Dev Genes Evol (2014) 224:175–181
DOI 10.1007/s00427-014-0473-0

SEQUENCE CORNER

Pan-metazoan phylogeny of the DMRT gene family: a framework for functional studies

Judith R. Wexler · David C. Plachetzki · Artyom Kopp

Received: 7 March 2014 / Accepted: 21 May 2014 / Published online: 6 June 2014

Abstract The family of Doublesex-Mab-3 Related Transcription factors (DMRTs) includes key regulators of sexual differentiation and neurogenesis. To help understand the functional diversification of this gene family, we examined DMRT gene complements from the whole genome sequences and predicted gene models of 32 animal species representing 12 different phyla and from several non-metazoan outgroups. DMRTs are present in all animals except the sponge *Amphimedon queenslandica*, but are not found in any of the outgroups, indicating that this gene family is specific to animals and has an ancient pre-eumetazoan origin. Our analyses suggest that DMRT genes diversified independently in bilaterian and non-bilaterian animals. Most clades in the DMRT gene tree, including those containing the well-characterized *DMRT1* and *doublesex* genes, have phylogenetically limited distributions.

Keywords Doublesex-mab-3 related transcription factor · Gene tree · Phylogenetics · Sex determination · Evolution

Communicated by: M. Q. Martindale

Judith R. Wexler and David C. Plachetzki contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00427-014-0473-0) contains supplementary material, which is available to authorized users.

J. R. Wexler (✉) · A. Kopp
Center for Population Biology, Department of Evolution and Ecology, University of California, One Shields Ave, Davis, CA 95616, USA
e-mail: jrwexler@ucdavis.edu

A. Kopp
e-mail: akopp@ucdavis.edu

D. C. Plachetzki
Department of Molecular, Cellular, & Biomedical Sciences, University of New Hampshire, Durham, NH, USA
e-mail: david.plachetzki@unh.edu

Introduction

Sexually dimorphic phenotypes are present in nearly all animals. Among the wide variety of genetic machinery responsible for the development of sex-specific traits, Doublesex-Mab-3 Related Transcription factors (DMRTs) are unique for their conservation in disparate animal phyla (Raymond et al. 1999; Volff et al. 2003). Most animal genomes contain multiple DMRT genes with different functions and expression patterns. Various subsets of this gene family have been examined in vertebrates, nematodes, arthropods, mollusks, cnidarians, and tunicates (Raymond et al. 1999; Mason et al. 2008; Burtis and Baker 1989; Yu et al. 2011; Miller et al. 2003; Tresser et al. 2010).

As regulators of sex-specific cell fates, DMRT genes have varied roles. One of the best studied members of the gene family—the insect *doublesex* gene—is required for both male and female development and is present in alternative splicing isoforms in the two sexes. Each isoform binds to the same recognition sequence but has sex-specific effects on transcriptional regulation (Coschigano and Wensink 1993). *doublesex* acts at the bottom of the sex determination pathway, but other DMRT genes, such as the male-specific DMY in medaka, are primary sex determiners (Matsuda et al. 2007). Some DMRT genes, such as DMRT1 in mice, are necessary only at specific developmental time points, while others must be expressed constitutively throughout adulthood to maintain sexual identity (Matson et al. 2011).

While most studies of DMRT gene function focus on the roles played by these transcription factors in sexual differentiation, these genes have also been shown to be essential for processes as varied as *Xenopus laevis* olfactory placode neurogenesis (Parlier et al. 2013) and zebrafish somite formation (Meng et al. 1999).

Interpretation of the similarities and differences in DMRT expression and function across animal lineages is complicated by the lack of a comprehensive phylogeny of the gene family. DMRT gene sequences are highly diverged, with detectable homology limited to the DNA-binding DM domain (Volff et al. 2003). The short length of this region (~70 amino acids) presents significant difficulty in resolving phylogenetic relationships among DMRT paralogs, and poor resolution of deep nodes in the DMRT tree has undermined our ability to infer the ancestral developmental function of this gene family.

Here, we use 33 genomes representing 12 phyla to examine the evolution of the DMRT gene family and to establish a phylogenetic framework for the ongoing analysis of the developmental functions of DMRT genes in different lineages. Our analysis shows that although this gene family evolved before the common ancestor of all eumetazoans, most DMRT paralogs, including the best studied ones, have restricted phylogenetic distribution.

Materials and methods

Dataset assembly

Our DMRT dataset was assembled using an in-house bioinformatics pipeline. First, amino acid sequences of the DM domains of nine well-characterized vertebrate and insect DMRT genes (Table S1) were used as queries in BLASTp (Altschul et al. 1997) searches of predicted gene models from whole genome sequences of 33 phylogenetically informative taxa (Table S2). BLAST searches were conducted with an Expect value of 0.01, and the 20 highest scoring sequences were retained from each taxon's gene model. Redundant sequences from the combined datasets were removed. BLAST hits were also removed if they did not contain, upon initial alignment, a canonical CCHHCC motif characteristic of the DMRT DNA-binding domain (Zhu et al. 2000). Our complete protein dataset is found in Supplemental Fig. 1. For each phylogenetic analysis, full-length sequences were aligned under the—auto setting in MAFFT (Katoh et al. 2002) and low-scoring regions of the alignment were removed using the *automated1* heuristic implemented in trimAl (Capella-Gutierrez et al. 2009). This approach maximizes the alignable portion of the dataset, involves no hand manipulations of the data and is reproducible. Alignments used are found in Supplemental Figs. 2 and 3.

Phylogenetic analyses

We conducted phylogenetic analyses using both maximum likelihood (ML) and Bayesian Monte Carlo Markov Chain (BMCMC) approaches. ML analyses were conducted in RAxML v7.2.8 (Stamatakis 2006) under the best-fit model determined using ProtTest3 (Abascal et al. 2005) (Table 1). In preliminary ML analyses, we noticed that the starting tree had an appreciable effect on the resulting ML topology. In order to better circumscribe the ML topology for animal DMRT genes, we conducted 100 random starts of the data matrix and selected the highest scoring topology as the best-known likelihood tree (BKL) as described in the RAxML documentation (Stamatakis 2006). We used the BKL tree to summarize the bipartitions from (1) all 100 random start ML trees, (2) 1000 bootstrap (BS) replicates, and (3) posterior proportions from BMCMC, resulting in the consensus BKL tree (Figs. 1 and 2). Bootstrap (BS) replicates were conducted under the best-fit model in RaxML (Stamatakis 2006). BMCMC analyses were run for 30,000,000 generations under the mixed amino acid model and default parameters in MrBayes 3.2 (Ronquist et al. 2012). Our total metazoan DMRT dataset included several loci on long branches (see below). For analyses of a pruned dataset (Fig. 2), branch length statistics were estimated using in-house python scripts. Table 1 summarizes all phylogenetic analyses conducted. Scripts used in the production of our datasets are available upon request.

We were unable to identify a likely outgroup with which to root phylogenetic analyses. DMRT genes bear no obvious informative homology to other zinc finger transcription factor DNA-binding domains, most of which bind DNA in the major groove, whereas the DM domain binds in the minor grove (Zhu et al. 2000). Phylogenetic results were midpoint-rooted. Using the BKL trees (Figs. S4, S5, and S6), we assigned orthology classes based on named, previously characterized DMRT loci that were used as queries in dataset construction (bold in Figs. 1 and 2).

Table 1 Summary of phylogenetic analyses

Dataset	Matrix length (amino acids)	# of genes in dataset	Model	Standard deviation of split frequencies after 3×10^7 generations (BMCMC)
Total DMRT (Fig. 1)	50	151	JTT+G	0.015944
Reduced DMRT (Fig. 2)	46	91	JTT+G	0.008122
Arthropod and vertebrate DMRT (S8)	242	67	WAG+G	0.002725

Results and discussion

Metazoan DMRT phylogeny

Alignable regions of homology in DMRT sequences localize to the DNA-binding DM domain. Phylogenetic analysis of this dataset presents considerable challenges, which derive from the short alignment and from the presence of several probable long-branch loci. Given these inherent challenges, our approach was to first identify sequences that were likely responsible for statistical errors in our analyses, and then reanalyze the data following their removal. Our initial phylogenetic analysis included the total metazoan DMRT dataset and included 151 sequences from 30 taxa. This dataset consisted of a matrix of 50 residues after alignment and trimming. Results from consensus BKL (Stamatakis 2006) (Fig. S4), bootstrap (BS) (Fig. S7), and Bayesian MCMC (BMCMC) (Fig. S8) analyses of this dataset showed poor resolution. Phylogenetic results for this dataset are summarized in Fig. 1.

Our analysis of all gene models from all taxa yielded a poorly supported group of long-branched sequences that contained genes from multiple phyla including the well-characterized *Caenorhabditis elegans mab-3* and *doublesex* genes from branchiopod crustaceans of the genus *Daphnia*. The placement of *Daphnia* sequences differed from previous analyses that used smaller taxon samples (Toyota et al. 2013; Kato et al. 2011) with longer amino acid alignments. To investigate potential problems of our dataset, we combined arthropod and vertebrate sequences from our data matrix with the dataset reported in a recent study of pancrustacean DMRT phylogeny (Toyota et al. 2013). This dataset consisted of 67 genes with a matrix length of 242 residues. BKL, BMCMC, and BS analyses resolved a well-supported clade of arthropod *dsx* genes that included the *Daphnia* and other crustacean sequences (summarized in Fig. S9). Improved support for the clades in this tree could be the result of removal of long-branched taxa, longer amino acid alignments, or both.

In order to address the effect of poorly supported long branching taxa in our pan-metazoan DMRT phylogeny, we removed all sequences with branch lengths greater than 1.0 standard deviations from the mean branch length across the tree. Results of BKL (Fig. S6), BS (Fig. S10) and BMCMC (Fig. S11) analyses of this reduced dataset are summarized in Fig. 2.

DMRT genes originated before the evolution of eumetazoans

We included gene models from seven non-bilaterian animals in our analyses: *Acropora millepora*, *Amphimedon queenslandica*, *Hydra magnipapillata*, *Mnemiopsis leidyi*, *Monosiga brevicollis*, *Nematostella vectensis*, and *Trichoplax adhaerans*. All of these organisms contained DMRT domains in their genomes with the exception of the sponge *Amphimedon* and the choanoflagellate *Monosiga*. In addition, tBLASTn analyses of other metazoan ougroups showed that no DM domain sequences are present in the choanoflagellate *Salpingoeca rosetta*, other single-celled Opisthokonts (*Sphaeroforma arctica* and *Capsaspora owczarzaki*), in any fungi including basal fungal lineages (*Allomyces macrogynus*, *Mortierella elongata*, and *Spizellomyces punctatus*), or non-opisthokont protists (e.g., *Naegleria gruberi*). The presence of DMRT genes in *Trichoplax* and *Mnemiopsis* indicates that this gene family arose early in metazoan diversification. *Mnemiopsis*, *Hydra*, and *Acropora*, but not *Trichoplax* DMRT genes, contain DMA domains, suggesting that this domain arose during the interval between *Trichoplax* and eumetazoans.

Phylogenetic distribution of DMRT paralogs

The phylogeny constructed from our reduced dataset reveals eight monophyletic groups containing previously named DMRT genes (Fig. 2). Most clades for which we find strong support have limited phylogenetic distribution (Fig. 3). Only two ortholog groups, DMRT93B and DMRT2/11E, contain sequences from deuterostomes, lophotrochozoans, and ecdyzosoans (Fig. 3). DMRT2 appears to play a wide variety of roles in diverse metazoan taxa. In vertebrates, DMRT2 functions in somitogenesis, and in zebrafish it also regulates left-right axial patterning (Saude et al. 2005). A study in the frog *Rana rugosa* found DMRT2 expression in the gonads of developing tadpoles, but no sexual dimorphism in expression levels was observed. In the oyster *Pinctada martensii*, DMRT2 is expressed in developing male germ cells (Kim et al. 2011). The only ecdysozoan data comes from the crustacean *Daphnia magna*, where DMRT11E transcription in the ovaries is higher than in testes (Kato et al. 2011). Much less is known about DMRT93B. Its expression has only been examined in the crustacean *D. magna*, where the levels of DMRT93B transcript change in response to juvenile hormone exposure (Kim et al. 2011).

The two DMRT paralogs with well-characterized roles in sexual differentiation—*DMRT1* and *dsx*—have mutually exclusive phylogenetic distributions (Figs. 2 and 3). *DMRT1* function is best studied in mice, where it is necessary for the maintenance of male-specific gonad differentiation (Matson et al. 2011). DMRT1 orthologs also function in male sexual development in other mammals (including humans) birds, amphibians, and teleost fishes (Kopp 2012; Zarkower 2013). Our analysis shows strong support for a DMRT1 clade containing only vertebrate sequences (Fig. 2). DMRT1 orthologs are found in all examined Osteichthyes (bony fish and tetrapods), but not in the genomes of the shark *Callorhinchus milii*, the urochordates *Ciona intestinalis* and *Oikopleura dioica*, the cephalochordate *Branchiostoma floridae*, or in any non-chordate deuterostome phyla.

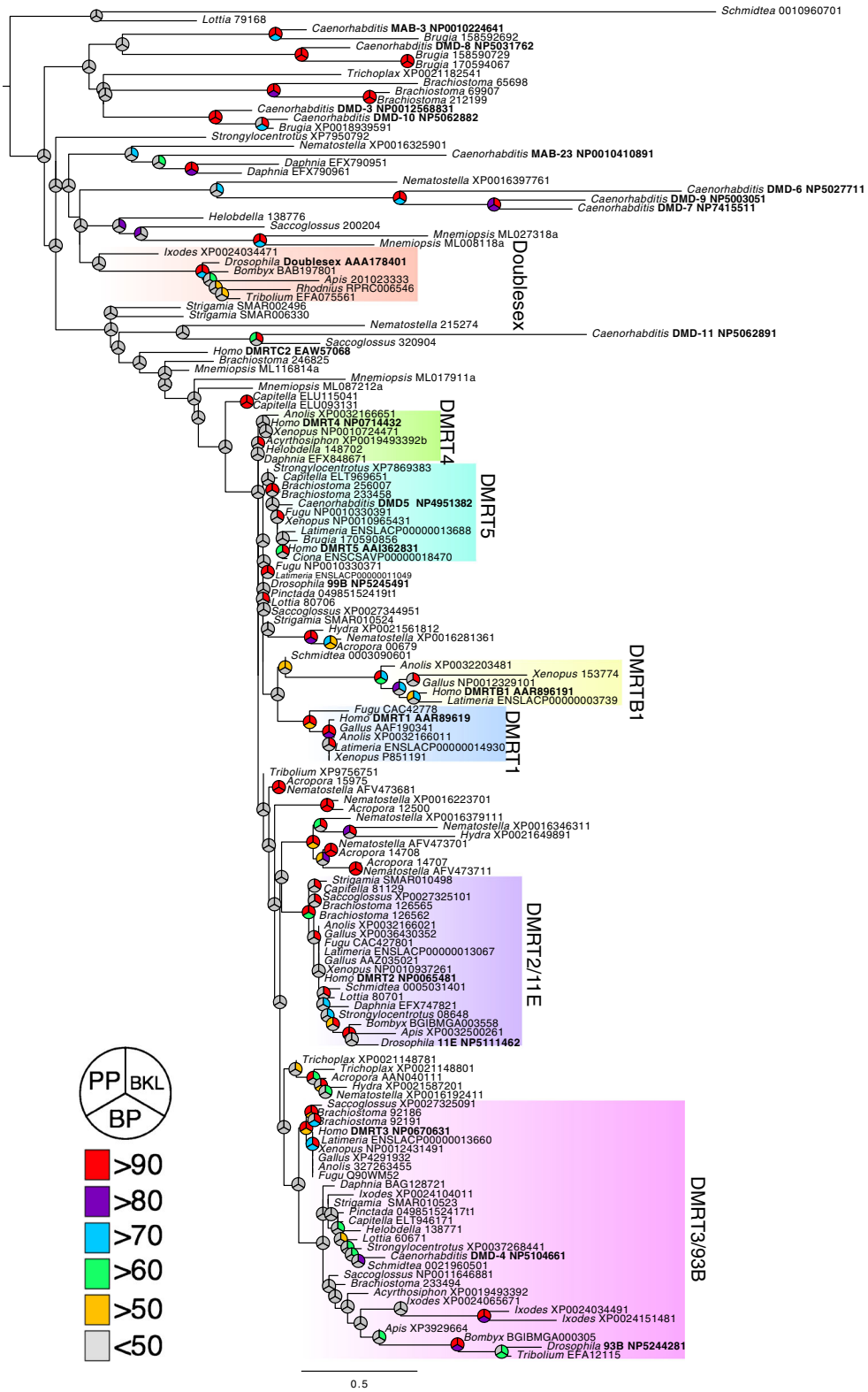

Fig. 1 Summary of phylogenetic analyses of the total metazoan DMRT dataset. Topology shown is the best-known likelihood (BKL) tree from 100 random parsimony starts (see "Materials and methods"). *Pie charts* indicate BKL support (*BKL*, *top right*), bootstrap support (*BP*, *bottom*) and Bayesian posterior probabilities (*PP*, *top left*)

dsx, which directs male and female development via alternatively spliced sex-specific isoforms in *Drosophila melanogaster* and other holometabolous insects, only has clear orthologs in arthropods (Fig. 3) (Kopp 2012; Zarkower 2013).

None of the DMRT genes from non-bilaterian animals fall out in any of the supported bilaterian DMRT clades (Figs. 1 and 2). However, several cnidarian DMRT genes are recovered as outgroups to bilaterian lineages, e.g., the

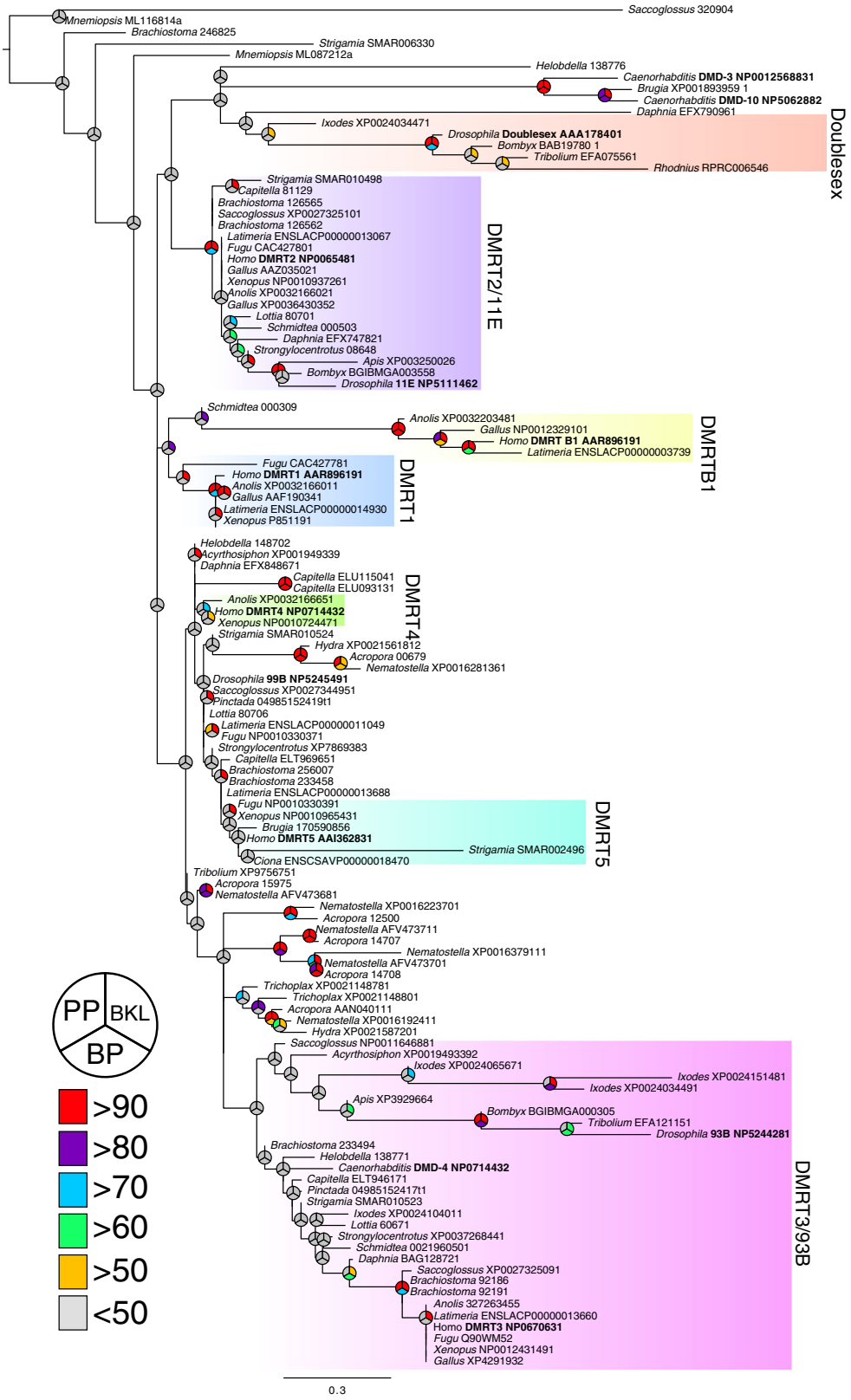

Fig. 2 Summary of phylogenetic analyses of the reduced metazoan DMRT dataset. Topology shown is the best-known likelihood (BKL) tree from 100 random parsimony starts (see "Materials and methods"). *Pie charts* indicate BKL support (*BKL*, *top right*), bootstrap support (*BP*, *bottom*) and Bayesian posterior probabilities (*PP*, *top left*)

pan-bilaterian DMRT2/11E clade. In the restricted dataset that excludes longer branching tips, most non-bilaterian genes cluster together at the base of the tree (Fig. 2), suggesting that DMRT genes diversified independently

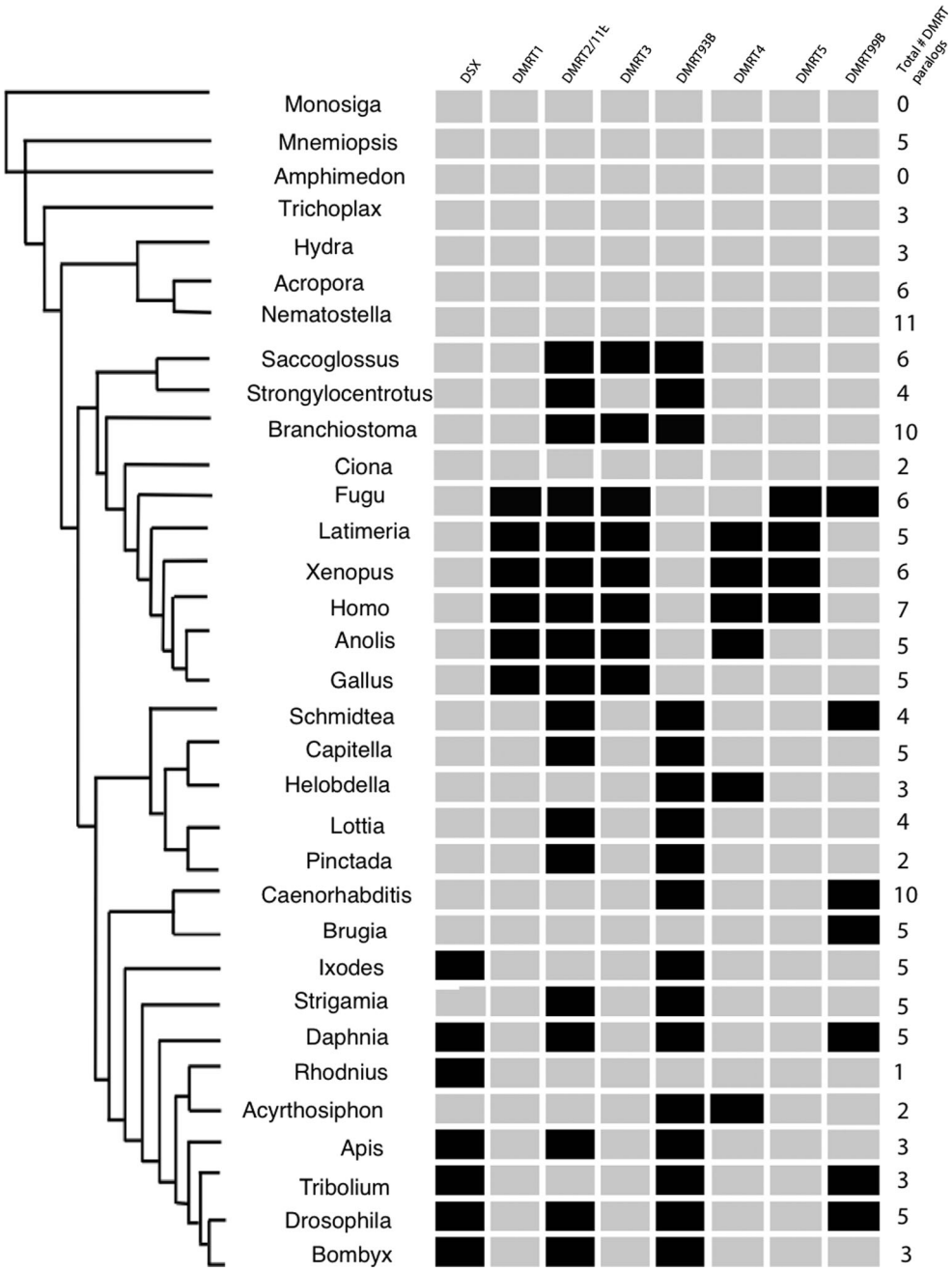

Fig. 3 Metazoan phylogeny indicating the presence (*dark gray*) or absence (*light gray*) of DMRT paralogs. A paralog group is defined as a monophyletic group containing at least three genes

in bilaterian and non-bilaterian animals. Functional analysis of these genes will be essential for reconstructing the diversification of the DMRT gene family, especially for understanding the origin of its roles in sexual differentiation and neurogenesis.

The domain structure of metazoan DMRTs

Our total DMRT dataset represents the complete complement of DM domain-containing genes for each of the taxa included in this study. These data allow the characterization of other protein domains present in DMRT genes, which can shed further light on the molecular function of these genes. The DMA domain, found in several previously described DMRT genes, is present in a wide range of metazoan phyla, including cnidarians and ctenophores. There is little phylogenetic signal to the presence of the DMA domain, although it appears that none of the *doublesex* orthologs possess a DMA domain. Functional data on the DMA domain is limited to one study in *Xenopus* which implicated the domain in neurological processes (Parlier et al. 2013).

Conclusion

This phylogenomic analysis provides a framework for experimental research aimed at understanding the function of DMRT genes, including the ancestral role of this gene family and the origin its functions in sexual differentiation. By presenting a full paralog complement of DMRT genes from multiple phyla, we also raise several new questions. For example, what roles do DMRT93B and DMRT2/11E—the only paralogs with broad phylogenetic distribution—play in different taxa? Lastly, we hope that these gene trees will help place current DMRT research in proper phylogenomic context.

Acknowledgments This work was supported by the National Institutes of Health (R01 GM082843 to A.K.); the Howard Hughes Medical Institute Life Sciences Research Foundation postdoctoral fellowship (to D.P.); and the UC—Davis Population Biology Graduate Group (to J.W.). We also thank two anonymous reviewers for their comments on the manuscript.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Burtis KC, Baker BS (1989) Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56:997–1010

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973

Coschigano KT, Wensink PC (1993) Sex-specific transcriptional regulation by the male and female doublesex proteins of Drosophila. Genes Dev 7:42–54

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Kato Y, Kobayashi K, Watanabe H, Iguchi T (2011) Environmental sex determination in the branchiopod crustacean *Daphnia magna*: deep conservation of a *Doublesex* gene in the sex-determining pathway. PLoS Genet 7:e1001345

Kim J, Kim Y, Lee S, Kwak K, Chung WJ, Choi K (2011) Determination of mRNA expression of *DMRT93B*, *vitellogenin*, and *cuticle 12* in *Daphnia magna* and their biomarker potential for endocrine disruption. Ecotoxicology 20:1741–1748

Kopp A (2012) Dmrt genes in the development and evolution of sexual dimorphism. Trends Genet 28:175–184

Mason DA, Rabinowitz JS, Portman DS (2008) *dmd-3*, a *doublesex*-related gene regulated by *tra-1*, governs sex-specific morphogenesis in *C. elegans*. Development 135:2373–2382

Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D (2011) *DMRT1* prevents female reprogramming in the postnatal mammalian testis. Nature 476:101–104

Matsuda M et al (2007) DMY gene induces male development in genetically female (XX) medaka fish. Proc Natl Acad Sci 104:3865–3870

Meng A, Moore B, Tang H, Yuan B, Lin S (1999) A *Drosophila doublesex*-related gene, *terra*, is involved in somitogenesis in vertebrates. Development 126:1259–1268

Miller SW, Hayward DC, Bunch TA, Miller DJ, Ball EE, Bardwell VJ, Zarkower D, Brower DL (2003) A DM domain protein from a coral, *Acropora millepora*, homologous to proteins important for sex determination. Evol Dev 5:251–258

Parlier D, Moers V, Van Campenhout C, Preillon J, Lecière L, Saulnier A, Sirakov M, Busgendal H, Kricha S, Marine JC, Rentzch F, Bellefroid EJ (2013) The Xenopus doublesex-related gene Dmrt5 is required for olfactory placode neurogenesis. Dev Biol 373:39–52

Raymond CS, Kettlewell JR, Hirsch B, Bardwell VJ, Zarkower D (1999) Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. Dev Biol 215:208–220

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542

Saude L, Lourenco R, Goncalves A, Palmeirim I (2005) *terra* is a left-right asymmetry gene required for left-right synchronization of the segmentation clock. Nat Cell Biol 7:918–920

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Toyota K, Kato Y, Sato M, Sugiura N, Miyagawa S, Miyakawa H, Watanabe H, Oda S, Ogino Y, Hiruta C, Mizutani T, Tatarazako N, Paland S, Jackson C, Colbourne J, Iguchi T (2013) Molecular cloning of *doublesex* genes of four cladocera (water flea) species. BMC Genomics 14:239

Tresser J, Chiba S, Veeman M, El-Nachef D, Newman-Smith E, Horie T, Tsuda M, Smith WC (2010) *doublesex/mab3 related-1 (dmrt1)* is essential for development of anterior neural plate derivatives in *Ciona*. Development 137:2197–2203

Volff JN, Zarkower D, Bardwell VJ, Schartl M (2003) Evolutionary dynamics of the DM domain gene family in metazoans. J Mol Evol 57(Suppl 1):S241–S249

Yu F-F, Wang M-F, Zhou L, Gui J-F, Yu X-Y (2011) Molecular cloning and expression characterization of *Dmrt2* in Akoya Pearl Oysters, *Pinctada martensii*. J Shellfish Res 30:247–254

Zarkower D (2013) DMRT genes in vertebrate gametogenesis. Curr Top Dev Biol 102:327–356

Zhu L, Wilken J, Phillips NB, Narendra U, Chan G, Stratton S, Kent S, Weiss M (2000) Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. Gene Dev 14: 1750–1764