

REGULATORY ELEMENTS IN THE UPSTREAM REGION OF METALLOTHIONEIN GENE IN TILAPIA SPECIES

MOHAMMAD SHAMIMUL ALAM*, ISRAT JAHAN¹, SADNIMAN RAHMAN,
HAWA JAHAN AND KANIZ FATEMA¹

*Genetics and Molecular Biology Laboratory, Department of Zoology,
University of Dhaka, Dhaka-1000, Bangladesh*

Key words: Metallothionein gene, Heavy metal pollution, Transcription factor binding sites, Tilapia, *Oreochromis*

Abstract

Tilapia is a hardy fish which can survive in water bodies polluted with heavy metals. Metal resistance is conferred by higher expression of *metallothionein* gene (*mt*) in many organisms. Level, time and tissue-specificity of gene expression is regulated through transcription factor binding sites (TFBS) which may be present in the upstream, downstream, or even in the introns of a gene. So, as a candidate regulatory region, the 5'upstream sequence of *mt* gene in three tilapia species, *Oreochromis aureus*, *O. niloticus* and *O. mossambicus* was studied. The targeted region was PCR-amplified and then sequenced using a pair of custom-designed primer. A total of only 2.7% variation was found in the sequenced genomic region among the three species. Metal-related TFBS were predicted from these sequences. A total of twenty eight TFBS were found in *O. aureus* and twenty nine in *O. mossambicus* and *O. niloticus*. The number of metal-related TFBS predicted in the targeted sequence was significantly higher compared to that found in randomly selected other genomic regions of same size from *O. niloticus* genome. Thus, the results suggest the presence of putative regulatory elements in the targeted upstream region which might have important role in the regulation of *mt* gene function.

Introduction

Tilapia is a hardy fish widely cultured in Asia. It has acquired second position (after carps) in terms of fish production worldwide⁽¹⁾. Tilapia culture has been expanded rapidly amongst the fish farmers because of its adaptability and tolerance. Tilapia species can adapt easily to salinity and oxygen level fluctuations. They consume food from different trophic levels and cope overcrowding⁽²⁾. Thus, they can survive in aquatic bodies with higher level of pollution. The water of Buriganga River has been polluted by different sources such as tannery and textile effluents, sewerage system and garbage⁽³⁾. Heavy metals, for instance, Cd, Hg, Cr, Pb and As are found in high concentrations in *Channa punctata*, *Heteropneustes fossilis* and *Oreochromis* spp. of Buriganga River⁽⁴⁾.

*Author for correspondence: <shamimul@du.ac.bd>. ¹Department of Fisheries, University of Dhaka, Dhaka-1000, Bangladesh.

Several genes are related to regulation of different metals in the body. Metallothionein (*mt*), first discovered in 1957, has important role as Cd binding protein⁽⁵⁾. The main function of this gene is detoxification of heavy metals, homeostasis of indispensable heavy metal (Zinc, Copper, etc.), protection against apoptosis, oxidative stress etc. It can help the organisms to survive in the heavily polluted area by detoxifying heavy metals⁽⁵⁾. Synthesis of *mt* gene is induced by many metals by a regulated process involving increased gene transcription⁽⁶⁾. Many physiological and nutritional factors including starvation and imposition of various types of physical or inflammatory stress can influence the expression of this gene via TFBS⁽⁶⁾. It is not known to us, how tilapia fishes of highly polluted water bodies like Buriganga survive and tolerate high level of metal concentrations. Thus, it is very important to understand the regulatory mechanism of *mt* gene. To know that, identifying the putative regulatory region along with occurrence of TFBS is very much essential. Therefore, in the present study, metal related TFBS from the upstream putative regulatory region of *mt* gene of three selected tilapia species from Buriganga, Bangladesh was predicted, analyzed and compared to find its uniqueness in metal regulation.

Materials and Methods

Sample collection and Identification: Three tilapia fish samples were collected from fisherman who captured the fishes from Showarighat point of Buriganga River, Dhaka, Bangladesh (23°42'41.2"N 90°23'40.5"E). Samples were labeled carefully after collection. Fish samples were identified morphologically following the published protocols⁽⁷⁾. Samples were identified as *Oreochromis aureus*, *O. niloticus* and *O. mossambicus* (Table 1).

DNA extraction and PCR amplification: DNA samples from three fish species were extracted by using CTAB method⁽⁸⁾. Targeted upstream region of *mt* gene was amplified by Polymerase Chain Reaction (PCR) using extracted DNA as template and visualized through electrophoresis in a 1.5% agarose gel⁽⁸⁾. Temperature regimes for PCR were 95°C for 1 minute in denaturation, 56°C for 30s in annealing, and 72°C for 30s in extension for 30 cycles with a final polymerization step at 72°C for five minutes. Primer pair used in this study were designed beforehand based on the available sequence (GenBank: AY257202.1) of the targeted genomic region. Forward and reverse primers designed for this study were 5' - GACTGAGGAGTCACACGAT -3' and 5' -TGGCTGGTTTGTTCAGAGTG-3', respectively. Then, PCR product was purified using GeneJet PCR Purification Kit (Thermo Fisher Scientific, USA) and sequenced using Sanger sequencing method (Macrogen, Korea).

Bioinformatics analysis: Serial Cloner 2.6 (http://serialbasics.free.fr/Serial_Cloner.html) and MEGA X⁽⁹⁾ were used to edit and analyze sequence data. Final nucleotide base pair length for all three species was 667 bp. To find out the nucleotide variation in the upstream regions of *mt* gene in these three fish species, multiple sequence alignment was

performed by using MEGA X⁽⁹⁾. Nucleotide sequence of these three fishes was locally aligned with the *mt* gene reference sequence of *O. mossambicus* from Hong Kong, China (GenBank: AY257202.1) by using MUSCLE program⁽¹⁰⁾. All the parameters were kept default.

To predict the transcription factor binding sites (TFBS) for the upstream region of *mt* gene of three tilapia species, three databases i.e. JASPAR, PATCH program of Biobase and ALGEN-PROMO were used⁽¹¹⁻¹³⁾. The resource data from these databases were obtained as of January 09, 2020. At first, predicted TFBS from all three databases were collected individually. TFBS related to metal regulation and common in all three databases were then selected for further evaluation. A threshold score of 7.0 was set. Finally, TFBS only with the threshold score ≥ 7.0 were counted. This process was repeated for all three tilapia species. Additionally, to compare the TFBS predicted in the *mt* gene regulatory region with other regions, five random portions of 667bp were chosen from the whole genome sequence of *O. niloticus* (GenBank: NC_031978.2) as control. Similarly, TFBS from these five regions were also predicted. The TFBS predicted from control regions were compared with those predicted from targeted upstream region of *mt* gene.

Results and Discussion

Taxonomic Identification: Three species were identified as *O. aureus* (Blue tilapia), *O. niloticus* (Nile tilapia) and *O. mossambicus* (Mozambique tilapia) based on morphological characteristics (Table 1).




Variation in upstream region of mt gene: Upstream nucleotide sequences of *mt* gene from three studied species were aligned with the sequence of *mt* gene of *O. mossambicus* (GenBank: AY257202.1) from Hong Kong, China. Out of 667, only 18 nucleotides showed variation among the three species (Table 2). Overall, only 2.7% polymorphism was found in the upstream putative regulatory region of *mt* gene. Percentage of variation between the species was also observed (Table 3). The targeted region of *O. mossambicus* showed 100% sequence similarity with that reported from *O. mossambicus* of Hong Kong, China.

Prediction of transcription factors binding sites: Regulation of gene expression at transcription level is an intricate process especially in multi-cellular organisms. Detection of transcription factor binding sites (TFBS) for any gene is needed for better understanding of its regulation mechanism⁽¹⁴⁾.

A total of twenty eight TFBS were found in *O. aureus*, and twenty nine in both *O. mossambicus* and *O. niloticus* (Table 4). Three GATA elements (GATA1, GATA5 and GATA6) were predicted from the sequenced upstream region of *mt* gene. Presence of GATA elements in the upstream regulatory region of *mt* gene in *Caenorhabditis elegans* was reported previously⁽¹⁵⁾. Besides, GATA family was found playing important role in

hematopoiesis and cardiogenesis in teleost fish, *Danio rerio*⁽¹⁶⁾. Findings show that GATA family plays important role throughout the life⁽¹⁷⁾.

Table 1. Morphological characteristics of three tilapia species.

| | <i>O. aureus</i> | <i>O. niloticus</i> | <i>O. mossambicus</i> |
|-------------------------------|--|--|---|
| Body coloration | bluish brown body; greenish upper lip; yellowish orange fins; no stripes in the caudal fin | grayish white body; yellowish orange fins; seven stripes in the body; also stripes in the caudal fin | dark black body; reddish black fins; mozambique type spots in every scales and operculum except caudal fin and body |
| Number of dorsal fin rays | 28 (17 spiny) | 28 (16 spiny) | 28 (15 spiny) |
| Number of pectoral fin rays | 12 | 13 | 12 |
| Number of pelvic fin rays | 6 | 6 | 6 |
| Number of anal fin rays | 12 (3 spiny) | 12 (3 spiny) | 13 (3 spiny) |
| Number of caudal fin rays | 17 | 18 | 18 |
| Number of lateral line scales | 31 | 32 | 35 |
| Images |  |  |  |

Apart from GATA family, Zinc Finger containing proteins (ZNFs) also play important role in different biological processes including development and metabolism of fishes⁽¹⁸⁾. Previous study showed that Zinc-finger family of transcription factors activated *mt* gene in response to metals and oxidative stress in mice⁽¹⁹⁾. In the present study, ZNF24, ZNF148, ZNF263, ZNF341 and ZNF384 were found in all three tilapia species, whereas ZNF317 was found in *O. aureus* and *O. mossambicus* in the nucleotide position of 628-639bp and 618-629bp respectively but not found in *O. niloticus*. On the other hand, Krüppel-like factor (KLF) family plays versatile role in development, differentiation, metabolism, autophagy, cell growth, proliferation, and differentiation⁽²⁰⁾. KLF10, KLF16 and KLF 17 were predicted from all three studied species. The SP family of transcription factors SP1, SP3, SP8 and SP9 were predicted from all three species in this study. They are the DNA-binding proteins that share a highly conserved, carboxyl-

terminal DNA binding domain featuring three zinc fingers. This family was found to regulate embryonic development in zebra fish during early stage⁽²¹⁾.

Table 2. Nucleotide variations in putative *mt* gene regulatory regions among three species.

| Nucleotide position | <i>O. aureus</i> | <i>O. mossambicus</i> | <i>O. niloticus</i> | Overall variation |
|---------------------|------------------|-----------------------|---------------------|-------------------|
| 42 | -- | A | A | 2.7% |
| 108 | G | A | G | |
| 178 | A | G | G | |
| 190 | T | -- | T | |
| 310 | A | G | G | |
| 331 | C | -- | C | |
| 332 | A | -- | A | |
| 333 | C | -- | C | |
| 334 | T | -- | T | |
| 335 | T | -- | T | |
| 361 | C | G | G | |
| 385 | G | T | G | |
| 386 | A | T | T | |
| 475 | G | A | G | |
| 493 | G | C | G | |
| 500 | A | G | G | |
| 639 | T | G | G | |
| 659 | T | A | A | |

Table 3. Nucleotide variation among the upstream sequences of *mt* gene in three tilapia species.

| Between species | Variation (%) |
|---|---------------|
| <i>O. aureus</i> and <i>O. mossambicus</i> | 2.7 |
| <i>O. mossambicus</i> and <i>O. niloticus</i> | 1.65 |
| <i>O. aureus</i> and <i>O. niloticus</i> | 1.2 |

The nuclear receptors (NRs) regulate inflammatory responses, cellular proliferation, and differentiation in organisms. Among five modules, the C module contains the DNA-binding domain (DBD) that encloses the zinc-fingers necessary for binding DNA at response elements in *Daphnia*⁽²²⁾. In our study, NR2C1 was predicted from *O. mossambicus* and *O. niloticus* in the nucleotide position of 373-381bp and 380-388bp, respectively (Table 4) but its occurrence was not found in *O. aureus*. NR3C1 was predicted to be occurred twice in the nucleotide position of 464-481bp and 521-535bp in *O. niloticus* (Table 4). On the other hand, Hepatocyte Nuclear Factors, HNF1A and HNF4A were predicted from all three species. This family plays important role in the expression of genes involved in pancreatic islet development and metabolism⁽²³⁾.

Table 4. Predicted TFBS in the upstream of *mt* gene of *O. aureus*, *O. mossambicus* and *O. niloticus* with their sequences.

| Name | Status in three fish species | | | TFBS Sequence |
|--------|------------------------------|-----------------------|---------------------|--|
| | <i>O. aureus</i> | <i>O. mossambicus</i> | <i>O. niloticus</i> | |
| AR | √ | √ | √ | GGGAACAGAGAGATC |
| GATA1 | √ | √ | √ | TTCTTATATAA ^(OA, OM) GGATGG ^(ON) |
| GATA5 | √ | √ | √ | TGATTACG |
| GATA6 | √ | √ | √ | CGTTCTTATATAA |
| HNF1A | √ | √ | √ | TCGTAATCATTAACC ^(OA) GGTTAATGATTACG ^(OM, ON) |
| HNF4A | √ | √ | √ | TTGCACTCGGCTCTG ^(OA) AGGCCATAGATCT ^(OM, ON) |
| RORA | √ | √ | √ | AAGAACGTCA ^(OA, ON) TATAAGAATGTCAA ^(OM) |
| SP1 | √ | √ | √ | CCCATCCACC |
| SP3 | √ | √ | √ | GCCACTCCTAC |
| SP8 | √ | √ | √ | GCCACTCCTACA |
| SP9 | √ | √ | √ | GCCACTCCTACA |
| NR112 | √ | √ | √ | GTGCACACGACAACCTT |
| NR2C1 | × | √ | √ | CAAGGTCAC ^(OM, ON) |
| NR3C1 | √ | √ | √ | GGAACAGAGAGATCC |
| NR3C1 | × | × | √ | AAGAACGTCAACTTCTGG ^(ON) |
| EGR2 | √ | √ | √ | GCCACTCCTACACCG ^(OA, ON) GCCACTCCTACACCA ^(OM) |
| ZNF24 | √ | √ | √ | TATTTTTTCATTC ^(OA, OM) CATTCATTCAAGT ^(ON) |
| ZNF148 | √ | √ | √ | CGCCCATCCACC |
| ZNF263 | √ | √ | √ | AGAGGATGTTGGGTGGATGGG |
| ZNF317 | √ | √ | × | GAACAACAGAGC |
| ZNF341 | √ | √ | √ | GGGAACAGAGAG ^(OA) TAAAAGAGCCAC ^(OM, ON) |
| ZNF384 | √ | √ | √ | AATGAAAAAATA ^(OA) TTACAAAAAAAC ^(OM, ON) |
| KLF10 | √ | √ | √ | GCCACTCCTAC |
| KLF16 | √ | √ | √ | GCCACTCCTAC |
| KLF17 | √ | √ | √ | GCCCATCCACCCAAC |
| SNAI2 | √ | √ | √ | ACACGCCTGTCAC |
| ZBTB7A | √ | √ | √ | GGCCTGAAGTGAA |
| ZSCAN4 | √ | √ | √ | CGCAAATGCTGAAAA |
| Zic2 | √ | √ | √ | CTCCATCAGGACG ^(OA, ON) CACGCAGCAGGCCA ^(OM) |
| OSR2 | √ | √ | √ | CTCCAGAAGTTG |

Here, √ and × indicates presence and absence of respected TFBS. Exact sequences of the TFBS from each species are indicated by alphabet superscripts (^{OA} = *O. aureus*, ^{OM} = *O. mossambicus*, and ^{ON} = *O. niloticus*). No superscript on a specific TFBS sequence indicates its presence in all positive species.

Some other transcription factors Zic2, ORS2, RORA and AR were also predicted. Zic2 acts as an enhancer-binding factor which is required for embryonic stem cell specification⁽²⁴⁾ whereas, androgen receptors (AR) which locates in the brain, plays important role in the control of reproduction in fish⁽²⁵⁾ and ORS2, member of the odd-skipped family of proteins is required for renal development in vertebrates⁽²⁶⁾. Therefore, from the previous reports, it appears that the predicted transcription factors play important role in development and other biological functions in addition to heavy metal resistance.

Table 5. Predicted TFBS from five random sites from the available whole genome sequence of *O. niloticus* (GenBank: NC_031978.2).

| Predicted Transcription Factor Binding Sites (TFBS) | | | | |
|---|--------------------|----------------------|----------------------|----------------------|
| Site 1 | Site 2 | Site 3 | Site 4 | Site 5 |
| [3978521..3979187] | [9558977..9559643] | [19239007..19239673] | [21178764..21179430] | [31278095..31278761] |
| GATA1 | GATA1 | GATA1 | GATA1 | GATA1 |
| GATA5 | GATA5 | GATA5 | GATA5 | GATA5 |
| ZNF24 | ZNF24 | EGR2 | ZNF24 | NR112 |
| Zic2 | NR112 | | NR2C1 | Zic2 |
| | Zic2 | | | |

Besides, the sequenced upstream region of *mt* gene was compared with randomly selected five other genomic regions in terms of predicted TFBS. To do that, five random portions of 667bp were chosen from the whole genome sequence of *O. niloticus* (GenBank: NC_031978.2) as controls. Only three to five metal related TFBS were predicted from these regions while twenty eight to twenty nine TFBS were predicted in the upstream region in case of studied *mt* gene (Table 5). Thus, this upstream region of *mt* gene might have important role in metal regulation. Since, only the TFBS related to heavy metal resistance were included in the study (see methodology), presence of those might be an indication of putative regulatory elements in the targeted region. Further extensive study is required to understand the regulatory elements in the upstream region of *mt* gene of tilapia.

References

1. Prabu E, CBT Rajagopalsamy, B Ahilan, IJMA Jeevagan and M Renuhadevi 2019. Tilapia - An excellent candidate species for world aquaculture: a review. Annual Research & Review in Biology **31**(3): 1-14.
2. Canonico GC, A Arthington, JK Mccrary and and ML Thieme 2005. The effects of introduced tilapia on native biodiversity. Aquatic Conserv: Mar. Freshw. Ecosyst. **15**: 463-483.
3. Bhuiyan MAH, SB Dampare, MA Islam and S Suzuki 2014. Source apportionment and pollution evaluation of heavy metals in water and sediments of Buriganga River, Bangladesh using multivariate analysis and pollution evaluation indices. Environ. Monit. Assess. **187**: 4075.

4. Ahmed MK, MA Baki, GK Kundu, MS Islam, MM Islam and MM Hossain 2016. Human health risks from heavy metals in fish of Buriganga River, Bangladesh. *SpringerPlus*. **5**: 1697.
5. Andrews GK 2000. Regulation of *metallothionein* gene expression by oxidative stress and metal ions. *Biochem. Pharmacol.* **69**(1): 95-104.
6. Bremner I 1991. Nutritional and physiological significance of metallothionein. *Methods. Enzymol.* **205**: 25-35.
7. Trewavas E 1941. Nyasa fishes of the genus *Tilapia* and a new species from Portuguese, East Africa. *J. Nat. Hist.* **7**: 294-306.
8. Alam MS, KA Ahmed, RA Begum and RM Shahjahan 2016. Identification of *Megaselia scalaris* (Diptera: Phoridae) based on morphology and mitochondrial *16S rRNA* and *COI* gene sequences. *Dhaka Univ. J. Biol. Sci.* **25**(2): 149-159.
9. Kumar S, G Stecher, M Li, C Knyaz and K Tamura 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6): 1547-1549.
10. Edgar RC 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* **5**:113.
11. Fornes O, JA Castro-Mondragon, A Khan, R van der Lee, X Zhang, PA Richmond, BP Modi, S Correard, M Gheorghe, D Baranašić, W Santana-Garcia, G Tan, J Chèneby, B Ballester, F Parcy, A Sandelin, B Lenhard, WW Wasserman and A Mathelier 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic. Acids. Res.* **48**(D1): D87-D92.
12. Messeguer X, R Escudero, D Farré, O Nuñez, J Martínez and MM Albà 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. *J. Bioinform.* **18**(2): 333-334.
13. Farré D, R Roset, M Huerta, JE Adsuara, L Roselló, MM Albà and X Messeguer 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic. Acids. Res.* **31**: 3651-3653.
14. Pierstorff N, CM Bergman and T Wiehe 2006. Identifying *cis* regulatory modules by combining comparative and compositional analysis of DNA. *J. Bioinform.* **22**(23): 2858-2864.
15. Moilanen LH, T Fukushige and JH Freedman 1999. Regulation of metallothionein gene transcription. Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from *Caenorhabditis elegans*. *J. Biol. Chem.* **274**(42): 29655-29665.
16. Heicklen-Klein A, LJ McReynolds and T Evans 2005. Using the zebrafish model to study GATA transcription factors. *Semin. Cell. Dev. Biol.* **16**(1): 95-106
17. Viger RS, SM Guittot, M Anttonen, DB Wilson and M Heikinheimo 2008. Role of the GATA family of transcription factors in endocrine development, function, and disease. *Mol. Endocrinol.* **22**(4): 781-798.
18. Jen J and YC Wang 2016. Zinc finger proteins in cancer progression. *J. Biomed. Sci.* **23**(1): 53.
19. Andrews GK 2000. Regulation of metallothionein gene expression by oxidative stress and metal ions. *Biochem Pharmacol.* **59**(1): 95-104.
20. Knoedler JR and RJ Denver 2014. Krüppel-like factors are effectors of nuclear receptor signaling. *Gen. Comp. Endocrinol.* **203**: 49-59.

21. Xie J, H Yin, TD Nichols, JA Yoder and JM 2010. Horowitz. Sp2 is a maternally inherited transcription factor required for embryonic development. *J. Biol. Chem.* **285**(6): 4153-4164.
22. Litoff EJ, TE Garriott, GK Ginjupalli, L Butler, C Gay, K Scott and WS Baldwin 2014. Annotation of the *Daphnia magna* nuclear receptors: Comparison to *Daphnia pulex*. *Gene.* **552**(1): 116-125.
23. Shih DQ, S Screenan, KN Munoz, L Philipson, M Pontoglio, M Yaniv, KS Polonsky and M Stoffel 2001. Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes* **50**(11): 2472-2480.
24. Luo Z, X Gao, C Lin, ER Smith, SA Marshall, SK Swanson, L Florens, MP Washburn and A Shilatifard 2015. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol. Cell.* **57**(4): 685-694.
25. Harbott LK, SS Burmeister, RB White, M Vagell and RD Fernald 2007. Androgen receptors in a cichlid fish, *Astatotilapia burtoni*: structure, localization, and expression levels. *J. Comp. Neurol.* **504**(1): 57-73.
26. Neto A, N Mercader and JL Gomez-Skarmeta 2011. The *osr1* and *osr2* genes act in the pronephric anlage downstream of retinoic acid signaling and upstream of *wnt2b* to maintain pectoral fin development. *Development* **139**(2): 301-311.

(Manuscript received on 6 September, 2020; revised on 29 October, 2020)