Multiple alternative splicing of Dmrt1 during gonadogenesis in Indian mugger, a species exhibiting temperature-dependent sex determination

Amit Anand, Minarbha Patel, Albert Lalremruata, Ajeet Pratap Singh, Raman Agrawal, Lalji Singh, Ramesh K. Aggarwal *

Centre for Cellular and Molecular Biology, Council of Scientific and Industrial Research, Uppal Road, Tarnaka, Hyderabad 500007, India

A R T I C L E   I N F O

Article history:
Received 28 April 2008
Received in revised form 4 August 2008
Accepted 5 August 2008
Available online 13 August 2008

Keywords:
Indian mugger
cpDmrt1
GSD
TSD
TSP
Alternative splicing

A B S T R A C T

Dmrt1 is an evolutionarily conserved gene having important role in the sex determination from lower vertebrates to mammals. Recent studies show transcriptional diversity for this important gene during gonadal differentiation in a few vertebrate species having genetic sex determination (GSD). In this study, we show for the first time that the transcriptional diversity of Dmrt1 is also found in the Indian mugger that exhibits temperature-dependent sex determination (TSD). We report here isolation and characterization of eight novel isoforms of Dmrt1 from Crocodylus palustris, along with its genomic locus that is referred as, cpDmrt1. Further, by sequence comparisons of cpDmrt1 and its expressed isoforms, we demonstrate that all the isoforms are generated by alternative splicing, exonization of intronic sequences and alternative polyA sites from the same locus. The eight transcripts range from 494 to 2060 bp and encode six predicted proteins having the characteristic DM domain of Dmrt1. The major heterogeneity in the isoforms and their predicted proteins is seen only in their C-termini and 3′-UTRs, which do not match with any similar sequences reported for other vertebrates. The cpDmrt1 expression was seen mainly in developing GAM (genital ridge-adrenal-mesonephros complex) with significant upregulation only in male embryos from the start of the temperature sensitive period (TSP). More significantly, ~70% of this expression was contributed only by one isoform (cpDmrt1e) that also has a unique 15 amino acid domain towards its C-terminal. cpDmrt1 expression was also detected at a lower level in brain and developing kidney. The study thus provides the first account of Dmrt1 locus, its transcriptional diversity and sex-specific expression in Indian mugger, a TSD species.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Sex determination is an important phenomenon occurring during early stages of embryonic development. Its underlying mechanism(s) and evolution still remain key questions for geneticists and evolutionary biologists. During euakaryotic evolution, a number of sex determination mechanisms have evolved (Valenzuela and Lance, 2004), ranging from those based on highly specialized genes and specific sex-chromosomes (called genetic sex determination, GSD), to simple environmental factors (environmental sex determination, ESD). Of the latter, TSD (temperature-dependent sex determination) is the most common one, wherein incubation temperature of the developing embryos decides the sex. TSD is prevalent in lower vertebrates, especially in lizards, turtles and all extant crocodilian species (Valenzuela and Lance, 2004). The period during which temperature seemingly decides the sex of the developing embryo is called the temperature sensitive period (TSP).

Our current knowledge about the molecular mechanisms underlying sex determination is based on the extensive studies carried out on GSD that have revealed a number of genes including a male-determining master regulator gene SRY in mammals (Swain and Lovell-Badge, 1999; Wilhelm et al., 2007). In comparison, our understanding of the molecular/genetic basis of TSD remains poor. However, it is now apparent that many of the genes involved in GSD are seemingly conserved even in TSD, though with significant differences in their temporal–spatial expression, suggesting their functions may not be conserved (Morrish and Sinclair, 2002; Sarre et al., 2004; Yao and Capel, 2005). Moreover, major questions such as how temperature is sensed during embryonic development and how the effect of the temperature is transduced for sex determination of developing embryo are yet to be answered.

In vertebrates, Dmrt1 (doublesex-/mab-3 related transcription factor-1) is one such important gene that plays an important role in
male sex determination/differentiation (Raymond et al., 2000; Matsuda et al., 2002; Hodgkin, 2002; Huang et al., 2005; Shan et al., 2000). The gene derives its name from structurally and functionally related genes found in invertebrates viz., doublesex in Drosophila and mab-3 in C. elegans, which encode the DM domain-containing transcription factors, and are involved in the sex determination and differentiation process (Hodgkin, 2002). Dmrt1 is expressed only in the gonads in mammals and is upregulated during early stages of male gonad development. Knockout studies show that Dmrt1 is required for postnatal testis differentiation after determination, but dispensable for ovary development in mouse (Mus musculus) (Raymond et al., 2000). In human (Homo sapiens), deletions of 9p24 region (where Dmrt1 is located) result in XY sex-reversal (Raymond et al., 1999). In the chicken (Gallus gallus), Dmrt1 is sex-linked on the Z-chromosome and is expressed at higher levels in male than in the female gonad suggesting it to be important for sexual development especially for the male gonad (Nanda et al., 2000; Zhao et al., 2007). In the teleost fish medaka (Oryzias latipes) two Dmrt1 copies are present (Dmrt1a and DMY/Dmrt1bY), of which the DMY is master regulator for male sex determination (Matsuda et al., 2002). Even in the Rice eel (Monopterus albus), which naturally shuffles sex from female to male in lifespan, Dmrt-1 expression is upregulated when the testis development starts from its precursor ovotestis (Huang et al., 2005). Recently, two Dmrt1 genes, one autosomal and the other on W-chromosome (DM-W), have been identified in clawed frog (Xenopus laevis), of which DM-W is suggested to be involved in the ovarian development (Yoshimoto et al., 2008). These and similar studies demonstrate that invariably among all GSD vertebrates, Dmrt1 is highly conserved having in general a male-specific expression in the developing male gonads (Hodgkin 2002); and that it plays an very important role from being a master regulator to a downstream gene required for the male gonad development. Furthermore, recent studies in GSD species (human, mouse, chicken, zebrafish, rice eel) have revealed differentially expressed multiple transcribed isoforms of Dmrt1 (Huang et al., 2005; Zhao et al., 2007; Guo et al., 2005; Cheng et al., 2006; Lu et al., 2007), suggesting that Dmrt1 is regulated by alternative splicing during male sexual development. Compared to GSD species, there are only a few preliminary studies on Dmrt1 in TSD species like turtles and alligators that also show very early expression and upregulation at male promoting temperatures (Kettlewell et al., 2000; Sreenivasulu et al., 2002; Torres Maldonado et al., 2002; Murdock and Wibbels, 2003, 2006; Hattori et al., 2007; Rhen et al., 2007; Shoemaker et al., 2007). However, diverse transcriptions of Dmrt1 and its regulation in the gonadogenesis in TSD species remain to be demonstrated.

Here, we present the first report of isolation of the Dmrt1 homologue from Indian mugger (Crocodylus palustris), a TSD species (Lang and Andrews, 1994), along with its partial genomic locus and show that it exhibits transcriptional diversity as seen in the GSD species. We show that cpDmrt1: 1) expression occurs since the beginning of the TSP in the bipotential GAM; 2) undergoes multiple alternate splicing leading to transcriptional diversity during gonadogenesis, with a few isoforms having domains seemingly unique to the reptiles; and, 3) transcribed isoforms show significant upregulation in the GAM of the male embryos since the beginning of TSP. The multiple splicing and differential expression patterns of cpDmrt1 isoforms may indicate important role for Dmrt1 in the male sex determination in TSD.

2. Materials and methods

2.1. Collection of eggs/study materials/RNA Isolation

Fertilized eggs of Indian mugger were collected from the Nehru Zoological Park, Hyderabad, India on the day of egg laying, and were incubated in the laboratory at 30.0 and 32.5 °C, female and male-supporting temperatures (FPT/MPT), respectively. Embryos were handled as per the guidelines approved by the Institutional Ethics Committee. Dissections were done on days corresponding to developmental stages 21 to 25, which define the TSP (Lang and Andrews, 1994). Embryonic GAM, heart, liver, kidney and brain, were excised, snap frozen in liquid nitrogen and stored at −70 °C. The stored embryonic tissues were used for total RNA isolation using the Trizol protocol.

![Fig. 1](image-url) (A) Schematic representation of the approach used for identification and isolation of cpDmrt1 (isoforms and locus). In the 1st step, partial cpDmrt1 was identified by RT-PCR with conserved primers, which was followed by 5′-3′ RACE to obtain different transcribed isoforms, and finally by genome walking to ascertain/isolate the possible intron(s). (B and C) End-specific RACE analysis suggesting the presence of multiple isoforms of cpDmrt1 expressed in GAM tissue of developing crocodile embryos. Note a single band in 5′ RACE (B), and multiple amplicons in 3′ RACE (C). Lanes: M—100 bp DNA ladder; GAM—Genital ridge-Adrenal-Mesonephros complex.
Table 1

Primers used in the study to isolate/characterize the cpDmrt1 locus and cpdmrt1 transcribed isoforms

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Primer name</th>
<th>Sequence (5'–3'end)</th>
<th>Source</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dmrt1F1</td>
<td>GAAGAAGGATGCAGCAGCGTCT</td>
<td>Designed from conserved domain of Dmrt1 sequences of human, mouse, Gallus, Trachemys, Pelodiscus (accession numbers NM_021951, NM_015826, AF211349, AY316537, AB179697) primers designed from Trachemys coordinates 15–535 bp</td>
<td>To identify/obtain the initial fragment of the crocodile homologue of Dmrt1</td>
</tr>
<tr>
<td>2</td>
<td>Dmrt1R1</td>
<td>CGACAGGAGGCTGAGCAGGCTGACTAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DMS5</td>
<td>GTTCTGAATACGACTCACTATAGGGCTGTCTCTG</td>
<td>New, present study; designed from the initial 500+ bp amplified/ transcribed cpdmrt1 fragment</td>
<td>Primer designed for the 5' RACE</td>
</tr>
<tr>
<td>4</td>
<td>DMS5N</td>
<td>GCTTCTGAGTACTCTTGCTCTGAGC</td>
<td>New, present study; primers designed from common region present in all isoforms</td>
<td>Nested primer to the DMS5</td>
</tr>
<tr>
<td>5</td>
<td>DMS3</td>
<td>GCCAGACAGGCGCCCTGTAGT</td>
<td>New, present study; to measure overall cpDmrt1 levels by real-time RT-PCR</td>
<td>Nested primer to the DMS3</td>
</tr>
<tr>
<td>6</td>
<td>DMS3A</td>
<td>NAGGAAAGCAGTTGAAGACCTGCGCT</td>
<td>New, present study; primers designed from the conserved DM-Ser region in all isoforms</td>
<td>To amplify a 322 bp probe sequence for Northern analysis of cpDmrt1/isoforms</td>
</tr>
<tr>
<td>7</td>
<td>DMM1</td>
<td>GTTAGATAGTACCGAGGCTGAGCAGGCTG</td>
<td>Isoform specific, present study</td>
<td>Primer for genome walking from exon-2</td>
</tr>
<tr>
<td>8</td>
<td>DMG6</td>
<td>CAGGAGGAGGCTGAGCAGGCTGACTAG</td>
<td>Primers designed from the conserved DM-GAM during TSP. All isoforms share the DM and isoforms other than a1/b1 have anterior alternative polyadenylation signals and shorter 3′-UTRs.</td>
<td>Forward primer for real-time PCR for cpDmrt1a/b1 designed from the 3′-UTR region</td>
</tr>
<tr>
<td>9</td>
<td>DMG6A</td>
<td>TCAAACTGAGGCTGAGCAGGCTGACTAG</td>
<td>Reverse primer for real-time PCR for cpdmrt1a/b1</td>
<td>Reverse primer for real-time PCR for cpdmrt1d</td>
</tr>
<tr>
<td>10</td>
<td>DMG6B</td>
<td>TCTAAGCTGAGGCTGAGCAGGCTGACTAG</td>
<td>Forward primer for real-time PCR for cpdmrt1c</td>
<td>Reverse primer for real-time PCR for cpDmrt1a/b1 designed from the 3′-UTR region</td>
</tr>
<tr>
<td>11</td>
<td>DMG6C</td>
<td>CTGTGAGGAGGCTGAGCAGGCTGACTAG</td>
<td>Forward primer for real-time PCR for cpdmrt1d</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
<tr>
<td>12</td>
<td>DMG6D</td>
<td>CTGTCAGGAGGCTGAGCAGGCTGACTAG</td>
<td>Reverse primer for real-time PCR for cpdmrt1f</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
<tr>
<td>13</td>
<td>DMG6E</td>
<td>ACTAGCTGAGGCTGAGCAGGCTGACTAG</td>
<td>Forward primer for real-time PCR for cpdmrt1g</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
<tr>
<td>14</td>
<td>DMG6F</td>
<td>ATAGCTGAGGCTGAGCAGGCTGACTAG</td>
<td>Reverse primer for real-time PCR for cpdmrt1h</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
<tr>
<td>15</td>
<td>DMG6G</td>
<td>GCTGAGGAGGCTGAGCAGGCTGACTAG</td>
<td>Forward primer for real-time PCR for cpdmrt1i</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
<tr>
<td>16</td>
<td>DMG6H</td>
<td>CTGTCAGGAGGCTGAGCAGGCTGACTAG</td>
<td>Reverse primer for real-time PCR for cpdmrt1j</td>
<td>Reverse primer for real-time PCR for isoform cpDmrt1e</td>
</tr>
</tbody>
</table>

2.2. Isolation and expression analysis of cpDmrt1 isoforms, and putative locus

The strategy used to isolate the cpDmrt1 is depicted in Fig. 1A. Briefly, a part of the crocodile homologue of Dmrt1 (cpDmrt1) was obtained initially by RT-PCR using primers designed for the conserved DM domain of published vertebrate sequences. For the purpose, Dmrt1 sequences of human, mouse, Gallus, Trachemys and Pelodiscus were retrieved from the NCBI database (accession numbers NM_021951, NM_015826, AF211349, AY316537, AB179697) and compared to design primers from the conserved region that spanned 15–535 bp region of the Dmrt1 locus and the alternative 3′-UTRs of cpDmrt1. Subsequently, the cpDmrt1 sequence obtained was used to design primers for RACE method (Invitrogen Corporation, USA) as per the manufacturer's instructions. Isolated RNA was dissolved in 0.1×TE buffer.

Fig. 2. (A) Graphical representation showing the relative pair-wise homology between the eight cpDmrt1 isoforms and the initial fragment of 537 bp revealed by RT-PCR. The tentative positions of the primers used for initial RACE reactions are also indicated by solid arrows. (B) Sequence alignment of predicted proteins of different cpDmrt1 isoforms. Short vertical bars indicate the point of change in reading frames due to alternate splicing. (C) Schematic representation of the cpDmrt1 genomic locus and the alternative splicing leading to eight multiple isoforms (GenBank accession numbers of the new sequences are: EU531727, EU526597–EU526604); White boxes represent the unique regions in the predicted genomic locus and the alternative 3′-UTRs. (Figure is not drawn to scale; for size details see Section 3.1.3. The solid black arrows indicate the primers and their locations which were used in genome walking PCRs to identify the intronic sequences; arrows 1 to 8 corresponds to the primers 17 to 24 detailed in Table 1).
reactions. The RACE products were cloned and sequenced to identify transcribed isoforms that were later analyzed for their expression using RT-PCR, quantitative real-time PCR and Northern analysis. Finally, the cpDmrt1 locus was isolated using the genome walking approach. All the primers used in the study are listed in the Table 1.

Further, all the RT-PCR and RACE reactions (for the isolation of cpDmrt1) were carried out using cDNA template prepared from the pooled male and female GAM tissues from the 21st to 25th developmental stages; in comparison, expression analysis was done using individual stage-/tissue-specific cDNA templates.
2.2.1. RT-PCR
The cDNA template for RT-PCR was prepared by reverse-transcription of total RNA (3 μg) from target tissue using the Superscript First-Strand cDNA synthesis system (Invitrogen Corporation, USA) following manufacturer's instructions. All RT-PCR amplifications were carried out using 1 μl cDNA as template in 20 μl reactions using the cycling conditions: 94 °C for 5 min followed by 33 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 90 s, and final extension at 72 °C for 3 min.

2.2.2. 5′/-3′-RACE analysis
5′/-3′-RACE ready cDNAs were prepared using 3 μg total RNA and the SMART-RACE cDNA amplification kit (Clontech Laboratories Inc., USA) as per manufacturer's recommendations. 5′-RACE was done in 2-rounds of PCR: primary PCR using DM5/-UPM and secondary PCR using nested primers DM5'/N/UP in each PCR reaction as per the manufacturer's instructions. The primary PCR was set using 1 μl of 10-times diluted RACE ready cDNA. Cycling conditions were: 95 °C for 2 min followed by 24 cycles of 95 °C for 15 s, 68 °C for 30 s and 72 °C for 2 min, and final extension at 72 °C for 5 min. Secondary PCR was performed using 1 μl of 30-times diluted primary-RACE-PCR product as template using reactions and cycling conditions similar to the one used for primary PCR but for only 14 amplification cycles. Similarly, 3′-RACE was done in two rounds, but using DM3′ and DM3′/N primers instead of DM5'/DM5'. In each case, RACE products were cloned in the pCRII vector using TA cloning kit (Invitrogen Corporation, USA) and then ∼ 300 recombinant clones were sequenced using the universal or gene-specific primers.

2.2.3. Northern blotting
Approximately 10 μg of total RNA was electrophoresed in MOPS-formaldehyde denaturing gel and transferred onto Hybond-N+ membrane (GE Healthcare, USA). A 322 bp region of cpDmrt1 (present in all isoforms) was radioabeled with 32PdATP by PCR using primers GDMF/GDMR. The membranes were hybridized with labeled probe overnight at 62 °C, followed by washings till final stringency of 1X SSC. Northern blots were read using Phosphorimager FLA 3000 (Fujifilm, Japan).

2.2.4. Real-time quantitative PCR
Real-time RT-PCR was carried out to quantify the expression levels of different cpDmrt1 isoforms using GAM-cDNA with primers designed from their unique regions. Every reaction was done in triplicate and repeated twice in 10 μl volume containing 1×SYBR GREEN Mix using 7900HT Fast Real-Time PCR System (Applied Biosystems, USA). The cycling conditions used were: 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s, 62 °C for 15 s, 65 °C for 30 s, and finally by dissociation curve analysis. cpDmrt1a1 and cpDmrt1b1 isoforms were amplified using the primers Dmrt1a1b1F and Dmrt1a1b1R; isoform cpDmrt1d was amplified using primers Dmrt1dR and Dmrt1.4; similarly, isoform cpDmrt1e was amplified using primers Dmrt1eR and Dmrt1.5. GAPDH was used as internal control. ABI software was used for the real-time data analysis.

2.2.5. Genome walking
Four genomewalker libraries were prepared from the crocodile genomic DNA (digested individually with DraI, EcoRV, Stul and PvuII) using the Universal genomewalker kit (Clontech Laboratories Inc., USA), following manufacturer's instructions. Genome walking was performed using two rounds of PCR. In the first round a locus-specific primer was used with API (adapter specific primer) and in the second round, PCR was done using nested primers (nested locus-specific primer and AP2) and 1 μl of 50-times diluted primary PCR product. Amplification conditions for the primary and secondary PCR were similar to those used for RACE analysis. The intron-1 region was amplified using primers DMGW1/DMGW2 (from exon-1 end) and DMGW3/DMGW4 (from exon-2 end), while intron-2 was amplified using primers DMGW5/DMGW6 from the exon-3 end and DMGW7/DMGW8 from the exon-2 end.

2.2.6. In-silico analysis
All primers in the present study were designed using Primer3 software ([http://frodo.wi.mit.edu](http://frodo.wi.mit.edu)). DNA sequences of different cpDmrt1 isoforms and its putative genomic locus were annotated using software ‘Codon Code Aligner’ ([www.codoncode.com](http://www.codoncode.com)). The analytical tool ‘Translate’ ([http://expasy.org/tools/dna.html](http://expasy.org/tools/dna.html)) was used to ascertain the coding regions/predicted proteins of the cpDmrt1 isoforms.

2.2.7. Phylogenetic analysis
Dmrt1 sequences obtained in the study were used to identify and retrieve related Dmrt1 sequences of other vertebrates from GenBank. The sequences were aligned using ClustalW software ([http://www.clustal.org/](http://www.clustal.org/)). The aligned sequences were used to generate neighbor-joining tree to infer their possible phylogenetic relationships using MEGA4.1 phylogenetic package ([http://www.megasoftware.net/](http://www.megasoftware.net/)).

3. Results
3.1. Isolation of cpDmrt1 (C. palustris homologue of Dmrt1)
3.1.1. Initial identification of partial cpDmrt-1
RT-PCR using primers Dmrt1F1/Dmrt1R1 ([Table 1](#table1)) and cDNA (prepared from GAM tissues pooled over developmental stages 21–25) amplified a 517 bp fragment (data not shown). The fragment was cloned and sequenced; it showed complete similarity to the Dmrt1 of vertebrates. This sequence was used to design gene-specific primers (DM5′/DM5′/N; [Table 1](#table1)) for RACE to pull out the complete cpDmrt1.

3.1.2. Multiple splice isoforms of cpDmrt-1
5′–RACE with male and female RACE ready cDNA resulted in a single product ([Fig. 1B](#fig1B)), while 3′–RACE gave several products ([Fig. 1C]) suggesting the presence of multiple isoforms of cpDmrt1. Sequencing analysis of ∼ 600 recombinant plasmids obtained from cloning of 5′/ 3′-RACE products revealed eight isoforms of cpDmrt1 that are named as cpDmrt1a1 (2060 bp, EU526604), cpDmrt1a2 (1119 bp, EU526603), cpDmrt1b1 (1793 bp, EU526602), cpDmrt1b2 (828 bp, EU526600), cpDmrt1c (945 bp, EU526599), cpDmrt1d (910 bp, EU526598), cpDmrt1e (1163 bp, EU526597) and cpDmrt1f (482 bp, EU526601). Sequence comparison of the isoforms suggested these to be splice variants ([Fig. 2A](#fig2A)).

In-silico analysis of the eight transcripts showed that while each of these has an identical short 5′-UTR, their coding regions and 3′-UTR varied significantly in size, ranging from 435 to 984 bp, and 23 to 1097 bp, respectively ([Table 2](#table2)). Further, the analysis predicted these isoforms to encode six proteins of 145–328 amino acids in size, all having the characteristic DM and serine-rich domains but unique C-termini owing to heterogeneity caused by splicing and/or shift in reading frame ([Fig. 2B](#fig2B)). The 313 aa long predicted protein coded by cpDmrt1a1 and cpDmrt1a2 isoforms shows the highest resemblance (up to 95%) to other vertebrate Dmrt1 counterparts with completely conserved DM domain, its flanking region and ser rich domains. Isoforms cpDmrt1b1/b2 are predicted to encode a shorter 155 aa protein of which first 113 aa are shared with the cpDmrt1a1/a2 protein and the remaining 43 aa are not conserved due to a shift in their reading frame caused by splicing. Similarly, the predicted proteins of cpDmrt1c (254 aa) and cpDmrt1d (288 aa) share 208 and 262 aa from their respective N-terminal with the protein coded by cpDmrt1a1/a2 isoforms, and have unique C-terminal ends of 46 and 26 aa, respectively, which do not match with any Dmrt1 protein sequences in public databases. Putative protein translation of cpDmrt1f though shortest in length (145 aa) still it retains the
3.13. Gene structure of cpDmrt1 and alternative splicing

To understand the genomic organization and origin of multiple isoforms, the cpDmrt1 locus was isolated by genome walking approach. The isolated cpDmrt1 locus (EU531727) comprises three exons (exon-1, exon-2, and exon-3 of 650, 208 and 1344 bp length, respectively), and two introns (1311 bp long intron-1, and 252 bp of partially sequenced intron-2). *In-silico* sequence comparisons unequivocally demonstrate that all the eight cpDmrt1 isoforms are generated from the same genomic locus by alternative splicing involving 5′-splice sites, exon skipping, exonization of intronic sequences and use of alternative polyadenylation sites (Fig. 2C). All the isoforms retain the part of exon-1 that codes for the conserved DM domain. It is noteworthy that contrary to the Dmrt1 gene in many vertebrates, the cpDmrt1 lacks any intron at the DM domain coding region. Isoforms cpDmrt1a2 and cpDmrt1b2 differ from isoforms cpDmrt1a1 and cpDmrt1b1 only in their 3′-UTR (shorter by ~1.0 kb) respectively, due to anterior alternative polyA signals. The longer 3′-UTRs of isoforms cpDmrt1a1 and cpDmrt1b1 are unique and do not match to other vertebrate Dmrt1 sequences. Isoforms cpDmrt1b1 and cpDmrt1b2 lack a part of exon-1 (284 bp) due to the use of an alternative 5′-splice site in exon-1, whereas isoform cpDmrt1e has additional 45 bp due to a 5′-alternative splice site in exon-2. Isoform cpDmrt1c lacks exon-2 (163 bp) and isoform cpDmrt1d has a completely different 3′-UTR due to an alternative 3′-splice site in exon-3. The shortest isoform cpDmrt1f was the most unique, comprising only part of exon-1 and exonization of a short sequence (34 nucleotides) of intron-1.

3.1.4. Homology analysis of cpDmrt1 isoforms

BLAST analysis revealed that at the nucleotide level, although all the isoforms were similar to the vertebrate Dmrt1 sequences, it was cpDmrt1a2 that was the most conserved isoform showing a homology of 82, 85 and ~75% with Dmrt1 sequences of Trachemys, Gallus and mammals, respectively, with completely conserved coding region and to some extent the 3′-UTR.

In comparison, all the other isoforms had some unique regions (comprising specific short intronic regions and/or different lengths of UTRs) resulting from the alternative splicing, which did not match to other vertebrate Dmrt1 sequences. Among these, the most notable regions were: the unique long 3′-UTR regions of cpDmrt1a1/cpDmrt1b1, 100 bp 3′ region of cpDmrt1d, and additional 45 bp exonic region of cpDmrt1e.

Phylogenetic analysis using cpDmrt1a2 with Dmrt1 sequences of other vertebrate species further confirms it to be a Dmrt1 sequence most closely related to that of turtles followed by Gallus (Fig. 3). Moreover, the analysis also provides support to the archosaurial lineage of evolution.

3.2. Expression analysis of cpDmrt1 and its isoforms

3.2.1. Tissue-specific expression of cpDmrt1

RT-PCR based expression analysis of cpDmrt1 in different embryonic tissues revealed that it is mainly expressed in the GAM and to a much lower level in the brain and kidney, but not in the heart and liver (Fig. 4A).

3.2.2. Upregulation of cpDmrt1 in male GAM

Semi-quantitative RT-PCR analysis for cpDmrt1 expression in embryonic tissues revealed significantly high expression in the GAM...
of male embryos (kept at MPT) in comparison to that of female embryos (kept at FPT) throughout TSP (Fig. 4B). Similar results were obtained by real-time PCR that showed >21-fold upregulation at 21st stage and 16–17 fold upregulation during stages 22 to 25 at MPT compared to FPT (Supplementary Fig. 1). However, no such sex-specific upregulation was seen in the brain and kidney during TSP (data not shown). Northern analysis further confirmed higher expression of *cpDmrt1* at the MPT, and also indicated presence of several isoforms in the GAM tissue, and maximum expression for an isoform of ~1 kb size (Fig. 4C). Similarly, the other observed bands matched with the isoforms having abundant expression.

### 3.2.3. Expression pattern of isoforms

Expression of isoforms *cpDmrt1a/b1* (combined), *cpDmrt1d* and *cpDmrt1e* at MPT and FPT during TSP was further analyzed using semi-quantitative, as well as, real-time RT-PCR (Fig. 5, Supplementary Fig. 1). The results revealed that *cpDmrt1e* was the most abundant isoform followed by *cpDmrt1a/b1* (combined expression) contributing ~70% and 26%, respectively to the overall *cpDmrt1* levels at MPT. Further, *cpDmrt1e* expression peaked at 21st developmental stage, while *cpDmrt1a/b1* showed highest upregulation at 22nd stage. In contrast, the expression of *cpDmrt1d* was <1% of the total *cpDmrt1* levels at both MPT and FPT, and its highest expression was at the 21st developmental stage. Overall, the analysis showed that all the isoforms were upregulated at MPT compared to FPT through TSP, but to different extents and at different time points.

## 4. Discussion

*Dmrt1* is a conserved gene suggested to be important in the sex determination and early gonad differentiation in vertebrates. Across the evolutionary strata, it is known to play varying role i.e. from being a master regulator of male sex in teleost Medaka (*Matsuda et al., 2002*) to an important sex regulator on the Z-chromosome of chicken (*Shan et al., 2000*) and to a gene involved in embryonic testis differentiation in mammals (*Raymond et al., 2000*). Recent studies have shown significant transcriptional diversity for this important gene during gonadal differentiation in many GSD species (*Huang et al., 2005; Zhao et al., 2007; Guo et al., 2005; Cheng et al., 2006; Lu et al., 2007*). Preliminary studies done in few, turtle species show upregulation of *Dmrt1* during male gonadogenesis, suggesting its possible role in male sex-development even in TSD species (*Kettlewell et al., 2000; Torres Maldonado et al., 2002; Shoemaker et al., 2007*). However in general, these studies lack organizational and functional details of the gene, which are important to ascertain its functional role and evolutionary significance in TSD.

In the present study, we have isolated *cpDmrt1*, the crocodile homologue of the *Dmrt1* gene, and also its eight novel isoforms generated by alternative splicing. We further show that these isoforms are significantly upregulated in the male GAM through TSP. This study provides the first evidence for the transcriptional diversity of *Dmrt1* in a TSD species as has earlier been seen in GSD vertebrates. Thus, these results may suggest *cpDmrt1* to be an important candidate early gene for male development even in TSD. This premise gets further support from the observation that some of the novel *cpDmrt1* isoforms are expressed at very high levels in GAM of male embryos from the beginning of TSP (21st developmental stage) when the gonadal primordia are bipotential in Indian mugger and other TSD species (*Wibbels et al., 1991; Lang and Andrews, 1994*).

Transcriptional diversity is expected to be for multiple functional roles of a gene. Recent reports suggest that *Dmrt1* may have a role in cell fate determination, postnatal differentiation of Sertoli cells and also in primordial germ cells maintenance and their meiotic progression (*Kim et al., 2007*). Such multiple roles for *Dmrt1* have also been suggested in the case of a lizard *Calotes versicolor* (*Sreenivasulu et al., 2002*) that exhibits hormonal/temperature dependent sex determination. Transcribed isoforms may contribute to the multiple roles of *Dmrt1* either by binding to different upstream and downstream targets or by differential regulation of specific isoforms. It is now well documented that DM domain genes in GSD vertebrate species (human, mouse, chicken, rice-eel and zebrafish) undergo splicing like their invertebrate counterparts (dsx and mab3) wherein male/female specific isoforms control different aspects of male and female development (*Hodgkin, 2002*). Splicing in above species is noted mainly in the 3′ region of the *Dmrt1*, and in many cases leads to change in the reading frame of the predicted proteins (*Lu et al., 2007*). Similar kind of splicing/change in reading frame has also been seen in the present study for isoforms...
**References**


