc species. Species ID can be found in Supp. Tab. S3.1. The tree has been midpoint rooted. Bootstrap values are shown for each node.



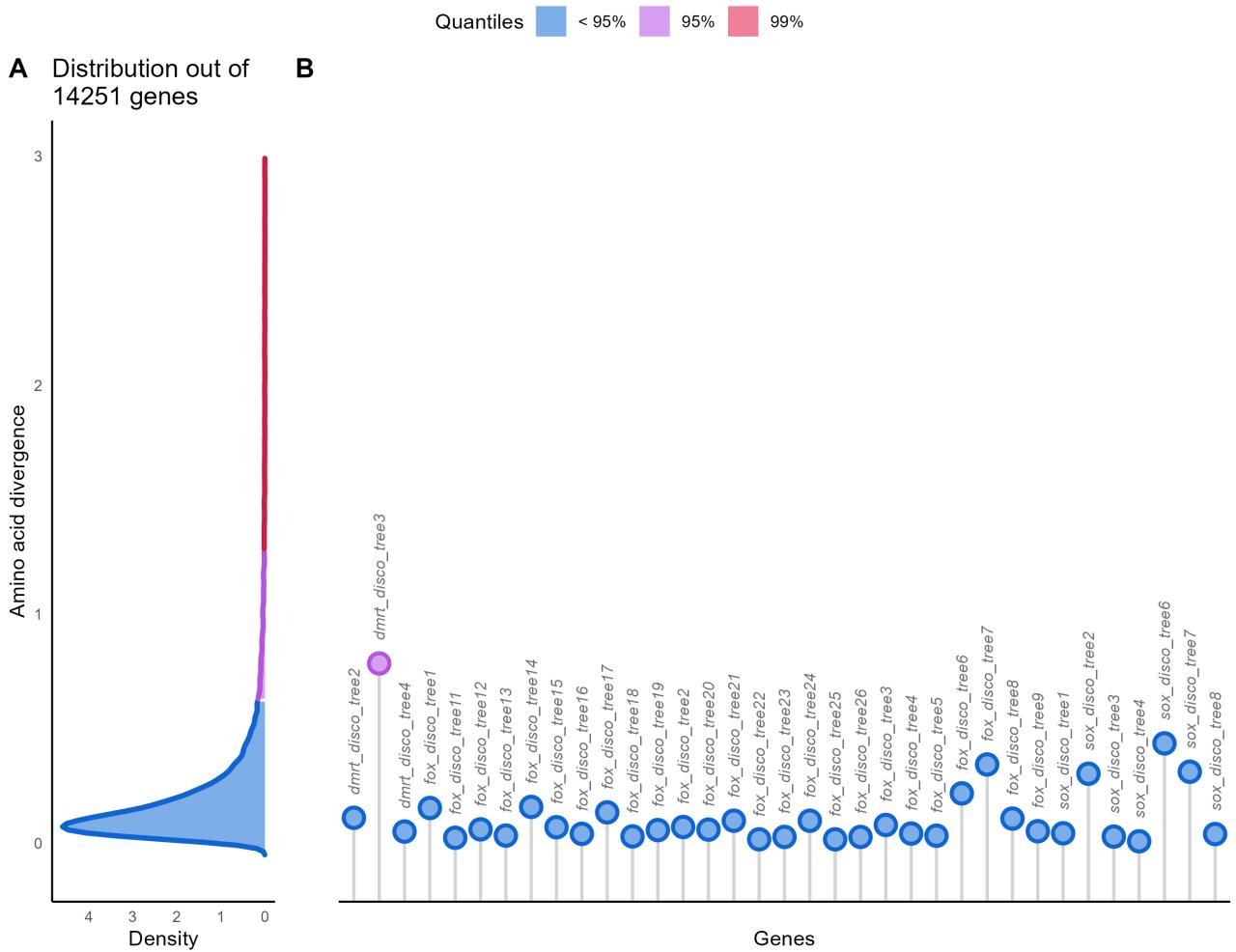
Supplementary Figure S3.12. ML phylogenetic tree of *Sox-B1* and *Sox-B2* genes in mollusc and reference species. Reference genes from *H. sapiens*, *C. elegans*, and *D. melanogaster* are marked with an asterisk at the beginning of the tip names. Species ID can be found in **Supp. Tab. S3.1**. Bootstrap values are shown for each node.



Supplementary Figure S3.13. 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 Supp. Tab. S3.1. Bootstrap values are shown for each node.



Supplementary Figure S3.14. ML phylogenetic tree of the Fox gene family in bivalves and the sea urchin *S. 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 **Supp. Tab. S3.1**. *S. purpuratus* genes are those given by **Tu et al., 2006**. Bootstrap values are shown for each node.



Supplementary Figure S3.15. Distribution of AASD of single-copy orthogroups in *C. gigas*, *C. angulata*, *C. ariakensis*, and *C. virginica* (A), including DSFG (B). The distribution of AASD in *Crassostrea* has been computed on the median values of pairwise distances of over 14k SCOs. Circle heights of DSFGs show the median value of their AASD. *Dmrt-1L* genes are indicated as ‘dmrt_disco_tree3’.

3.6.2 Supplementary Tables

All the supplementary tables are available in a parsable version at the following GitHub repository: [LINK](#) [LINK](#) [LINK](#).

Supplementary Table S3.1. Genomic and transcriptomic data of bivalves and other molluscs.

Gene family	PANTHER/CDD	ID	Description

Supplementary Table S3.2. DSFG family and domain identifiers (IDs) in PANTHER and CDD, respectively. After having retrieved putative DSFGs on the basis of HMM profiles, IDs have been used to retain only reliable hits.

Gene family	PANTHER/CDD	ID	Description
Dmrt	CDD	gnl—CDD—214606	Doublesex DNA-binding motif
	CDD	gnl—CDD—4258550	DM DNA binding domain
	PANTHER	PTHR12322	DOUBLESEX AND MAB-3 RELATED TRANSCRIPTION FACTOR DMRT PROTEIN CBR-MAB-23
	PANTHER	PTHR12322;SF115	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR 1
	PANTHER	PTHR12322;SF116	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR DMD-4
	PANTHER	PTHR12322;SF118	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR 2
	PANTHER	PTHR12322;SF123	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR A1
	PANTHER	PTHR12322;SF53	DOUBLESEX- AND MAB-3-RELATED TRANSCRIPTION FACTOR A1
	PANTHER	PTHR12322;SF71	STRESS RESPONSE PROTEIN NST1
	PANTHER	PTHR16897;SF2	RIBONUCLEASE H
Sox	CDD	gnl—CDD—432488	SOX transcription factor
	CDD	gnl—CDD—432558	Sox developmental protein N terminal
	CDD	gnl—CDD—438790	high mobility group (HMG)-box found in group B SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438820	high mobility group (HMG)-box found in sex-determining region Y (SRY)-box (SOX) family transcription factors
	CDD	gnl—CDD—438837	high mobility group (HMG)-box found in group A, group B and group G of SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438838	high mobility group (HMG)-box found in group C SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438839	high mobility group (HMG)-box found in group D SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438840	high mobility group (HMG)-box found in group E SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438841	high mobility group (HMG)-box found in group F SRY-related high-mobility group (HMG) box (Sox) transcription factors
	CDD	gnl—CDD—438842	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 30 (SOX30) and similar proteins
	CDD	gnl—CDD—438843	high mobility group (HMG)-box found in sex-determining region Y protein (SRY) and similar proteins
	CDD	gnl—CDD—438844	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 15 (SOX15) and similar proteins
	CDD	gnl—CDD—438845	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 4 (SOX4) and similar proteins
	CDD	gnl—CDD—438846	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 11 (SOX11) and similar proteins
	CDD	gnl—CDD—438847	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 12 (SOX12) and similar proteins
	CDD	gnl—CDD—438849	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 7 (SOX7) and similar proteins
	CDD	gnl—CDD—438850	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 17 (SOX17) and similar proteins
	CDD	gnl—CDD—438851	high mobility group (HMG)-box found in sex determining region Y (SRY)-box 18 (SOX18) and similar proteins
	PANTHER	PTHR10270;SF107	TRANSCRIPTION FACTOR SOX-14
	PANTHER	PTHR10270;SF161	SOX DOMAIN-CONTAINING PROTEIN DICHAETE-RELATED
	PANTHER	PTHR10270;SF199	SEX-DETERMINING REGION Y PROTEIN
	PANTHER	PTHR10270;SF231	TRANSCRIPTION FACTOR SOX-2
	PANTHER	PTHR10270;SF27	TRANSCRIPTION FACTOR SOX-4
	PANTHER	PTHR10270;SF313	TRANSCRIPTION FACTOR SOX-21
	PANTHER	PTHR10270;SF315	TRANSCRIPTION FACTOR SOX-1A-RELATED
	PANTHER	PTHR10270;SF317	TRANSCRIPTION FACTOR SOX-15-RELATED
	PANTHER	PTHR10270;SF322	TRANSCRIPTION FACTOR SOX-3
	PANTHER	PTHR10270;SF324	TRANSCRIPTION FACTOR SOX-3
	PANTHER	PTHR10270;SF326	TRANSCRIPTION FACTOR SOX

Table S3.2 continued from previous page

Gene family	PANTHER/CDD	ID	Description
Sox	PANTHER	PTHR10270	SOX TRANSCRIPTION FACTOR
	PANTHER	PTHR45789	F118025P1
	PANTHER	PTHR45789:SF2	F118025P1
	PANTHER	PTHR45803:SF1	"TRANSCRIPTION FACTOR SOX-9
	PANTHER	PTHR45803:SF2	"TRANSCRIPTION FACTOR SOX-8
	PANTHER	PTHR45803:SF5	SOX100B
	PANTHER	PTHR45803	SOX100B
	PANTHER	PTHR4729:SF1	"TRANSCRIPTION FACTOR SOX-30
	PANTHER	PTHR4729	"TRANSCRIPTION FACTOR SOX-30
	CDD	gnl—CDD—410788	Forkhead (FH) domain found in Forkhead box (FOX) family of transcription factors and similar proteins
Fox	CDD	gnl—CDD—410789	Forkhead (FH) domain found in the Forkhead box protein A (FOXA) subfamily
	CDD	gnl—CDD—410790	Forkhead (FH) domain found in the Forkhead box protein B (FOXB) subfamily
	CDD	gnl—CDD—410791	Forkhead (FH) domain found in the Forkhead box protein C (FOXC) subfamily
	CDD	gnl—CDD—410792	Forkhead (FH) domain found in the Forkhead box protein D (FOXD) subfamily
	CDD	gnl—CDD—410793	Forkhead (FH) domain found in the Forkhead box protein E (FOXE) subfamily
	CDD	gnl—CDD—410794	Forkhead (FH) domain found in the Forkhead box protein F (FOXF) subfamily
	CDD	gnl—CDD—410795	Forkhead (FH) domain found in the Forkhead box protein G (FOXG) subfamily
	CDD	gnl—CDD—410796	Forkhead (FH) domain found in the Forkhead box protein H (FOXH) subfamily
	CDD	gnl—CDD—410797	Forkhead (FH) domain found in Forkhead box protein J1 (FOXJ1) and similar proteins
	CDD	gnl—CDD—410798	Forkhead (FH) domain found in Forkhead box proteins, FOXJ2, FOXJ3 and similar proteins
	CDD	gnl—CDD—410799	Forkhead (FH) domain found in the Forkhead box protein I (FOXI) subfamily
	CDD	gnl—CDD—410800	Forkhead (FH) domain found in the Forkhead box protein K (FOXKK) subfamily
	CDD	gnl—CDD—410801	Forkhead (FH) domain found in Forkhead box protein L1 (FOXL1) and similar proteins
	CDD	gnl—CDD—410802	Forkhead (FH) domain found in Forkhead box protein L2 (FOXL2) and similar proteins
	CDD	gnl—CDD—410803	Forkhead (FH) domain found in the Forkhead box protein M (FOXM) subfamily
	CDD	gnl—CDD—410804	Forkhead (FH) domain found in Forkhead box protein N1 (FOXNL1) and similar proteins
	CDD	gnl—CDD—410805	Forkhead (FH) domain found in Forkhead box protein N2 (FOXNL2) and similar proteins
	CDD	gnl—CDD—410806	Forkhead (FH) domain found in the Forkhead box protein O (FOXO) subfamily
	CDD	gnl—CDD—410807	Forkhead (FH) domain found in the Forkhead box protein P (FOXP) subfamily
	CDD	gnl—CDD—410808	Forkhead (FH) domain found in Forkhead box protein Q1 (FOXQ1) and similar proteins
	CDD	gnl—CDD—410809	Forkhead (FH) domain found in Forkhead box protein Q2 (FOXQ2) and similar proteins
	CDD	gnl—CDD—410810	Forkhead (FH) domain found in the Forkhead box protein R (FOXR) subfamily
	CDD	gnl—CDD—410811	Forkhead (FH) domain found in the Forkhead box protein S1 (FOXS1)
	CDD	gnl—CDD—410812	Forkhead (FH) domain found in Forkhead box protein A1 (FOXA1) and similar proteins
	CDD	gnl—CDD—410813	Forkhead (FH) domain found in Forkhead box protein A2 (FOXA2) and similar proteins
	CDD	gnl—CDD—410814	Forkhead (FH) domain found in Forkhead box protein A3 (FOXA3) and similar proteins
	CDD	gnl—CDD—410816	Forkhead (FH) domain found in Forkhead box protein B1 (FOXB1) and similar proteins
	CDD	gnl—CDD—410817	Forkhead (FH) domain found in Forkhead box protein B2 (FOXB2) and similar proteins
	CDD	gnl—CDD—410818	Forkhead (FH) domain found in Forkhead box protein C1 (FOXC1) and similar proteins
	CDD	gnl—CDD—410819	Forkhead (FH) domain found in Forkhead box protein C2 (FOXC2) and similar proteins
	CDD	gnl—CDD—410820	Forkhead (FH) domain found in Forkhead box proteins FOXD1, FOXD2 and similar proteins
	CDD	gnl—CDD—410821	Forkhead (FH) domain found in Forkhead box protein D3 (FOXD3) and similar proteins

Table S3.2 continued from previous page

Gene family	PANTHER/CDD	ID	Description
CDD	gnl—CDD—410822	Forkhead (FH) domain found in Forkhead box protein D4 (FOXD4) and similar proteins	
CDD	gnl—CDD—410823	Forkhead (FH) domain found in Forkhead box protein F1 (FOXF1) and similar proteins	
CDD	gnl—CDD—410824	Forkhead (FH) domain found in Forkhead box protein F2 (FOXF2) and similar proteins	
CDD	gnl—CDD—410825	Forkhead (FH) domain found in Forkhead box protein J2 (FOXJ2) and similar proteins	
CDD	gnl—CDD—410826	Forkhead (FH) domain found in Forkhead box protein J3 (FOXJ3) and similar proteins	
CDD	gnl—CDD—410827	Forkhead (FH) domain found in Forkhead box protein II (FOXII) and similar proteins	
CDD	gnl—CDD—410828	Forkhead (FH) domain found in Forkhead box protein K1 (FOXK1) and similar proteins	
CDD	gnl—CDD—410829	Forkhead (FH) domain found in Forkhead box protein K2 (FOXK2) and similar proteins	
CDD	gnl—CDD—410830	Forkhead (FH) domain found in Forkhead box protein N1 (FOXN1)	
CDD	gnl—CDD—410831	Forkhead (FH) domain found in Forkhead box protein N4 (FOXN4)	
CDD	gnl—CDD—410832	Forkhead (FH) domain found in Forkhead box protein N2 (FOXN2)	
CDD	gnl—CDD—410833	Forkhead (FH) domain found in Forkhead box protein N3 (FOXN3)	
CDD	gnl—CDD—410834	Forkhead (FH) domain found in Forkhead box protein O1 (FOXO1)	
CDD	gnl—CDD—410835	Forkhead (FH) domain found in Forkhead box protein O3 (FOXO3)	
CDD	gnl—CDD—410836	Forkhead (FH) domain found in Forkhead box protein O4 (FOXO4) and similar proteins	
CDD	gnl—CDD—410837	Forkhead (FH) domain found in Forkhead box protein O6 (FOXO6) and similar proteins	
CDD	gnl—CDD—410838	Forkhead (FH) domain found in Forkhead box protein P1 (FOXP1)	
CDD	gnl—CDD—410839	Forkhead (FH) domain found in Forkhead box protein P2 (FOXP2)	
CDD	gnl—CDD—410840	Forkhead (FH) domain found in Forkhead box protein P3 (FOXP3) and similar proteins	
CDD	gnl—CDD—410841	Forkhead (FH) domain found in Forkhead box protein P4 (FOXP4) and similar proteins	
PANTHER	PTHR11829	FORKHEAD BOX PROTEIN	
PANTHER	PTHR11829:SF142	FOXC22 PROTEIN	
PANTHER	PTHR11829:SF156	FORKHEAD BOX PROTEIN E3	
PANTHER	PTHR11829:SF206	FORKHEAD BOX PROTEIN Q1	
PANTHER	PTHR11829:SF209	FORKHEAD BOX PROTEIN B1	
PANTHER	PTHR11829:SF335	FORKHEAD BOX PROTEIN D2	
PANTHER	PTHR11829:SF340	FORKHEAD BOX PROTEIN H1	
PANTHER	PTHR11829:SF342	FORKHEAD BOX PROTEIN L2	
PANTHER	PTHR11829:SF348	FORKHEAD BOX PROTEIN D1	
PANTHER	PTHR11829:SF361	FORKHEAD BOX PROTEIN D3	
PANTHER	PTHR11829:SF398	FORKHEAD BOX PROTEIN PES-1	
PANTHER	PTHR11829:SF399	FORKHEAD TRANSCRIPTION FACTOR FKH-9	
PANTHER	PTHR11829:SF401	FORKHEAD BOX C1-A-RELATED	
PANTHER	PTHR13962	FORKHEAD BOX PROTEIN N3-LIKE PROTEIN-RELATED	
PANTHER	PTHR13962:SF17	FORKHEAD BOX PROTEIN N4	
PANTHER	PTHR13962:SF19	FORKHEAD BOX PROTEIN N2	
PANTHER	PTHR13962:SF20	FORKHEAD BOX PROTEIN N3	
PANTHER	PTHR13962:SF22	FORKHEAD BOX PROTEIN N3-LIKE PROTEIN	
PANTHER	PTHR13962:SF26	FORKHEAD BOX PROTEIN N2	
PANTHER	PTHR45767	FORKHEAD BOX PROTEIN O	
PANTHER	PTHR45767:SF2	FORKHEAD BOX PROTEIN O	
PANTHER	PTHR45796	FORKHEAD BOX P, ISOFORM C	
PANTHER	PTHR45796:SF3	FORKHEAD BOX PROTEIN P1	

Table S3.2 continued from previous page

Gene family	PANTHER/CDD	ID	Description
Panther	PANTHER	PTHR45796:SF4	FORKHEAD BOX P, ISOFORM C
	PANTHER	PTHR45881:SF3	FORKHEAD BOX PROTEIN K2
	PANTHER	PTHR45881:SF4	FORKHEAD BOX PROTEIN K1
	PANTHER	PTHR46078	FORKHEAD BOX PROTEIN J2 FAMILY MEMBER
	PANTHER	PTHR46262	FORKHEAD BOX PROTEIN BINIOU
	PANTHER	PTHR46262:SF2	FORKHEAD BOX PROTEIN BINIOU
	PANTHER	PTHR46617	FORKHEAD BOX PROTEIN G1
	PANTHER	PTHR46617:SF3	FORKHEAD BOX PROTEIN G1
	PANTHER	PTHR46721	FORKHEAD BOX PROTEIN N1
	PANTHER	PTHR46721:SF2	FORKHEAD BOX N1
Fox	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 S3.3. 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)	-	B2
<i>SOX21</i> (NP_009015.1)	<i>sox21b</i> (NP_648695.1)		
<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/FKHL7</i> (NP_001444.2)			
<i>FOXC2/MFH1</i> (NP_005242.1)	<i>crocodile/fd1</i> (NP_524202.1)	-	C

Table S3.3 continued from previous page

<i>Homo sapiens</i>	<i>Drosophila melanogaster</i>	<i>Caenorhabditis elegans</i>	Group
Fox gene family			
<i>FOXD1/FREAC4</i> (<i>NP_004463.1</i>)			
<i>FOXD2/FREAC9</i> (<i>NP_004465.3</i>)	<i>fd59A/fd3</i> (<i>NP_523814.1</i>)	<i>unc-130</i> (<i>NP_496411.1</i>)	D
<i>FOXD3</i> (<i>NP_036315.1</i>)			
<i>FOXD4</i> (<i>NP_997188.2</i>)			
<i>FOXE1/TITF2</i> (<i>NP_004464.2</i>)	-	-	E
<i>FOXE3</i> (<i>NP_036318.1</i>)			
<i>FOXF1</i> (<i>NP_001442.2</i>)	<i>binious/FoxF</i> (<i>NP_523950.2</i>)	<i>let-381/F26B1.7</i> (<i>NP_491826.1</i>)	F
<i>FOXF2</i> (<i>NP_001443.1</i>)			
	<i>slp1</i> (<i>NP_476730.1</i>)		
<i>FOXG1/BF1/HBF2</i> (<i>NP_005240.3</i>)	<i>slp2</i> (<i>NP_476834.1</i>)	<i>fkh2/T14G12.4</i> (<i>NP_508644.1</i>)	G
	<i>fd19B/cg9571</i> (<i>NP_608369.1</i>)		
<i>FOXH1/FAST1</i> (<i>NP_003914.1</i>)	-	-	H
<i>FOXI1/FREAC6/HFH3</i> (<i>NP_036320.2</i>)	-	-	I
<i>FOXJ1</i> (<i>NP_001445.2</i>)	-	-	J1
<i>FOXJ2</i> (<i>XP_011519063.1</i>)	-	-	J2
<i>FOXJ3</i> (<i>XP_005270689.1</i>)	-	-	J3
<i>FOXK1/ILF1</i> (<i>NP_001032242.1</i>)	<i>foxK/LD16137</i> (<i>NP_001261701.1</i>)	-	K
<i>FOXK2</i> (<i>NP_004505.2</i>)			
<i>FOXL1</i> (<i>NP_005241.1</i>)	<i>foxL1/fd2</i> (<i>NP_523912.1</i>)	-	L1
<i>FOXL2</i> (<i>NP_075555.1</i>)	-	-	L2
<i>FOXM1</i> (<i>NP_001400854.1</i>)	-	-	M
<i>FOXN1/WHN</i> (<i>NP_001356298.1</i>)	<i>jumeau</i> (<i>NP_524302.1</i>)	-	N1/4
<i>FOXN4</i> (<i>NP_998761.2</i>)			
<i>FOXN2/HTLF</i> (<i>NP_001362376.1</i>)	<i>ches-1</i> (<i>NP_511071.3</i>)	-	N2/3
<i>FOXN3/CHES1</i> (<i>NP_001078940.1</i>)			
<i>FOXO1</i> (<i>NP_002006.2</i>)			
<i>FOXO3</i> (<i>NP_963853.1</i>)	-	<i>daf-16</i> (<i>NP_001364785.1</i>)	O
<i>FOXO3B</i> (<i>NP_001355064.1</i>)			
<i>FOXP1</i> (<i>NP_001231739.1</i>)			
<i>FOXP2</i> (<i>NP_683696.2</i>)	<i>foxP/cg16899</i> (<i>NP_001247011.1</i>)	<i>F26D12.1</i> (<i>NP_001293813.1</i>)	P
<i>FOXP3</i> (<i>NP_054728.2</i>)			
<i>FOXP4</i> (<i>XP_011512591.1</i>)			
<i>FOXQ/HFH11</i> (<i>NP_150285.3</i>)	-	-	Q1
-	<i>fd102C/cd11152</i> (<i>NP_651951.1</i>)	<i>fkh-10/C25A1.2</i> (<i>NP_492676.2</i>)	Q2
<i>FOXSI/FREAC10</i> (<i>NP_004109.1</i>)	-	-	S
-	-	<i>PES-1</i> (<i>NP_001023406.1</i>)	-
-	-	<i>B0286.5/FKH-6</i> (<i>NP_494775.1</i>)	-
-	-	<i>F40H3.4/FKH-8</i> (<i>NP_001254107.1</i>)	-
-	-	<i>C29F7.4/FKH-3</i> (<i>NP_001294822.1</i>)	-
-	-	<i>K03C7.2/FKH-9</i> (<i>NP_001024760.1</i>)	-

Supplementary Table S3.4. Genomic data of mammals 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	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	Afroscoricia	Genome	C:98.0%[S:97.4%,D:0.6%],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:0.6%],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>Oryctopus 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	N/A
<i>Ochetona princeps</i>	Opri	Mammalia	Euarchontoglires	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	Euarchontoglires	Primates	Genome	C:97.3%[S:95.1%,D:2.2%],F:1.7%,M:1.0%	GCF_0011604975.1	Orkin et al., 2021
<i>Homo sapiens</i>	Hsap	Mammalia	Euarchontoglires	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	Euarchontoglires	Primates	Genome	C:98.3%[S:97.2%,D:1.1%],F:0.4%,M:1.3%	GCF_020740605.2	Vertebrate Genome Project
<i>Cavia porcellus</i>	Cpor	Mammalia	Euarchontoglires	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	Euarchontoglires	Rodentia	Genome	C:99.4%[S:98.7%],D:0.7%,F:0.2%,M:0.4%	GCF_000001635.27	Genome Reference Consortium
<i>Scirurus carolinensis</i>	Scar	Mammalia	Euarchontoglires	Rodentia	Genome	C:99.1%[S:96.9%],D:2.2%,F:0.3%,M:0.6%	GCF_902686445.1	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_0199233935.1	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>	Pafu	Mammalia	Laurasiatheria	Artiodactyla	Genome	C:98.8%[S:98.3%],D:0.5%,F:0.6%,M:0.6%	GCF_016906955.1	N/A
<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_011762395.1	Xiong et al., 2009
<i>Ailuropoda melanoleuca</i>	Amel	Mammalia	Laurasiatheria	Carnivora	Genome	C:97.3%[S:96.6%],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_011100855.1	Wang et al., 2021
<i>Mirounga angustirostris</i>	Mang	Mammalia	Laurasiatheria	Carnivora	Genome	C:96.7%[S:94.5%],D:2.2%,F:1.9%,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_022682495.1	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	Fourret 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	N/A
<i>Equus quagga</i>	Equa	Mammalia	Laurasiatheria	Perissodactyla	Genome	C:98.5%[S:95.0%],D:3.5%,F:0.5%,M:1.0%	GCF_021613505.1	Vibstrup 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_014570355.1	N/A
<i>Sarcophilus harrisi</i>	Shar	Mammalia	Metatheria	Dasyuromorphia	Genome	C:95.5%[S:94.5%],D:1.0%,F:0.9%,M:3.6%	GCF_902635055.1	Stammnitz 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>Ornithorhynchus 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	zhou2021platypus
<i>Dasypus novemcinctus</i>	Dnow	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_015220355.1	Vertebrate Genome Project

Supplementary Table S3.5. 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>		Culicidae	Celicia	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	Mahaian 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 S3.6. Complete set of DSFGs in bivalves. For each gene, the species ID (Sp. ID) as in Supp. Tab. S3.1, the accession number (Gene ID), the Possvm-based annotation, and the CDD domains (including their Pssm-ID) are indicated.

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Airc	Contig6.279	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Apur	sccaffold_235_403	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Cang	XP_052698016..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Cari	EVM0027346..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Cgig	XP_011441049..2	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Cvir	XP_022333988..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Gaeg	XP_041358115..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Hbia	M00000013645	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Hbia	M00000045261	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Hruf	XP_046372338..2	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Hruf	XP_046335704..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Mcal	XP_052068518..1	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Mcor	CAC5397186..1	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Medu	CAG2232556..1	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Mgal	VDI03798..1	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Mmar	MuAM00000008302	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Mner	g120437..t1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Obim	XP_014782618..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Oedu	XP_048736174..2	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Osin	XP_036368911..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Osin	XP_036366646..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Osin	XP_029647701..2	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pcan	XP_025090051..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pcan	XP_025111744..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pmar	DN30657..e0..g1..i1..p1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pmax	XP_033733655..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Ppur	DN2992..c0..g1..i1..p1	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pstr	KAK3599448..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pvir	s01850g168	Dmrt	Dmrt-OG4..NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Pyes	XP_021353714..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Sbro	EVM0020695..1	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Sglo	Sg011295	Dmrt	Dmrt-OG0/NA	Doublesex DNA-binding motif (2146:06)	N/A	Annotated as Dmrt-1L
Airc	Contig172..94	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	N/A	-
Apec	DN21321..c1..g1..i1..p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	CUE-like DMA domain (270553)	-
Bbla	XP_055887190..1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	CUE-like DMA domain (270553, 270601, 270602)	-
Cpli	DN116454..c0..g1..i1..p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	CUE-like DMA domain (270553)	-
Gaeg	XP_041359971..1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	CUE-like DMA domain (270553, 270601, 270602)	-
Hruf	XP_048255484..1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	CUE-like DMA domain (270553, 270602, 270601)	-
Mcor	CAC5404148..1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	N/A	-
Medu	CAG2252366..1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (2146:06)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mgal	VD142071.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Mmer	XP_045156965.2	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Mphi	scat_304771.9	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Obim	XP_052832484.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553)	-
Osin	XP_029650766.2	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553)	-
Pcan	XP_025111540.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pcor	DN165632.c0.g1.i1.p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pcor	DN115233.c0.g1.i1.p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pmar	DN28033.c0.g1.i1.p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pmax	XP_03373864.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553)	-
Pstr	KAK3603675.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pvir	s00097235	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Pyes	XP_021368788.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553)	-
Rphi	XP_060589226.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Sbro	EVM0001645.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Scon	Chr8.1365	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Tgra	KAJ8306274.1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Tsqu	DN51730.c0.g1.i1.p1	Dmrt	Dmrt-2	Doublesex DNA-binding motif (214606)	N/A	-
Acal	XP_005096932.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Airc	Contig349.40	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Amar	Amal2564	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Apur	scatfield.95.76	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Bbla	XP_013077145.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Cang	XP_052687934.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Cari	EVM0028466.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Cgig	XP_011427033.2	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Cpli	DN37429.c0.g2.i1.p1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	N/A	-
Csin	Hic.asm.11.1174	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Cvir	XP_02237913.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Dpol	XP_052246678.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Gaeg	XP_041357684.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Hbia	M0000034631	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270602)	-
Hruf	XP_046367747.2	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270602)	-
Mare	XP_052777877.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mcal	XP_052076466.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mcor	CAC5360634.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Medu	CAG226664.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mgal	VD132052.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mmar	MMAM0000039146	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mmer	XP_045157038.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Mner	g243052.t1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270602)	-
Mphi	scat42165.0.7	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-
Oedu	XP_048761857.1	Dmrt	Dmrt-3	Doublesex DNA-binding motif (214606)	CUE-like DMA domain (270553, 270601, 270602, 270600)	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pcan	XP_025110327_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602)
Pcor	DN33000.c0.g1.i1.p1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Pmax	XP_033737545_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Pvir	s000973272	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Pyes	XP_021377273_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Rphi	XP_060576862_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Sbro	EVM0006488_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602)
Scon	Chr8:2435	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Sglo	Sg1014397	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602, 270600)
Tgra	KAJ8305800_1	Dmrt	Dmrt-3	Dmrt-3	Dmrt-3	CUE-like DMA domain (270553, 270601, 270602)
Acal	XP_005096931_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Airc	Contig349:42	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602, 270465)
Amar	Ama12441	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601)
Amar	Ama12902	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602)
Apec	DN10969.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602)
Apur	scatfold_95_77	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Bgla	XP_013089804_2	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Cang	XP_052685010_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cari	EVM000086233_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cgig	NP_001295834_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cpli	DN67594.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cpli	DN55944.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cpli	DN54583.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cpli	DN9274.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Cpli	DN67594.c0.g2.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Csin	Hic.asm.11.400	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Csin	Hic.asm.11.338	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Cvir	XP_022315926_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Dpol	XP_052255215_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Dpol	XP_0522224782_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Dpol	XP_052257495_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Gaeg	XP_041357898_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270466)
Hbia	M00000038945	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Hbia	M00000038948	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Hbia	M00000038946	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Lorb	DN147341.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Hbia	M00000018813	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Hbia	M00000038947	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Hruf	XP_046367734_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Lorb	DN40542.c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mare	XP_052776891_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mare	XP_052774906_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mare	XP_052776885_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mcal	XP_052092340_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mchi	DN35166_c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mcor	CAc5398878_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Medu	CAGP2039978_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mgal	VDI124477_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mmar	MMAM00000040448	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mmar	MMAM00000008515	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mmar	MMAM0000016566	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mmar	MMAM0000047004	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mmar	MMAM0000044361	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Mmer	XP_045157593_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mmer	XP_045157053_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Mmer	XP_045159713_2	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mmod	DN116828_c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mner	g34603.t1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Mner	g58531.t1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Mner	g241174.t1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mner	g192820.t1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Mphi	scaf.68796.0.3	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Oedu	XP_04876391_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Pcan	XP_025110328_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Pcor	DN12587_c0.g3.i3.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pcor	DN12587_c0.g2.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Pcor	DN1623_c0.g1.i3.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pcor	DN12587_c0.g1.i4.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Pgen	DN68344_c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Pgen	DN35556_c0.g1.i2.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pmax	XP_033737544_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270465)
Poku	DN35178_c0.g1.i2.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Pstr	KAK3612677_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pstr	KAK3583105_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Pstr	KAK3583112_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Pstr	KAK3583110_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pstr	KAK3583109_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Pvir	s00733g2	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Pyes	XP_021377274_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602, 270465)
Rdec	DN26973_c0.g1.i1.p1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Rphi	XP_066060638_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)
Rphi	XP_066060746_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A
Sbro	EVM0004355_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Sbro	EVM0021940_1	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)
Scon	Chr8.1999	Dmrt	Dmrt-4/5	Dmrt-4/5	Dmrt-4/5	N/A

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Scon	Chr8:2143	Dmrt	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602)	-
Sglo	SgI006392	Dmrt	Dmrt	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600, 270602, 270465)	-
Sgra	DN54078_c0.g1.i1.p1	Dmrt	Dmrt	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)	-
Sgra	DN6659_c0.g1.i3.p1	Dmrt	Dmrt	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270600)	-
Tgra	KAJ8307799.1	Dmrt	Dmrt	Dmrt-4/5	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602)
Tsqu	DN75749_c0.g1.i1.p1	Dmrt	Dmrt	Dmrt-4/5	Dmrt-4/5	N/A
Apec	DN5372_c0.g4.i1.p1	Dmrt	N/A	Dmrt	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602)
Mchi	DN34711_c0.g1.i2.p1	Dmrt	N/A	N/A	N/A	N/A
Scon	Chr8:738	Dmrt	N/A	Dmrt	Dmrt-4/5	CUE-like DMA domain (270553, 270601, 270602)
Acal	XP_005097243.2	Fox	Fox-A	Fox	Fox-A	Forkhead N-terminal (369872); Forkhead C-terminal domain (430552)
Aire	Contig6.157	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Amar	Arna08751	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Apec	DN107972_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Apur	scaffold.124.7	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Bglu	XP_013067134.2	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cang	XP_052701295.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cari	EVM0004613.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Chu	DN101169_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cgig	XP_011413445.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cpli	DN47094_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Csin	Hic.asm.10.638	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Csin	Hic.asm.10.437	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cvir	XP_022333332.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Cvir	XP_022334050.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Dpol	XP_052272379.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Gaeg	XP_041352454.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Hbia	M00000018167	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Hruf	XP_046371021.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mare	XP_052765228.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Meal	XP_052106467.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mchi	DN23553_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mcor	CAc5374046.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Medu	CAG2201348.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mgal	VDI17457.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mmar	MMAM0000008663	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mmer	XP_045173733.1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mmod	DN103780_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mner	g192217.t1	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Mphi	scat.4682.0.0	Fox	Fox-A	Fox	Fox-A	Forkhead domain A1 (410812)
Obim	XP_014785201.2	Fox	Fox-A	Fox	Fox-A	Forkhead N-terminal (369872); Forkhead C-terminal domain (430552)
Oedu	XP_048735259.1	Fox	Fox-A	Fox	Fox-A	Forkhead N-terminal (369872); Forkhead C-terminal domain (430552)
Pcan	XP_025090786.1	Fox	Fox-A	Fox	Fox-A	Forkhead N-terminal (369872); Forkhead C-terminal domain (430552)
Pcor	DN3042_c0.g1.i1.p1	Fox	Fox-A	Fox	Fox-A	Forkhead N-terminal (369872); Forkhead C-terminal domain (430552)

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pgen	DN174637.c0.g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872; partial); Forkhead C-terminal (369872);	-
Pmar	DN30866.c0.g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872);	-
Pmax	XP_033734080..1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Pstr	KAK3597847.1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552; partial)	-
Pvir	s00068g447	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Pyes	XP_021361791..1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Rphi	XP_060590755..1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Sbro	EVM0003194..1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872);	-
Scon	Chr4.24670	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Sglo	SgI008464	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872); HNF3 C-terminal domain (430552)	-
Sgra	DN78052.c0.g1.i1.p1	Fox	Fox-A	Forkhead domain A1 (410812)	Forkhead N-terminal (369872; partial); HNF3 C-terminal domain (430552)	-
Acal	XP_005089018..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Airc	Contig636..38	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Apur	scaffold..313..50	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Bgla	XP_013078204..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Cang	XP_052700333..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Cari	EVM0003536..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Chu	DN98613.c0.g1.i1.p1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Cgig	XP_011445364..2	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Csin	Hic_asm..16..347	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Cvir	XP_0222334612..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Dpol	XP_052233250..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Dpol	XP_052256324..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Dpol	XP_052281977..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Gaeg	XP_041361159..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Hbia	M0000029836	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Hruf	XP_046358590..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Lorb	DN5589.c3.g1.i1.p1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mare	XP_052791461..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mcal	XP_052100219..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mcor	CAC5382565..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Medu	CAG2229716..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mgal	V1D169670..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mmar	MMAM0000015629	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mmer	XP_045215505..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mner	g250725.t1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Mphi	scat.10920..0..0	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Obim	XP_052832317..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Oedu	XP_048732871..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Osin	XP_029653697..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Pcan	XP_02507261..1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Pcor	DN23979.c0.g1.i1.p1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-
Pcor	DN23979.c1.g1.i1.p1	Fox	Fox-B	Forkhead domain B2 (410812)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pmax	XP_033749587_1	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Pstr	KAK3607900_1	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Pvir	s001895215	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Pyes	XP_021357620_1	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Rphi	XP_060564412_1	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Scon	Chr5_12	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Sglo	Sg012012	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Tgra	KAJ8304921_1	Fox	Fox-B	Forkhead domain B2 (410817)	N/A	-
Acal	XP_005106277_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Airc	Contig58_63	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Amar	Amar17094	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Apur	scaffold_577_50	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Bbla	XP_055890240_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Cang	XP_052715579_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Cari	EVM0002771_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Chu	DN96576_c0.g1.i1.p1	Fox	Fox-C	Forkhead domain C(410791)	N/A	-
Cgig	XP_011477585_2	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Cpli	DN157725_c0.g1.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Csin	Hic_asm_17.1357	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Csin	Hic_asm_17.1443	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Cvir	XP_022346235_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Dpol	XP_052247387_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Gaeg	XP_041377087_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Hbia	M00000331058	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Hruf	XP_046372770_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mare	XP_052819073_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mcal	XP_052063314_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mchi	DN13809_c0.g1.i1.p1	Fox	Fox-C	Forkhead domain C(410791)	N/A	-
Mcor	CAC537404_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Medu	CAG2206844_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mgal	V1D122482_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mmar	MMAM00000035616	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mmer	XP_045194706_2	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mner	g82158.t1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Mphi	scat69950.1.0	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Obim	XP_014786040_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Oedu	XP_048762038_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Osin	XP_029653806_2	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pcan	XP_025115697_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pcor	DN14158_c0.g2.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pcor	DN14158_c0.g5.i1.p1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pmax	XP_033755061_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pstr	KAK3590993_1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pvir	s02023g12	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Pyes	XP_021346967..1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Rphi	XP_060597004..1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Sbro	EVM0022192..1	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Scon	Chr1..448	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Sglo	Sg009485	Fox	Fox-C	Forkhead domain L1 (410801)	N/A	-
Tgra	KAJ8303551..1	Fox	Fox-C	Forkhead domain C2 (410819)	N/A	-
Acal	XP_035824261..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Airc	Contig1003..15	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Amar	Amar1686	Fox	Fox-D	Forkhead domain D3 (410821)	N/A	-
Apec	DN87882..c0..g1..i1..p1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Apur	scatfold..13962..1..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Bbla	XP_013096936..2	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Cang	XP_052688370..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Cari	EVM0005770..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Cgig	XP_011446328..2	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Cpli	DN23774..c0..g1..i1..p1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Csin	Hic..asm..11..1..425	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Cvir	XP_022231646..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Dpol	XP_052256035..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Dpol	XP_052256591..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Gaeg	XP_041356731..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Hbia	M00000030583	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Hruf	XP_046329290..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Lorb	DN224803..c0..g1..i1..p1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mare	XP_052777467..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mare	XP_052777725..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mcal	XP_052095202..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mcal	XP_052075808..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mcor	CAC5382691..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mcor	CAC5407497..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Medu	CAG2204666..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Medu	CAG2248150..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Medu	CAG2203862..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mgal	VDI107735..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mgal	VDIH3066..1	Fox	Fox-D	Forkhead domain D3 (410821)	N/A	-
Mmar	MMAM00000002467	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mmer	XP_045157253..2	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mner	g192986..t1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mphi	scat..69..89..1..2	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Mphi	scat..42..85..6..0..4	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Obim	XP_052826256..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Oedu	XP_048762457..1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pcan	XP_025110523.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pcor	DN15187_c0_g1.i1.p1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pmar	DN31265_c0_g1.i1.p1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pmax	XP_033737945.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pstr	KAK3592139.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pvir	s01402691	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pvir	s000975340	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Pyes	XP_021345225.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Rphi	XP_0660585828.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Sbro	EVM0002351.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Scon	Chr8:2069	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Sglo	Sg1013024	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Tgra	KAJ8306624.1	Fox	Fox-D	Forkhead domain D4 (410822)	N/A	-
Airc	Contig989.19	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Amar	Amal2850	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Apur	scatfold_253.23	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Cang	XP_052688878.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Cari	EVM0003339.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Chu	DN108936_c2_g1.i1.p1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Cgig	XP_011444776.2	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Cvir	XP_022319236.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Dpol	XP_052286560.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Hbia	M00000038943	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Hruf	XP_046355575.2	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mare	XP_052775423.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mcal	XP_052075782.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mcor	CAC5384360.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Medu	CAG2217852.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Medu	CAG2194171.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Medu	CAG2199036.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mgal	VDH0460.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mmar	MMAM0000033594	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mmer	XP_045157592.2	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mmod	DN117568_c0_g1.i1.p1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mner	g241620.t1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Mphi	scat_31587.0.4	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Oedu	XP_048762291.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Pmar	DN27017_c0_g1.i1.p1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Pmax	XP_033737819.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Pstr	KAK3583103.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Pvir	s00145654	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Pyes	XP_021378858.1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Rdec	DN24595_c4_g1.i1.p1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Rphi	XP_060578687_1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Sbro	EVM0010028_1	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Sglo	Sg009305	Fox	Fox-E	Forkhead domain E (410793)	N/A	-
Acal	XP_005105963_2	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Airc	Contig1133_18	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Amar	Ama39500	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Amar	Amaa2615	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Apec	DN75342_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Apur	scaffold_860_37	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Bbla	XP_055892380_1	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Cang	XP_052712246_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Cari	EVM0011190_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Cgig	XP_011445317_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Cpli	DN7628_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Csin	Hic_asm_17_158	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Cvir	XP_022335664_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Dpol	XP_052232755_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Gaeg	XP_041375666_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Hbia	M00000007664	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Hruf	XP_046372649_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mare	XP_052815084_1	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Mcal	XP_052060477_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mcor	CAC5387332_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Medu	CAG2252875_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mgal	V1D121852_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mmar	MMAM00000030848	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mmer	XP_045194642_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mmod	DN104261_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mner	g106129_t1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Mphi	scat40546_0_2	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Obim	XP_014777539_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Oedu	XP_048732202_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pcan	XP_025116010_1	Fox	Fox-F	Forkhead domain F1 (410823)	N/A	-
Pcor	DN180603_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pcor	DN129940_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pmar	DN14344_c0_g1_i1_p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pmax	XP_033755005_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pstr	KAK3601654_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pvir	s133835g10	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Pyes	XP_021358008_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Rphi	XP_066061663_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Sbro	EVM0015186_1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Scon	Chr11_927	Fox	Fox-F	Forkhead domain F (410794)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Scon	Chr11.810	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Sglo	SgI005267	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Tgra	KAJ8302829.1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Tsqu	DN137576.c0.g1.i1.p1	Fox	Fox-F	Forkhead domain F (410794)	N/A	-
Acal	XP_005099252..2	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Acal	XP_005099253..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Airc	Contig625..38	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Amar	Arna.0381	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Apec	DN10836.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Apur	scaffold..36470..28	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Bbla	XP_055879295..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Cang	XP_052699015..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Cari	EVM0001891..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Chu	DN104980.c0.g1.i2.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Cgig	XP_011427689..2	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Cpli	DN58419.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Csin	Hic.asm..10..1034	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Cvir	XP_022233454..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Dpol	XP_052270224..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Dpol	XP_052270147..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Gaeg	XP_041354930..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Gaeg	XP_041354700..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Hbia	M00000035850	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Hruf	XP_046371537..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Hruf	XP_048259351..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mcal	XP_052104484..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mcor	CAC5405696..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Medu	CAG2193433..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mgal	V1D124297..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mmar	MMAM00000030730	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mmer	XP_045162348..2	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mmod	DN60588.c0.g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mner	g133265..t1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Mphi	scat..15017..0..3	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Obim	XP_052824454..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Oedu	XP_048737541..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pcan	XP_025105677..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pcan	XP_025106039..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pcan	XP_025105724..1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pcor	DN81635.c0.g2..i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pcor	DN81635.c0.g1..i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pgen	DN112984.c0..g1..i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pmar	DN28516.c0.g1..i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pmax	XP_033734631_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Poku	DN41090_c0_g2.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pstr	KAK3604690_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pvir	s00383635	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Pyes	XP_021365790_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Rphi	XP_060589805_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Sbro	EVM0011335_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Sbro	EVM0012606_1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Scon	Chr3_2805	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Sglo	Sg014601	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Sgra	DN49488_c0_g1.i1.p1	Fox	Fox-G	Forkhead domain G (410795)	N/A	-
Airc	Contig_18244_2	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Airc	Contig_178_106	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Amar	Ama1_6564	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Amar	Ama05868	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Cang	XP_052684756_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Cari	EVM0021377_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Chu	DN138466_c0_g1.i1.p1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Cgig	XP_034313225_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Csin	Hic_asm_14_811	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Cvir	XP_022314590_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Hbia	M00000018729	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Hruf	XP_048254913_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Hruf	XP_048255113_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mcor	CAC5408624_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mcor	CAC5397897_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mcor	CAC5397906_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mcor	CAC5403969_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Medu	CAG2228903_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Medu	CAG2188004_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Medu	CAG2252853_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Medu	CAG2202596_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mgal	V1D162725_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mgal	V1DH3947_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mgal	V1D120844_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mmar	MMAM00000022684	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mmer	XP_053378216_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mmer	XP_045194303_2	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mmer	XP_045198985_2	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mner	g213542_t1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mphi	scat_17325_0_4	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Mphi	scat_28666_1_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Oedu	XP_048759429_2	Fox	Fox-H	Forkhead domain H (410796)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pcan	XP_025075954_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Pcor	DN116957.c0.g1.i1.p1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Pmax	XP_033755807_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Pstr	KAK3603859.1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Pvir	s7596633	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Pvir	s00234g131	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Rphi	XP_060558970_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Rphi	XP_060567331_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Rphi	XP_060604067_1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Sbro	EVM0016618.1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Sbro	EVM0013817.1	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Scon	Chr11.1359	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Scon	Chr2.2082	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Sglo	Sg013003	Fox	Fox-H	Forkhead domain H (410796)	N/A	-
Acal	XP_005108651_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Airc	Contig775.5	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Amar	Amad2822	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Apec	DN20109.c0.g1.i6.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Apur	scaffold.797.10	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Bbla	XP_013064514.2	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cang	XP_052684275_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cari	EVM0003558.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Chu	DN127407.c0.g2.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cgig	XP_011444234.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cpli	DN65792.c0.g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Csin	Hic.asm.0.1540	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cvir	XP_0222319181_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Cvir	XP_022231268_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Dpol	XP_052265485_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Gaeg	XP_041362703_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Hbia	M00000003225	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Hruf	XP_046330175_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Lorb	DN146717.c0.g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052764656_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052816854_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052764667_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052775202_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052775217_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mare	XP_052775230_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Meal	XP_052065038_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mchi	DN41583.c0.g1.i5.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mcor	CAC5403074_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Medu	CAG2242807_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mgal	VDI12691.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mmar	MMAM00000019873	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mmer	XP_045212565_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mmod	DN23659_c0_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mner	g198765.t1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Mphi	scat_33310.2.9	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Obim	XP_052824622_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Oedu	XP_04876213_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Osin	XP_029638410_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pcan	XP_025095423_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pcor	DN84891_c0_g1.i2.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pcor	DN480_c0_g2.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pcor	DN84891_c0_g2.i2.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pgen	DN2399_c8_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pmar	DN32837_c1_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pmax	XP_033752519_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Poku	DN19777_c2_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Ppur	DN2521_c0_g1.i4.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pstr	KAK357929.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pvir	s01693g10	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Pyes	XP_0213351058_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Rdec	DN22834_c9_g7.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Rphi	XP_060587750_1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Sbro	EVM0018668.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Scon	Chr1.3201	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Scon	Chr1.3198	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Sglo	Sgl000050	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Sgra	DN10939_c0_g1.i1.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Tgra	KAJ8318321.1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Tsqu	DN6625_c2_g1.i2.p1	Fox	Fox-J1	Forkhead domain J1 (410797)	N/A	-
Acal	XP_005111247_3	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Airc	Contig53_201	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Amar	Ama34942	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Apur	scatfield_11801_24	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Bbla	XP_013070049_1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Cang	XP_052716138_1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Cari	EVM0008910.1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Cflu	DN58808_c0_g1.i1.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Cgig	XP_011422959_2	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Cpli	DN78731_c5_g1.i2.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Csin	Hic_asm_4_381	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Cvir	XP_0222341777_1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Dpol	XP_052285372_1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Gaeg	XP_041378546..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Hbia	M0000016754	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Hruf	XP_048247606..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mare	XP_052759824..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Mare	XP_052759802..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Mcal	XP_052082445..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mchi	DN34970.c0.g1.i2.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mcor	CAC5378041..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Medu	CAG2221519..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mgal	VD157447..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mgal	VD157448..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mmar	MMAM0000017129	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mmer	XP_053378821..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mmer	XP_053379966..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mmod	DN38610.c0.g1.i1.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Mphi	scat_40576.0..4	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Obim	XP_052832979..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Oedu	XP_048739234..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Osin	XP_029651657..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Pcan	XP_025081218..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Pcor	DN2942.c0.g1.i5.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Pcor	DN2942.c0.g4.i3.p1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pcor	DN2942.c0.g4.i3.p2	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pgen	DN5381.c0.g1.i7.p1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pmar	DN41364.c0.g1.i3.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Pmax	XP_033763328..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Poku	DN14771.c0.g2.i3.p1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Ppur	DN2181.c0.g1.i5.p1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pstr	KAK3583417..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Pvir	s00019g6	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pyes	XP_021374633..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Rphi	XP_060577634..1	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Sbro	EVM0013081..1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Scon	Chr12.800	Fox	Fox-J2/3	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Sgio	Sgi003279	Fox	Fox-K	Forkhead domain J3 (410826)	N/A	-
Sgra	DN10442.c0.g1.i2.p1	Fox	Fox-J2/3	Forkhead domain J3 (410826)	N/A	-
Tgra	KAJ8320220..1	Fox	Fox-K	Forkhead domain J3 (410826)	N/A	-
Tsqu	DN10376.c0.g1.i2.p1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Acal	XP_005092494..1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Apur	scatfold_855..43	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Bbla	XP_013090285..1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Cang	XP_052688140..1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Cari	EVM0009067..1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Cgig	XP_011416099_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Cpli	DN64350.c0.g1.i1.p1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Cvir	XP_022316096_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Gaeg	XP_041362451_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Hbia	M0000008333	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Hruf	XP_048248693_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Mmar	MMAM00000012630	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Mphi	scat.14580.0.11	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828; partial)	-
Obim	XP_014772374_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Oedu	XP_048761057_2	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Osin	XP_029646877_1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Rdec	DN21696.c3.g1.i1.p1	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828; partial)	-
Sglo	Sg000589	Fox	Fox-K	Forkhead domain K (410800)	Forkhead associated (FHA) domain (410828)	-
Acal	XP_012940028_1	Fox	Fox-L1	Forkhead domain L1 (410800)	N/A	-
Airc	Contig58.64	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Amar	Amal.791.4	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Apec	DN74037.c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Apur	scaffold.122.2	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Bgla	XP_055890278_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Cang	XP_052718686_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Cari	EVM0019009.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Cgig	XP_011417586_2	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Cpli	DN157469.c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Csin	Hic.asm.17.1.225	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Cvir	XP_022346240_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Dpol	XP_052252043_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Gaeg	XP_041375667_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Hbia	M00000331057	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Hruf	XP_046344397_2	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Lorb	DN104458.c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mare	XP_052817971_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mare	XP_052817977_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mcal	XP_052063315_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mcor	CAC5374005_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Medu	CAG2206845_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mgal	V1D22484_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mgal	V1DH7507_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mmar	MMAM0000028776	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mmer	XP_053402988_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mmod	DN51324.c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mner	g268924.t1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Mphi	scat.69950.0.0	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Obim	XP_0147785001_1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Oedu	XP_048762056.2	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pcan	XP_025076243.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pcor	DN28326_c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pcor	DN187497_c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pmar	DN30135_c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pmax	XP_033755354.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Ppur	DN73831_c0.g1.i1.p1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pstr	KAK3590991.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pvir	s02023g11	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Pyes	XP_021346965.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Rphi	XP_0660608039.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Sbro	EVM0016190.1	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Scon	Chr11_868	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Sglo	Sg009486	Fox	Fox-L1	Forkhead domain L1 (410801)	N/A	-
Acal	XP_005101910.2	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Airc	Contig51.34	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Amar	Am34673	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Apur	scaffold.84.159	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Bbla	NP_055865110.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cang	XP_052718506.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cari	EVM0021728.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Chu	DN127322_c6_g2_i2.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cgig	NP_001295827.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cpli	DN75086_c5_g1_i2_p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Csin	Hic.asm.4.274	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cvir	XP_022345405.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Cvir	XP_022345173.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Dpol	XP_052212727.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Gaeg	XP_04137252.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Hbia	M00000035173	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Hruf	XP_048250285.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Lorb	DN129129_c0_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mare	XP_052760962.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mcal	XP_052082415.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mcor	CAC5401149.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Medu	CAG239672.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mgal	V1D149865.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mgal	V1D149864.1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mmar	MMAM0000016212	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mmer	XP_045161614.2	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mmod	DN2410_c0.g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mner	g83235.t1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Mphi	scat.50301.0.3	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Obim	XP_014785648_2	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Oedu	XP_048729555_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pcan	XP_025083514_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pcor	DN35937_c0_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pgen	DN134171_c0_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pmar	DN32846_c0_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pmax	XP_035724493_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pstr	KAK3602726_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pvir	s002468193	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Pyes	XP_021353421_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Rdec	DN21003_c0_g1.i2.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Rphi	XP_060585301_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Sbro	EVM0017513_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Sbro	EVM0014371_1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Scon	Chr12:1684	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Sglo	Sg1005363	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Tsqu	DN37_c29_g1.i1.p1	Fox	Fox-L2	Forkhead domain L2 (410802)	N/A	-
Acal	XP_005091040_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Airc	Contig281:47	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Amar	Ama2:3426	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Apec	DN27027_c0_g1.i2.p1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Apur	scaf0ld.17.163	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Bbla	XP_055899100_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Cang	XP_052711314_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Cari	EVM0024311_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Cgig	XP_034303195_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Cpli	DN78931_c1_g1.i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Csin	Hic.asm.2:101	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Cvir	XP_022292787_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Dpol	XP_052270473_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Gaeg	XP_041365083_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mcal	XP_052066630_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mchi	M0000027642	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Hruf	XP_048241610_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mare	XP_052801997_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mgal	XP_052066630_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mgal	V1DH3462_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mcor	CAC5383890_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Medu	CAG2257106_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mgal	V1DH93464_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mgal	V1DH3462_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mgal	V1DH93463_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mmar	MMAM0000018109	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-
Mmer	XP_045177340_1	Fox	Fox-N1/4	Forkhead domain N1 (410802)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mmod	DN2607.c0.g2.i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Mphi	scaf_69935.0.10	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Obim	XP_052825413..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Oedu	XP_055996035..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Osin	XP_029638459..2	Fox	Fox-N1/4	Forkhead domain N4 (410831)	N/A	-
Pcan	XP_025087495..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pcor	DN55558..c0.g2.i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pcor	DN55558..c0.g1.i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pgen	DN145626.c0.g..1..p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pmar	DN30748.c0.g1..i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pmax	XP_033751425..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Poku	DN12531..c0.g1..i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Ppur	DN19563..c0.g..1..p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pstr	KAK3587366..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pvir	s24333845	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Pyes	XP_021371548..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Rdec	DN20122..c0.g1..i1.p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Rphi	XP_066060682..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Sbro	EVM0009578..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Scon	Chr14.2061	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Sglo	SgI044456	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Sgra	DN11935..c0.g2..i3..p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Tgra	KAJ8298705..1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Tsqu	DN22139..c0.g1..i1..p1	Fox	Fox-N1/4	Forkhead domain N1 (410804)	N/A	-
Acal	XP_005099217..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Airc	Contig117..153	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Amar	Ama0979	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Apec	DN11918..c0.g1..i1..p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Apur	scaf0ld_489..22	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Bgla	XP_013084252..2	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Cang	XP_052698143..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Cari	EVM0016469..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Chu	DN125734..c1..g..1..p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Cgig	XP_034324255..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Cpli	DN79231..c0.g1..i1..p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Csin	Hic..asm..10..136	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Cvir	XP_0222331167..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Dpol	XP_052270965..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Gaeg	XP_041353111..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Hbia	M0000030949	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Hruf	XP_046351344..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mare	XP_052767715..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mcal	XP_052107190..1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mchi	DN39446_c1.g1.i1.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mcor	CAC5378437.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Medu	CAG2235611.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mgal	VDI180286.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mgal	VDI180289.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mgal	VDI180287.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mmar	MMAM00000016688	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mmer	XP_053370884.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mmod	DN2418_c0.g1.i31.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mner	g156153_i2	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Mphi	scaf_37509_0.5	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Obim	XP_052822674.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Oedu	XP_048735022.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Osin	XP_029633348.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pcan	XP_025106088.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pcor	DN195240_c0.g1.i1.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pcor	DN5191_c0.g2.i7.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pcor	DN18451_c0.g2.i3.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pgen	DN14328_c0.g1.i3.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pmar	DN42157_c1.g1.i3.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pmax	XP_033734749.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Poku	DN17429_c4.g1.i2.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Ppur	DN5075_c0.g1.i1.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pstr	KAK3595953.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pvir	s00410g95	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Pyes	XP_021366964.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Rdec	DN22296_c2.g1.i1.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Rphi	XP_060552999.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Sgio	Sgi013452	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Sgra	DN13133_c0.g2.i8.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Tgra	KAJ18308641.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Tgra	KAJ18316727.1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Tsqu	DN75347_c0.g1.i2.p1	Fox	Fox-N2/3	Forkhead domain N3 (410833)	N/A	-
Acal	XP_005112460.1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Airc	Contig116.24	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Amar	Ama178.4	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Apec	DN23636_c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Apur	scaffold_712.6	Fox	Fox-O	Forkhead domain O1 (410834)	N/A	-
Bbla	XP_013095255.2	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cang	XP_052708272.1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cari	EVM00000968.1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cflu	DN112955_c5.g3.i5.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cgig	XP_011414359.1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Cpli	DN72415.e5.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cvir	XP_022287692_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Cvir	XP_022287423_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Dpol	XP_057243903_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Gaeg	XP_041351714_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Hbia	M00000024236	Fox	Fox-O	Forkhead domain O1 (410834)	N/A	-
Hruf	XP_046374293_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Lorb	DN142512.c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mare	XP_052808484_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mcal	XP_052087664_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mchi	DN32908.c0.g1.i3.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mcor	CAC5392548_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mmar	MMAM00000004017	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mmer	XP_045195791_2	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mmod	DN145.c2.g1.i21.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Mner	g187283.t1	Fox	Fox-O	Forkhead domain O3 (410835)	N/A	-
Obim	XP_052832486_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Oedu	XP_048766900_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Osin	XP_029657040_1	Fox	Fox-O	Forkhead domain O1 (410806)	N/A	-
Pcan	XP_025097678_1	Fox	Fox-O	Forkhead domain O3 (410835)	N/A	-
Pcor	DN20008.c0.g1.i2.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pcor	DN7549.c0.g2.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pcor	DN7549.c0.g3.i2.p1	Fox	Fox-O	Forkhead domain O1 (410834)	N/A	-
Pgen	DN24871.c0.g1.i3.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pgen	DN24871.c0.g1.i3.p2	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pmar	DN44399.c0.g5.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pmax	XP_033740844_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Poku	DN10962.c1.g2.i2.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Ppur	DN72510.c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pstr	KAK357695.1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pvir	s00079g102	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Pyes	XP_021377366_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Rdec	DN30721.c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Rphi	XP_060589384_1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Scon	Chr2.491	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Sglo	Sglo18351	Fox	Fox-O	Forkhead domain O1 (410834)	N/A	-
Sgra	DN5576.c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Tsqu	DN138852.c0.g1.i1.p1	Fox	Fox-O	Forkhead domain O (410806)	N/A	-
Airc	Contig3461.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Airc	Contig330.72	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Apur	scatfold_576.108	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Cang	XP_052673828_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Cari	EVM0015778.1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Cgig	XP_011412452_2	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Cvir	XP_022300144_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Dpol	XP_052234997_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Dpol	XP_052237166_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Gaeg	XP_041362068_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Hbia	M00000010651	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Hruf	XP_048236781_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Hruf	XP_046328651_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mcal	XP_052088402_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mcor	CAC5388114_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Medu	CAG2250347_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mgal	VD105563_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mgal	VD105564_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mmar	MMAM00000027087	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mmer	XP_045182963_2	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mmod	DN9753_c0_g1_i1_p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Mphi	scat_66119_0_21	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Oedu	XP_048731527_2	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Pmar	DN32892_c0_g1_i1_p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Pmax	XP_033727511_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Pstr	KAK3609024_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Pvir	s01298551	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Pyes	XP_021354438_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Rdec	DN23702_c2_g1_i2_p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Rphi	XP_060566633_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Sbro	EVM0009544_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Scon	Chr1.409	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Sglo	Sglo04484	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Tgra	KAJ8322379_1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Tsqu	DN207442_c0_g1_i1_p1	Fox	Fox-OG13/NA	Forkhead domain P (410807)	N/A	-
Acal	XP_005106916_3	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	Forkhead domain Q2 (410809)	Annotated as Fox-Q2b
Airc	Contig5_21	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Amar	Amar19770	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Apur	scaf0ld_360_14	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Bbla	XP_013071662_2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Cang	XP_05267257_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Cari	EVM0001823_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Cgig	XP_011438389_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Cpli	DN35479_c0_g1_i1_p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Csin	Hic_asm_12_159	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Cvir	XP_022296913_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Dpol	XP_052253230_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Dpol	XP_052253240_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Gaeg	XP_041347345_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Hbia	M00000015843	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Hruf	XP_046382017_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mare	XP_05277846_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mcal	XP_052062481_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mcor	CAC5419515_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Medu	CAG2199155_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mgal	VID150805_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mmar	MMAM00000012410	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mmer	XP_045166371_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Mphi	scat.67833.0_2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Obim	XP_014773941_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Oedu	XP_048742700_2	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pcan	XP_025096321_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pcor	DNT7667.c0.g1.i1.p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pcor	DNT34039.c0.g1.i1.p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pcor	DNT7667.c0.g3.i1.p1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pmax	XP_033744896_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pstr	KAK3589497_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Pyes	XP_021371037_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Rphi	XP_060586724_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Sbro	EVM0000506_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Sbro	EVM0006433_1	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Scon	Chr7.1624	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Sglo	Sg024307	Fox	Fox-OG15/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2b
Airc	Contig1425_7	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Apur	scaffold.604.173	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Cang	XP_052678026_1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410788)	N/A	-
Cari	EVM0023364_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Chu	DN107758.c5.g.1.i2.p1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Cgig	XP_019927657_2	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Cpli	DNT0609.c0.g1.i2.p1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Cvir	XP_022321288_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Cvir	XP_022295893_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Dpol	XP_052222623_1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Gaeg	XP_041365712_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Hbia	M00000030826	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Hruf	XP_04636140_2	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Lorb	DN243786.c0.g.1.i1.p1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Mare	XP_052802309_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Mcor	CAC5370465_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Medu	CAG2202185_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-
Medu	CAG2246856_1	Fox	Fox-OG16/NA	Forkhead domain (410788)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mgal	VDI24665.1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Mmar	MMAM00000032793	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Mmer	XP_045177123.2	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Mphi	scat1.70200.0.4	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Obim	XP_014771053.1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Oedu	XP_048728661.2	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Osin	XP_036362897.1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Pmar	DN13205_c4_g1_i12.p1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Pmax	XP_033752233.1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Ppur	DN14666_c0_g1_i2.p1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Pstr	KAK3581527.1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Pvir	s00119451	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Pyes	XP_021353413.1	Fox	Fox-OG16/NA	Forkhead domain Q2 (410809)	N/A	-
Rdec	DN22502_c0_g4_i1.p1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Rphi	XP_060599562.1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Rphi	XP_060601370.1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Sbro	EVM0002125.1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Scon	Chr14.1628	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Tsqu	DN2384_c1_g1_i1.p1	Fox	Fox-OG16/NA	Forkhead domain (41078)	N/A	-
Airc	Contig138.5	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Apec	DN23140_c1_g1_i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Apur	scaf001.333.74	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Cang	XP_052712245.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Cari	EVM0012682.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Chu	DN113069_c3_g1_i4.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Ccli	DN76173_c0_g1_i2.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Csin	Hic.asm.2.1802	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Csin	Hic.asm.2.1455	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Cvir	XP_022286391.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Dpol	XP_052282333.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Dpol	XP_052234537.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Gaeg	XP_041366058.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Hbia	M0000014061	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Hruf	XP_046335487.2	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mare	XP_052798013.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mcal	XP_052064572.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mcor	CAC5375062.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Medu	CAG2224977.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mmar	MMAM0000037791	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mmer	XP_05338136.1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mmod	DN27089_c0_g1_i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Mner	g144243.t1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Oedu	XP_048731286.2	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pcor	DN5679_c0.g2.i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pcor	DN5679_c0.g1.i15.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pcor	DN13056_c0.g1.i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pcor	DN13056_c0.g2.i2.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pgen	DN28413_c1.g1.i2.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pmar	DN44947_c1.g2.i3.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pmax	XP_033751305_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Poku	DN37223_c1.g1.i8.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Ppur	DN3451_c0.g1.i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pstr	KAK3597624_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pvir	s00219g11	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Pyes	XP_021377259_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Rdec	DN22152_c4.g4.i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Rphi	XP_066956990_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Sbro	EVM0023670_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Sglo	Sg1005561	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Sgra	DN54780_c0.g1.i1.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Tgra	KAJ8299135_1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Tsqu	DN6434_c0.g1.i7.p1	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-M
Airc	Contig465_41	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-Q2c
Amar	Ama25953	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-Q2c
Apur	scafold_867_41	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-Q2c
Apur	scafold_381_16	Fox	Fox-OG2/NA	Forkhead domain M (410803)	N/A	Annotated as Fox-Q2c
Cang	XP_052700156_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Cari	EVM0004465_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Cgig	XP_011435457_2	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Csin	Hic_asm_16.939	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Cvir	XP_022333263_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Dpol	XP_052278575_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Dpol	XP_052278576_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Dpol	XP_052278604_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Hruf	XP_046341176_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mare	XP_052791887_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mare	XP_052791890_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mare	XP_052791888_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419385_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5380823_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5379920_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419389_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419386_1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mcor	CAC5419381.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419382.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419383.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419387.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419380.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mcor	CAC5419388.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG2214460.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22194706.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198066.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198058.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198055.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG2214461.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198060.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198057.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198065.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198059.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198056.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG2234548.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198063.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198061.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG22198064.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Medu	CAG2234548.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI102350.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI139859.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI15906.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI102348.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI115903.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI102347.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI115905.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI102349.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mgal	VDI115904.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mmer	XP_053405097.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Mmod	DN6982_c0_g1.i3.p1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
MPhi	scat.15444.0.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Oedu	XP_056021213.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Pmar	DN30963_c0_g1.i1.p1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Pmax	XP_033751006.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Pmax	XP_033749723.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Pvir	s03437648	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Pyes	XP_021366588.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Rphi	XP_060585777.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Sbro	EVM0013029.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Scon	Chr5.396.1	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Sglo	Sglo13625	Fox	Fox-OG28/NA	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2c
Acal	XP_005102249.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Airc	Contig636.41	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Amar	Amar16012	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Apur	scaffold_31352	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Bgla	XP_055874345.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Cang	XP_052699279.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Cari	EVM0018541.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Cgig	XP_011441298.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Cvir	XP_022334408.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Cvir	XP_022334077.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Dpol	XP_052275569.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Gaeg	XP_041375984.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hbia	M00000015535	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Hruf	XP_046350707.2	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350710.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350686.2	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350709.2	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Hruf	XP_046350688.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350660.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350687.2	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350714.2	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350712.2	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Hruf	XP_046350708.2	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Mare	XP_052795236.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Mcal	XP_052102496.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Mcor	CAC5414394.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Medu	CAG2193762.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Mgal	V1D135942.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Mmar	MMAM00000023830	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Mmer	XP_045215157.2	Fox	Fox-OG39/NA	Forkhead domain E (410801)	N/A	Annotated as Fox-AB
Mphi	scat7111.0.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Oedu	XP_048737442.2	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Pcan	XP_025076030.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Pmax	XP_033756900.1	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Pstr	KAK3601439.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Pstr	KAK3601419.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Pvir	s00189g393	Fox	Fox-OG39/NA	Forkhead domain E (410793)	N/A	Annotated as Fox-AB
Pyes	XP_021357612.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Rphi	XP_060589081.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Sbro	EVM0003782.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Scon	Chr5.18	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Sglo	Sglo004401	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Tgra	KA18304916.1	Fox	Fox-OG39/NA	Forkhead domain L1 (410801)	N/A	Annotated as Fox-AB
Airc	Contig120.132	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Apur	scaffold_507.3	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Bgla	XP_052883746.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Cang	XP_052680529.1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Cari	EVM0015348.1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Chu	DN101403_c0.g2.i3.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Cgig	XP_011419330.2	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Cpli	DNS80534_c6.g1.i4.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Csin	Hic.asm..12.158	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Cvir	XP_022295865.1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Dpol	XP_052258768.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Gaeg	XP_041347582.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Hruf	XP_048239143.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Lorb	DN59652_c1.g1.i4.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mare	XP_052780914.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mcal	XP_052101037.1	Fox	Fox-P	Forkhead domain P2 (410839)	N/A	-
Mcal	XP_052083749.1	Fox	Fox-P	Forkhead domain P2 (410839)	N/A	-
Mcal	XP_052100969.1	Fox	Fox-P	Forkhead domain P2 (410839)	N/A	-
Mgal	XP_052061774.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mchi	DN46200_c0.g1.i1.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mcor	CAC5419517.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Medu	CAG2199156.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mgal	V1D14555.1	Fox	Fox-P	Forkhead domain P2 (410839)	N/A	-
Mgal	V1D150808.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mgal	V1D150806.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mgal	V1D150807.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mmar	MMAM0000012411	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mmer	XP_053376718.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mmod	DN2151_c0.g1.i2.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mner	g93547.t1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Mphi	scat_39347_0.7	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Obim	XP_052822821.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Oedu	XP_055998748.1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Osin	XP_036357858.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pcan	XP_028106713.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pcor	DN1820_c0.g1.i26.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pgen	DN35611_c0.g1.i6.p1	Fox	Fox-P	Forkhead domain P2 (410839)	N/A	-
Pmar	DN41268_c0.g1.i1.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pmax	XP_033745371.1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Poku	DN88526_c2.g1.i1.p1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036)	-
Ppur	DN59408_c0.g1.i2.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pstr	KAK3589495.1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Pvir	s01329g124	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Pyes	XP_021363304..1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Rdec	DNS8028.c0.g1.i1.p1	Fox	Fox-P	Forkhead domain P (410807)	FOXP coiled-coil domain (465036; partial)	-
Rphi	XP_0660586741..1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Sbro	EYMM004295..1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Scon	Chr7:2129	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036; partial)	-
Scon	Chr7:2133	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Sglo	Sg011345	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Sgra	DN17101.c0.g1.i13.p1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Tgra	KAJ8302344..1	Fox	Fox-P	Forkhead domain P2 (410839)	FOXP coiled-coil domain (465036)	-
Acal	XP_005099459..2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Airc	Contig1420..28	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Amar	Amz29905	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Apur	scaffold_832..35	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Bgl2	XP_055865367..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cang	XP_052699620..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cari	EYMM002665..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cgig	XP_011425762..2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cpli	DN105612.c0.g1..11.p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Csin	Hic_asm..16..4	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Cvir	XP_0222333968..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Dpol	XP_052289896..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	
Gaeg	XP_041363041..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Hbia	M00000035328	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Hruf	XP_046373579..2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mcal	XP_052101305..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mchi	DN22486.c0.g2..11.p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mcor	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	
Mgal	V1D174621..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mmar	MMAM0000000686	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mmer	XP_045215524..2	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mner	g209553..t1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Mphi	scat:22910..1..1	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	
Pcan	XP_025078472..1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pcor	DN99235.c0.g1..11..p1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	Annotated as Fox-Q2a	
Pmar	DN49466.c0.g1..11..p1	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	

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Rphi	XP_060571531.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2a
Sbro	EVM0023378.1	Fox	Fox-Q2	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2a
Scon	Chr5.2105	Fox	Fox-Q2	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2a
Scon	Sg009183	Fox	Fox-Q2	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2a
Sglo	XP_005109004.3	Fox	N/A	Forkhead domain Q2 (410809)	N/A	Annotated as Fox-Q2a
Acal	Contig879.9	Fox	N/A	Forkhead domain Q2 (410788)	N/A	-
Airc	Amar	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Cang	Ama25952	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Amar	Ama7375	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Cang	XP_052676682.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cang	XP_052680288.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cang	XP_052677368.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cari	EVM0001935.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cari	EVM0027332.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cgig	XP_034306826.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cgig	XP_011447567.2	Fox	N/A	Forkhead domain H (410796)	N/A	-
Cpli	DN157619.c0.g.11.p1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Cvir	XP_0222300767.1	Fox	N/A	Forkhead domain (410788)	N/A	-
Cvir	XP_0222300750.1	Fox	N/A	Forkhead domain H (410796)	N/A	-
Dpol	XP_052277921.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Dpol	XP_052227296.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Gaeg	XP_041366967.1	Fox	N/A	Forkhead domain Q2 (410788)	N/A	-
Gaeg	XP_041378820.1	Fox	N/A	Forkhead domain (410788)	N/A	-
Gaeg	XP_041347225.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Gaeg	XP_041375925.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Gaeg	XP_041375913.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Gaeg	XP_041378820.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Hbia	M0000018946	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mare	XP_052791886.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mare	XP_052771066.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mcal	XP_052098820.1	Fox	N/A	Forkhead domain N1 (410804)	N/A	-
Mgal	VDI152978.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mgal	CAC5413379.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mmar	MMAM00000049704	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Medu	CAG2194707.1	Fox	N/A	Forkhead domain L1 (410801)	N/A	-
Medu	CAG2208945.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mgal	VDI15902.1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mner	g159704.t1	Fox	N/A	Forkhead domain L1 (410801)	N/A	-
Mphi	scat.46189.0.0	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mphi	scat.15444.0.2	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Mphi	scat.27787.1.10	Fox	N/A	Forkhead domain L1 (410801)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Obim	XP_014777604_1	Fox	N/A	Forkhead domain M (410803)	N/A	-
Oedu	XP_048739629_2	Fox	N/A	Forkhead domain H (410796)	N/A	-
Osin	XP_036359188_1	Fox	N/A	Forkhead domain M (410803)	N/A	-
Osin	XP_029655092_1	Fox	N/A	Forkhead domain (41078)	N/A	-
Osin	XP_029655092_1	Fox	N/A	Forkhead domain (41078)	N/A	-
Pcor	DN89866_c4_g1.i1.p1	Fox	N/A	Forkhead domain (41078)	N/A	-
Pcor	DN55206_c2_g1.i1.p2	Fox	N/A	Forkhead domain FOXJ2, FOXJ3 (410798)	N/A	-
Pcor	DN155905_c0_g1.i1.p1	Fox	N/A	Forkhead domain P (410807)	N/A	-
Pmax	XP_033750561_1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Pstr	KAK3585306_1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Pvir	s00585648	Fox	N/A	Forkhead domain L1 (410801)	N/A	-
Pyes	XP_021348419_1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Rdec	DN23525_c0_g1.i1.p1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Rphi	XP_069585776_1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Rphi	XP_069551131_1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Scon	Chr5_397	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Sglo	Sg021575	Fox	N/A	Forkhead domain (41078)	N/A	-
Tsqu	DN23960_c0_g1.i1.p1	Fox	N/A	Forkhead domain Q2 (410809)	N/A	-
Acal	XP_035824685_1	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Acal	XP_012946205_1	Sox	N/A	High mobility group box (438820)	N/A	-
Acal	XP_005105939_1	Sox	N/A	High mobility group box (438820)	N/A	-
Apec	DN48806_c0_g1.i1.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Apec	DN108003_c0_g1.i1.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Apur	scatfold_391_70	Sox	N/A	High mobility group box (438820)	N/A	-
Bbla	XP_013078241_2	Sox	N/A	High mobility group box (438820)	N/A	-
Bbla	XP_013078156_1	Sox	N/A	High mobility group box (438820)	N/A	-
Cang	XP_052697278_1	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Cang	XP_052713692_1	Sox	N/A	High mobility group box (438820)	N/A	-
Cari	EVM0018891_1	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674; partial)	-
Cari	EVM0005567_1	Sox	N/A	High mobility group box (438820)	N/A	-
Cgig	XP_011425869_2	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Cgig	XP_034335819_1	Sox	N/A	High mobility group box (438820)	CW-type Zinc Finger (462181)	-
Cvir	XP_0222336758_1	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Cvir	XP_022233079_1	Sox	N/A	High mobility group box (438820)	N/A	-
Dpol	XP_052271004_1	Sox	N/A	High mobility group box (438820)	N/A	-
Gaeg	XP_041377139_1	Sox	N/A	High mobility group box (438820)	N/A	-
Hbia	M00000038049	Sox	N/A	High mobility group box (438820)	N/A	-
Hbia	M00000004998	Sox	N/A	High mobility group box (438820)	N/A	-
Hruf	XP_046329595_1	Sox	N/A	High mobility group box (438820)	N/A	-
Hruf	XP_048239511_1	Sox	N/A	High mobility group box (438820)	N/A	-
Mcor	CAC5384832_1	Sox	N/A	High mobility group box (438820)	N/A	-
Medu	CAG2253429_1	Sox	N/A	High mobility group box (438820)	N/A	-
Mgal	V1D178477_1	Sox	N/A	High mobility group box (438820)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mmod	DN113112.c0,g1.11.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Obim	XP_014776519..1	Sox	N/A	High mobility group box (438820)	N/A	-
Oedu	XP_048738250..2	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Oedu	XP_048752144..2	Sox	N/A	High mobility group box (438820)	N/A	-
Osin	XP_036355200..1	Sox	N/A	High mobility group box (438820)	N/A	-
Osin	XP_029654541..1	Sox	N/A	High mobility group box (438820)	N/A	-
Osin	XP_029656568..1	Sox	N/A	High mobility group box (438820)	N/A	-
Osin	XP_029657644..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Osin	XP_029656220..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Osin	XP_029657648..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Osin	XP_029655056..1	Sox	N/A	High mobility group box B (438820)	N/A	-
Osin	XP_029655785..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Osin	XP_029656129..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Osin	XP_029654991..1	Sox	N/A	High mobility group box A, B and G (438837)	N/A	-
Pcan	XP_025104729..1	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Pcor	DN32781.c0,g1.i1.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Pcor	DN21964.c0,g1.i2.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Pmar	DN33290.c0,g1.i2.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Pmar	DN35008.c0,g1.i4.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Pmax	XP_033755821..1	Sox	N/A	High mobility group box (438820)	N/A	-
Ppur	DN4784.c0,g1.i4.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Pyes	XP_0211347051..1	Sox	N/A	High mobility group box (438820)	N/A	-
Sbro	EVM0018224..1	Sox	N/A	High mobility group box (438820)	N/A	-
Sglo	Sg1009175	Sox	N/A	High mobility group box (438820)	Helix loop helix domain (197674)	-
Sglo	Sg1012029	Sox	N/A	High mobility group box (438820)	N/A	-
Tsqu	DN639.c0,g1..1.p1	Sox	N/A	High mobility group box (438820)	N/A	-
Acal	XP_005108230..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Acal	XP_035824438..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Airc	Contig49.126	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Airc	Contig44.115	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Amar	Ama33032	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Amar	Am333828	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Apec	DN29410.c0,g1.i2.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Apec	DN12297.c0,g1.i3.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Apur	scaffold.15489..10	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Apur	scaffold.865..4	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Bbla	XP_013075432..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Bbla	XP_055868106..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cang	XP_052706368..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cang	XP_052705551..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cari	EVM0026792..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cari	EVM0013965..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cflu	DN118670.c2,g1..12.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Chu	DN99542.c1.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cgig	XP_011433975..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cgig	XP_011455662..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cpli	DN31343.c0.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cpli	DN98511.c0.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Csin	Hic.asm.6.930	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Csin	Hic.asm.6.233..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cvir	XP_02228516..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Cvir	XP_022345230..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Dpol	XP_052214544..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Dpol	XP_052217420..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Gaeg	XP_041353075..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Gaeg	XP_041357874..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Hbia	M00000002798	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Hbia	M00000033682	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Hruf	XP_046370193..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Hruf	XP_046326733..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Lorb	DN80278.c0.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Lorb	DN14..c4.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mare	XP_052784929..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mare	XP_052784720..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mcal	XP_052105617..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mcal	XP_052104911..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mchi	DN33632.c0.g1.i2.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mchi	DN45716.c0.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mcor	CAC540.01077..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mcor	CAC5413203..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Medu	CAG2229644..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Medu	CAG2206403..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mgal	V1D133296..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mgal	V1D169660..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmar	MMAM00000041532	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmar	MMAM00000023253	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmer	XP_045201594..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmer	XP_045201080..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmod	DN1981..c0.g1.i2.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mmod	DN78279..c0.g1.i1.p1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mner	g140596..t1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mner	g157489..t1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mphi	scat.66349.0..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Mphi	scat.24206.0..4	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Obim	XP_014789971..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-
Obim	XP_014780771..1	Sox	Sox-B1/2	High mobility group box B (438790)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Oedu	XP_048746651.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Oedu	XP_048746633.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Osin	XP_029654000.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Osin	XP_029659838.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pcan	XP_02507293.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pcan	XP_02507598.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pcor	DN9087.c0.g1.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pcor	DN14753.c0.g1.i2.p1	Sox	SoxB1/2	High mobility group box A, B and G (438837)	N/A	-
Pcor	DN5688.c0.g1.i0.p2	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pcor	DN5688.c0.g3.i1.p1	Sox	SoxB1/2	High mobility group box A, B and G (438837)	N/A	-
Pcor	DN9742.c0.g1.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pgen	DN5232.c0.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pgen	DN1199.c0.g1.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pmar	DN30477.c0.g2.i1.p1	Sox	SoxB1/2	High mobility group box A, B and G (438837)	N/A	-
Pmar	DN30459.c0.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pmax	XP_033760667.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pmax	XP_033759382.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Poku	DN51870.c1.g1.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Poku	DN10677.c0.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Ppur	DN202737.c0.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pstr	KAK358631.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pstr	KAK3589936.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pvir	s00319g159	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pvir	s00037g281	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pyes	XP_021356125.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pyes	XP_021344413.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Pyes	XP_021372128.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Rdec	DN21477.c2.g7.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Rphi	XP_060556101.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Rphi	XP_060561544.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sbro	EVM0016386.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sbro	EVM0007529.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Scon	Chr9.1352	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Scon	Chr9.1522	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Scon	Chr9.1514	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sglo	Sglo10100	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sglo	Sglo20107	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sgra	DN3782.c0.g1.i3.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Sgra	DN357.cl.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Tgra	KA18310140.1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Tsqu	DN97880.c0.g1.i1.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Tsqu	DN50556.c0.g1.i2.p1	Sox	SoxB1/2	High mobility group box B (438790)	N/A	-
Airc	Contig80.70	Sox	SoxC	High mobility group box C (438838)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Amar	Amal12726	Sox	SoxC	High mobility group box C (438838)	N/A	-
Apec	DN12286_c0_g3.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Apur	scaffold_16.61	Sox	SoxC	High mobility group box C (438838)	N/A	-
Cang	XP_052689209.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Cari	EVM0025846.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Chu	DN126276_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Cgig	XP_011445203.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Cpli	DN19112_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Csin	Hic.asm_11.1.009	Sox	SoxC	High mobility group box C (438838)	N/A	-
Cvir	XP_0222317619.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Dpol	XP_052257395.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Gaeg	XP_041353324.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Hbia	M00000037669	Sox	SoxC	High mobility group box C (438838)	N/A	-
Hruf	XP_046365064.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Lorb	DN14941_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mare	XP_052777703.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mcal	XP_052087802.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mchi	DN44798_c0_g4.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mcor	CAC5424030.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Medu	CAG2189937.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mgal	VD141453.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mgal	VD114462.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mmar	MMAM00000036315	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mmer	XP_045158937.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mmod	DN104308_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mner	g26404.t1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Mphi	scat.17954.1.5	Sox	SoxC	High mobility group box C (438838)	N/A	-
Oedu	XP_048762549.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Osin	XP_028654195.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pcan	XP_025110204.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pcor	DN2429_c2_g1.i2.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pcor	DN29124_c0_g2.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pmax	XP_033737425.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Poku	DN71015_c0_g2.i17.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Ppur	DN89859_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pstr	KAK3610995.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pvir	s00145g243	Sox	SoxC	High mobility group box C (438838)	N/A	-
Pyes	XP_021356242.1	Sox	SoxC	High mobility group box C (438838)	N/A	-
Rdec	DN52924_c0_g1.i1.p1	Sox	SoxC	High mobility group box C (438838)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Rphi	XP_060555827_1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Sbro	EVM0006311.1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Scon	Chr8.1790	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Sglo	Sg1000072	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Sgra	DN6210.c0.g1.i1.p1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Tgra	KAJ8306264.1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Tgra	KAJ8306266.1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Tsqu	DN11669.c1.g1.i2.p1	Sox	Sox-C	High mobility group box C (438838)	N/A	-
Acal	XP_035824396.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Airc	Contig290.5.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Amar	Ama23921	Sox	Sox-D	High mobility group box (438839)	N/A	-
Apec	DN1990.c0.g1.i0.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Apur	scaffold_393.10	Sox	Sox-D	High mobility group box (438839)	N/A	-
Bbla	XP_055899647.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Cari	EVM0012405.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Chu	DN124552.c0.g1.i15.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Cgig	XP_011425377.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Cpli	DN64448.c0.g1.i1.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Csin	Hic.asm.2.1656	Sox	Sox-D	High mobility group box (438839)	N/A	-
Csin	Hic.asm.2.1600.2	Sox	Sox-D	High mobility group box (438839)	N/A	-
Cvir	XP_0222302926.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Dpol	XP_052213125.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Gaeg	XP_041367101.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Hbia	M00000014008	Sox	Sox-D	High mobility group box (438839)	N/A	-
Hruf	XP_046329046.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Lorb	DN5537.c0.g2.i3.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mare	XP_052800695.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mcal	XP_052063962.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mchi	DN38691.c1.g1.i4.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mcor	CAC5366270.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Medu	CAG2197887.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mgal	V1D147525.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mgal	V1D147529.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mgal	V1D147528.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mgal	V1D147527.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mmod	V1D147526.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mner	V1D147530.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mmar	MMAM00000004319	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mmer	XP_053384959.1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mmod	DN588.c0.g1.i9.p1	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mner	g103147.i2	Sox	Sox-D	High mobility group box (438839)	N/A	-
Mphi	scat.4218.i.1.3	Sox	Sox-D	High mobility group box (438839)	N/A	-
Obim	XP_052828391.1	Sox	Sox-D	High mobility group box (438839)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Oedu	XP_048779633..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Osin	XP_029644081..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pcan	XP_025088657..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pcor	DN353.c2.g2.i1.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pcor	DN1386.c1.g1.i1.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pcor	DN13571.c0.g1.i4.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pcor	DN1386.c1.g2.i2.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pgen	DN24654.c0.g1.i2.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pmar	DN40112.c0.g1.i4.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pmax	XP_033751614..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Poku	DN371.c1.g4.i1.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Ppur	DN33319.c0.g1.i1.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Pstr	KAK3605548..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Rphi	XP_0660604110..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Sbro	EVM0000795..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Scon	Chr14.562..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Sgra	DN6138.c0.g1.i6.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Tgra	KAJ8298781..1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Tsqu	DN55031.c0.g1.i1.p1	Sox	Sox-ID	High mobility group box (438839)	N/A	-
Acal	XP_005102100..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Airc	Contig52.209	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Amar	Amad1107	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Apec	DN4330.c0.g1.i1.p1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Apur	scaffold.488.7	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Bgla	XP_013091187..2	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Cang	XP_052689355..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Cari	EVM0005846..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Cflu	DN106407.c5.g2.i1.p1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Cgig	NP_001299801..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Cpli	DN71393.c0.g2.i1.p1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Cpli	DN71393.c0.g1.i2.p1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Csin	Hic.asm.0..353	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Cvir	XP_022312895..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Dpol	XP_052264587..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Gaeg	XP_041362638..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Hbia	M00000012324	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Hbia	M00000012325	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Hruf	XP_04635366..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mare	XP_052786944..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Mare	XP_052783666..1	Sox	Sox-E	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Mcal	XP_052068536.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Michi	DN42011.c0.g1.i12.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mcor	CAC5402442.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Medu	CAG2231021.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mgal	VDI182092.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mgal	VDI182090.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mmar	MMAM00000042410	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Mmer	XP_045213795.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mmod	DN78330.c0.g1.i11.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Mphi	scat.25414.0.6	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Oedu	XP_056019113.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Pcan	XP_025091262.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Pcor	DN4274.c0.g1.i3.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Pcor	DN96098.c0.g1.i1.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Pcor	DN4274.c0.g3.i1.p1	Sox	SoxE	High mobility group box E (438840)	N/A	-
Pmar	DN30335.c0.g1.i1.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Pmax	XP_033739301.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Poku	DN87807.c0.g1.i7.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Ppur	DN46000.c0.g1.i1.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Pstr	KAK3600863.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Pstr	KAK3610785.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Pvir	s136484g74	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Pyes	XP_021348843.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Rphi	XP_0660604697.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Sbro	EVM0002110.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Scon	Chr1.75	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Sglo	Sglo024297	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Sgra	DN22463.c0.g1.i1.p1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Tgra	KAJ8317914.1	Sox	SoxE	High mobility group box E (438840)	Sox developmental protein N terminal (463586; partial)	-
Tsqu	DN8973.c2.g1.i2.p1	Sox	SoxF	High mobility group box E (438840)	Sox developmental protein N terminal (463586)	-
Acal	XP_005107482.1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Airc	Contig80.101	Sox	SoxF	High mobility group box F (438841)	N/A	-
Amar	Ama1616	Sox	SoxF	High mobility group box F (438841)	N/A	-
Apur	scatfield.546.32	Sox	SoxF	High mobility group box F (438841)	N/A	-
Bbla	XP_013074628.2	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cang	XP_052685434.1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cari	EVM000623.1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cflu	DN139006.c0.g1.i1.p1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cgig	XP_011448074.2	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cpli	DN4414.c0.g1.i1.p1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Csin	Hic.asm.11.549	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cvir	XP_0222319962.1	Sox	SoxF	High mobility group box F (438841)	N/A	-
Cvir	XP_0222314364.1	Sox	SoxF	High mobility group box F (438841)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Dpol	XP_052274104.1	Sox	Sox-F	High mobility group box (438820)	N/A	-
Gaeg	XP_041359436.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Hbia	M00000015459	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Hruf	XP_042357912.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mare	XP_042774544.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mare	XP_052774361.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mcal	XP_052061059.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mcor	CAC5414609.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Medu	CAG2242031.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Medu	CAG2187650.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mgal	VDI50271.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mgal	VDI50270.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mmar	MMAM00000025810	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mmer	XP_053395054.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mmod	DN80495.c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mner	g15494.t1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Mphi	scat_61114.0.13	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Obim	XP_052825684.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Oedu	XP_048764319.2	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pcan	XP_028109598.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pcor	DN11375.c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pcor	DN29649.c0.g1.i3.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pcor	DN5688.c2.g1.i3.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pgen	DN144332.c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pmar	DN24748.c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pmax	XP_03373287.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Poku	DN41229.c1.g2.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pstr	KAK3583243.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pvir	s001376284	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pyes	XP_021378109.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Rdec	DN53443.c0.g1.i1.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Rphi	XP_060559438.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Sbro	EVM00000861.1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Pyes	Chr8.897	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Sglo	Sglo005442	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Sgra	DN5013.c3.g1.i7.p1	Sox	Sox-F	High mobility group box F (438841)	N/A	-
Airc	Contig1525.38	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Amar	Ama26724	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Apec	DN93182.c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Apur	scatfold_768.3	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Cang	XP_052703370.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Cari	EVM0018164.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Cgig	XP_011415859.3	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-

Table S3.6 continued from previous page

Sp ID	Gene ID	Group	Annotation	Main catalytic domain (Pssm-ID)	Additional domains (Pssm-ID)	Notes
Cpli	DN80002.c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Csin	Hic.asm.15.1.471	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Cvir	XP_022333738.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Dpol	XP_057226448.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Gaeg	XP_041370217.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Gaeg	XP_041369137.1	Sox	Sox-H	High mobility group box (438820)	N/A	-
Hbia	M00000001.184	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Hruf	XP_046358520.2	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mcal	XP_05209860.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mcor	CAC5406014.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Medu	CAG2257203.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mgal	VD130824.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mgal	VD130823.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mmar	MMAM00000015662	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mmer	XP_053407277.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mner	g125234.t1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mphi	scat.12010.0.4	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Mphi	scat.59202.0.9	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Obim	XP_052832677.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Oedu	XP_056006679.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Osin	XP_036368794.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pcor	DN186456.c0.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pmar	DN40950.c1.g1.i2.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pmax	XP_033755818.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Poku	DN16718.c0.g1.i6.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Ppur	DN7268.c0.g1.i11.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pstr	KAK3582760.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pvir	s00451g108	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Pyes	XP_021340986.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Rdec	DN22482.c4.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Rphi	XP_060578490.1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Scon	Chr15.1899	Sox	Sox-H	High mobility group box (438820)	N/A	-
Sglo	Sg1010047	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-
Tsqu	DN874.c7.g1.i1.p1	Sox	Sox-H	High mobility group box Sox-30 (438842)	N/A	-

Supplementary Table S3.7. Proportions of missing data in both DSFGs and bivalve species. Bivalve species represented by transcriptomic data are highlighted with an asterisk ('*').

Group	Genes	% missing data (out of 43 bivalve species)
	<i>Dmrt-1L</i>	48.837209
	<i>Dmrt-3</i>	30.232558
	<i>Dmrt-2</i>	55.813953
	<i>Dmrt-4/5</i>	6.976744
	<i>Fox-A</i>	13.953488
	<i>Fox-B</i>	30.232558
	<i>Fox-C</i>	23.255814
	<i>Fox-D</i>	20.930233
	<i>Fox-E</i>	30.232558
	<i>Fox-F</i>	18.604651
	<i>Fox-G</i>	16.279070
	<i>Fox-H</i>	41.860465
	<i>Fox-J1</i>	0.000000
	<i>Fox-J2/3</i>	9.302326
	<i>Fox-L1</i>	18.604651
	<i>Fox-L2</i>	13.953488
	<i>Fox-N1/4</i>	6.976744
	<i>Fox-N2/3</i>	6.976744
	<i>Fox-O</i>	13.953488
	<i>Fox-P</i>	9.302326
	<i>Fox-Q2</i>	30.232558
	<i>Fox-OG13/NA</i>	32.558140
	<i>Fox-OG15/NA</i>	34.883721
	<i>Fox-OG16/NA</i>	30.232558
	<i>Fox-OG2/NA</i>	16.279070
	<i>Fox-OG28/NA</i>	39.534884
	<i>Fox-OG39/NA</i>	37.209302
	<i>Sox-B1/2</i>	0.000000
	<i>Sox-C</i>	0.000000
	<i>Sox-D</i>	4.651163
	<i>Sox-E</i>	9.302326
	<i>Sox-F</i>	13.953488
	<i>Sox-H</i>	20.930233

Species	% missing data (out of 33 DSFGs)
Species	
<i>A. irridians concentricus</i>	0.000000
<i>A. marissinica</i>	21.212121
<i>A. pectinata</i> *	48.484848
<i>A. purpuratus</i>	6.060606
<i>C. angulata</i>	6.060606
<i>C. ariakensis</i>	3.030303
<i>C. fluminea</i> *	42.424242
<i>C. gigas</i>	6.060606
<i>C. plicata</i> *	21.212121
<i>C. sinensis</i>	21.212121
<i>C. virginica</i>	3.030303
<i>D. polymorpha</i>	9.090909
<i>H. bialata</i>	9.090909
<i>L. orbiculatus</i> *	63.636364
<i>M. arenaria</i>	21.212121
<i>M. californianus</i>	9.090909
<i>M. chinensis</i> *	57.575758
<i>M. coruscus</i>	0.000000
<i>M. edulis</i>	3.030303
<i>M. galloprovincialis</i>	6.060606
<i>M. margaritifera</i>	6.060606
<i>M. mercenaria</i>	3.030303
<i>M. modiolus</i> *	36.363636
<i>M. nervosa</i>	27.272727
<i>M. philippinarum</i>	9.090909
<i>O. edulis</i>	6.060606
<i>P. coreanum</i> *	18.181818
<i>P. generosa</i> *	54.545455
<i>P. margaritifera</i> *	21.212121
<i>P. maximus</i>	0.000000
<i>P. okutanii</i> *	54.545455
<i>P. purpuratus</i> *	54.545455
<i>P. streckersoni</i>	6.060606
<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

Chapter 4

Expression patterns of three sex-related genes and the germline marker *Vasa* in early developmental stages of *Mytilus galloprovincialis* embryos

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.*

In preparation.

4.1 Introduction

In preparation.

4.2 Materials and Methods

4.2.1 Time-series gene expression

Miglioli et al., 2024 recently produced one of the very first detailed developmental transcriptome of *M. galloprovincialis*, spanning from the unfertilized oocyte to the larval stage at 72 hpf, with time points sampled every 4 hpf. A total of 30 different mRNA libraries was sequenced, consisting of fifteen developmental time points per two technical replicates. These data are very useful to thoroughly investigate the transcription patterns of genes throughout the first three days of development in *M. galloprovincialis* and to obtain hints on the expected outcomes of mRNA-ISH experiments.

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 normalized using the built-in function `vst` of the package DESeq2 (Love et al., 2014). The function `plotPCA` was then used to run a principal component analysis (PCA) on read mapping counts and visualize the corresponding results. Normalized gene counts were also used to plot expression values of target genes (i.e., *Vasa*, *Dmrt1L*, *SoxH* and *FoxL2*), as well as in maSigPro (Conesa et al., 2006) to run a differential gene expression analysis in a time course experiment.

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

4.2.2 Sample collection, MitoTracker staining and fixation

Adult Mediterranean mussels (*M. galloprovincialis*) were hand collected from various locations surrounding the AltaSea institute at the port of Los Angeles (CA, USA). Sampling took place during the late spawning season of the species in California, i.e., from October 2023 to early January 2024. Specimens were checked for species and sexual maturity before usage.

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 acclimatized 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 30-40 minutes. As soon as individual mussels started spawning, they were promptly removed from the common tank, carefully washed and then allowed to continue spawning in isolated containers of about 250 ml 16°C FASW.

Sperm from six males and oocytes from six females were separately mixed to increase the number of crosses. An 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 minutes to let them assume a proper circular shape. Oocyte abundance was estimated under a stereo microscope, by counting the number of gametes in five aliquotes of 1 mL and then calculating the mean value. Sperm mitochondria were labeled with MitoTracker™ Red CMXRos (Thermo Fisher Scientific) at a working concentration of 500 nM for 30 minutes. MitoTracker is a vital and fixation-resistant mitochondrial dye and was used to be able 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.

Fertilization was performed by mixing oocytes and sperm at a ratio of 1:10. Fertilization success was checked after 20-30 minutes by the formation of polar bodies. The suspension was then carefully washed 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°C in the dark. Water was changed every 24 hours. After 48 hpf, larvae were fed with *Isochrysis galbana* at a final concentration of circa 100,000 cells/mL following **Helm et al., 2004**.

Embryos/larvae were sampled at 1, 2, 3 and 4 hpf, and then every 12 hours until 72 hpf,

Target	Amplifier	Fluorophore	No. of probe pairs
<i>Vasa</i>	B1	ALEXA-488	33
<i>Dmrt1L</i>	B2	ALEXA-647	18
<i>SoxH</i>	B3	ALEXA-546	22
<i>FoxL2</i>	B4	ALEXA-700	28

Table 4.1. List of genes targeted through HCR, with the corresponding amplifiers, fluorophores and number of generated probe pairs.

every time after checking for proper development and vitality. After concentration in a mesh of proper size, embryos/larvae were fixed in 3.2% paraformaldehyde (PFA) in 1× PBS at 4°C overnight under constant and gentle shaking. Fixed samples were washed 3 × 20 minutes in 1× PBS 0.1% Tween 20 (PBST) and then dehydrated 3 × 30 minutes in absolute methanol at room temperature (RT). Dehydrated samples were stored at -20°C until usage.

4.2.3 mRNA *in-situ* Hybridization Chain Reaction (HCR)

HCR probe design

Vasa, *Dmrt1L*, *SoxH*, and *FoxL2* spliced-transcript nucleotide sequences of *M. galloprovincialis* were obtained from previous analyses with OrthoFinder v2.5.5 (Emms and Kelly, 2019) and 30 annotated bivalve genomes (see Chapter 3). Accession numbers of spliced transcripts are 10B017427, 10B093608, 10B014180, and 10B094018, respectively. The `insitu_probe_generator` script from 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 in Tab. 4.1. Resulting probes were synthetized by Integrated DNA Technologies (IDT™) in different oligo pools.

Fluorescent *in-situ* hybridization through hybridization chain reaction and microscope imaging

HCR mRNA-FISH 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.

Target	Dye	Excitation (nm)	Emission (nm)
dsDNA (nuclei)	DAPI	360	460
Sperm mitochondria	MitoTracker™ Red CMXRos	575	600
<i>Vasa</i>	ALEXA-488	499	520
<i>Dmrt1L</i>	ALEXA-647	653	670
<i>SoxH</i>	ALEXA-546	557	575
<i>FoxL2</i>	ALEXA-700	685	700

Table 4.2. List of dyes used for every target, together with the excitation and emission peaks as returned by the Las X software.

Dehydrated samples stored in methanol were washed 4 times per 5 minutes and 1 time per 10 minutes in a phosphate-buffered saline solution (PBS; 128 mM NaCl, 2 mM KCl, 8 mM Na₂HPO₄ · 2H₂O, 2 mM KH₂PO₄) with 0.1% Tween 20 (PBST). Samples were then permeabilized for 30 minutes in a detergent solution (1.0% SDS, 0.5% Tween 20, 50 mM Tris-HCl, 1.0 mM ethylenediaminetetraacetic acid (EDTA), 150.0 mM NaCl) and washed again 2 times per 5 minutes in PBST. Samples were prepared for the HCR detection stage by incubation in probe hybridization buffer for 30 minutes 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 was removed by washing 4 times per 20 minutes with probe wash buffer at 37 °C and 3 times per 5 minutes with 5× saline-sodium citrate Tween 20 buffer (SSCT; 5× SSC, 0.1% Tween 20) at room temperature. Samples were incubated for 30 minutes in amplification buffer at room temperature. Hairpins were heated at 95 °C for 90 seconds and then snap-cooled at room temperature for 30 minutes. The amplification step of HCR was performed with 6 pmol of each hairpin in amplification buffer overnight (>12 h) at room temperature.

Excess hairpins was removed by washing 2 times per 5 minutes, 2 times per 30 minutes, and 1 time per 5 minutes with SSCT. If not immediately mounted on slides, samples were stored in SSCT at +4 °C. Otherwise, samples were immersed in 50% and 75% glycerol for 30-60 minutes each, 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 any crosstalks. **Tab. 4.2** summarises the excitation and emission peaks for each dye. Images were then manipulated and post-produced using Fiji v2.14.0.

4.2.4 Immunolocalization of Vasa

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 from fading.

Dehydrated samples stored in methanol were rinsed 3 times per 10 minutes and 1 time for 2 hours in Tris-buffered saline (TBS; 10 mM Tris-HCl, 155 mM NaCl), following an additional wash for 10 minutes with PBS. Samples were then digested for 6 minutes and 30 seconds with 0.01% pronase E (Merck) in PBS, and washed again 2 times for 5 minutes in PBS. Permeabilization was then performed in TBS-Triton (TBST) 0.1% for 5 minutes at RT and in TBST 1% overnight at 4°C.

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

Excess primary antibody was rinsed from samples with 4 washes of 30 minutes in TBST 0.1%, while non-specific protein-binding sites were blocked again with an incubation of 1 hour in TBST 0.1% containing 3% BSA. Samples were then incubated at 4°C for 24-32 hours with secondary antibody HRP anti-rabbit in goat (Santa Cruz Biotechnology Inc.) dilutied 1:400. Excess secondary antibody was rinsed with 4 washes of 30 minutes in TBST 0.1% and 1 wash of 1 hour in TBST 1%.

Samples were immersed in 50% and 75% glycerol for 30-60 minutes each, and then mounted with VECTASHIELD® PLUS Antifade Mounting Medium with DAPI (H-2000). Slides were imaged COMPLETECOMPLETECOMPLETECOMPLETE. Each dye was imaged sequentially in a separate channel, to enhance the yield and avoid any crosstalks. **Tab. 4.2** summarises the excitation and emission peaks for each dye. Images were then manipulated and post-produced using Fiji v2.14.0.

4.3 Results

In preparation.

4.4 Discussion

In preparation.

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.
- 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.
- 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.
- 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 comparative 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.
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using diamond. *Nature methods*, 12(1), 59–60.

- 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., Adrián, J. R., Liao, Y., Montooth, 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.
- 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.
- Cutter, A. D., & Ward, S. (2005). Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Molecular Biology and Evolution*, 22(1), 178–188.
- 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.
- 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.
- 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.
- 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., 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.
- 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.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from rna-seq data. *Nature biotechnology*, 29(7), 644.
- 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.
- Haas, B. J. (n.d.). *Transdecoder*. <https://github.com/TransDecoder>
- 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.
- 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.
- 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.
- 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.

- 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.
- Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., & Paabo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, 309(5742), 1850–1854.
- 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.
- Kopp, A. (2012). Dmrt genes in the development and evolution of sexual dimorphism. *Trends in Genetics*, 28(4), 175–184.
- Kousathanas, A., Halligan, D. L., & Keightley, P. D. (2014). Faster-x adaptive protein evolution in house mice. *Genetics*, 196(4), 1131–1143.
- 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.
- 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, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics*, 39(1), btad014.
- 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, 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*, 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.
- Lozano-Fernandez, J. (2022). A practical guide to design and assess a phylogenomic study. *Genome Biology and Evolution*, 14(9), evac129.
- 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*, 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.
- 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.
- 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.
- 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., & Sourdaïne, P. (2009). Molecular cloning and gene expression of cg-foxl2 during the development and the adult gametogenetic 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.
- Natsidis, P., Kapli, P., Schiffer, P. H., & Telford, M. J. (2021). Systematic errors in orthology inference and their effects on evolutionary analyses. *Iscience*, 24(2).
- 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 brachiopoda with emphasis on tadpole shrimps. *Frontiers in Ecology and Evolution*, 11, 1046960.
- 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.
- of Sex Consortium, T., et al. (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., Fontseré, 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.
- 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.
- Peñaloza, 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.

- Perez-Garcia, C., Moran, P., & Pasantes, J. J. (2011). Cytogenetic characterization of the invasive mussel species *Xenostrobus securis* lmk. (bivalvia: Mytilidae). *Genome*, 54(09), 771–778.
- 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 ophox complexes in bivalves. *Molecular Biology and Evolution*, 38(6), 2597–2614.
- 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.
- 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.
- 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.

- 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.
- Sanchez-Flores, A., Peñaloza, F., Carpineteyro-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.
- 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.
- 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.
- Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature reviews genetics*, 3(2), 137–144.
- 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.
- 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*, 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*, 2020, vol. 20, num. 5, p. 1445-1452.
- 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.

- 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.
- 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.
- 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.
- 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.

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

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

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

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

⁴*School 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.

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. dagleishae* 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

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

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

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

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

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

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 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 profiles 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 million years ago, 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.