# DIVERSITY AND EVOLUTION OF *CYP* MITOCHONDRIAL GENE IN NILE TILAPIA (*Oreochromis niloticus* L.)

**M. A. RASHED[1], AMIRA EL-KERADY[2] AND M. MAGDY[1]**

1. *Genetics Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt*
2. *Genetics Department, Faculty of Agriculture, Tanta University, Cairo, Egypt*

The cytochrome P450 (CYP) superfamily of heme-containing enzymes catalyzes Phase I biotransformation of endogenous and xenobiotic compounds, including fatty acids, steroids, drugs, and environmental contaminants. Over 1900 CYP isoforms with overlapping substrates have been identified in animals (P450 Nomenclature Committee, 2006).

Enzymes produced from the *cytochrome P450* genes are involved in the formation (synthesis) and breakdown (metabolism) of various molecules and chemicals within cells. Cytochrome P450 enzymes play a role in the synthesis of many molecules including steroid hormones, certain fats (cholesterol and other fatty acids), and acids used to digest fats (bile acids). Additional cytochrome P450 enzymes metabolize external substances, such as medications that are ingested and internal substances, such as toxins that are formed within cells (Lynch and Price, 2007).

Cytochrome P450 enzymes are primarily found in liver cells but are also located in cells throughout the body. Within cells, cytochrome P450 enzymes are located in a structure involved in protein processing and transport (endoplasmic reticulum) and the energy-producing centers of cells (mitochondria). The enzymes found in mitochondria are generally involved in the synthesis and metabolism of internal substances, while enzymes in the endoplasmic reticulum usually metabolize external substances, primarily medications and environmental pollutants (Guengerich, 2007).

Most of them have been described in mammals and only recently substantial numbers of CYPs have been identified in fish. Because activity of these enzymes is often the determining factor in susceptibility of organisms to toxicity, knowledge of substrate specificity and regulation of CYPs has been useful in predicting toxicity across species. Unfortunately, mammalian models of CYP metabolism have not always successfully transferred to fish (Miranda *et al.*, 1998).

Common variations (polymorphisms) in *cytochrome P450* genes can affect the function of the enzymes. The

*Egypt. J. Genet. Cytol., 46: 297-304, July, 2017*
Web Site (*www.esg.net.eg*)

effects of polymorphisms are most prominently seen in the breakdown of medications. Depending on the gene and the polymorphism, drugs can be metabolized quickly or slowly. If a cytochrome P450 enzyme metabolizes a drug slowly, the drug stays active longer and less is needed to get the desired effect. A drug that is quickly metabolized is broken down sooner and a higher dose might be needed to be effective. Cytochrome P450 enzymes account for 70 percent to 80 percent of enzymes involved in drug metabolism (Hannemann *et al*., 2007).

Each *cytochrome P450* gene is named with CYP, indicating that it is part of the gene family. The gene is also given a number associated with a specific group within the gene family, a letter representing the gene's subfamily, and a number assigned to the specific gene within the subfamily. For example, the *cytochrome P450* gene that is in group 27, subfamily A, gene 1 is written as *CYP27A1*.

In fish, members of the CYP2 and CYP3A families play a major role in the metabolism of xenobiotics and endogenous compounds, including steroids (Miranda *et al*., 1989; Brian *et al*., 1990; Thibaut *et al*., 2002). To obtain a better understanding of P450 regulation and activity in fish, the database of available sequences in fish needs to be expanded. Therefore, the current study aimed to isolate and compare *cytochrome P450 3A40* gene to understand its diversity and evolution among fresh water fish species including those available in the GenBank database.

## MATERIALS AND METHODS

### *Materials*

A total of 15 samples of Nile tilapia (*Oreochromis niloticus*) were captured alive from the Fish Farm of Animal Department, Faculty of Agriculture, Ain shams University.

### *Methods*

#### *RNA extraction and partial CYP gene amplification*

RNA was extracted from fish livers using Wizard® RNA Purification Kit (PROMEGA, USA) by following the manufacturer's manual. Extraction success was examined using agarose gel (2%) electrophoresis standard protocol and visualized using InGenius3 gel documentation system (SYNGENE, UK). RNA concentration was measured using Quantus fluorometer (Promega, USA) and transcript reversely using SensiFAST cDNA Kit (BIO-65054, Bioline, UK) according to the kit manual. Approximately 100 ng of cDNA (bulk from 15 samples) was added to 50 µl total volume of 1x MyTaq Red Mix (BIO- 25043, Bioline, UK), 1 mM of the newly designed forward primer (On-CYPF: 5'-ATGTACTGGGGGACAATAGCCA-3') and 1 mM of oligo d(T)18 as reverse primer. Techne-512 (Techne, UK) was programmed for long PCR amplification at: 95°C 5 min, 32 cycles of 95°C/30 sec,

55°C/30 sec and 72°C/2 min, and a final step of 72°C for 15 min. Product was tested using agrose gel (1.5%) electrophoresis standard protocol and the molecular weight was calibrated using O'Generuler 100 bp DNA ladder (Thermo Scientifics, Lithuania). When successful, amplicons were purified and sequenced for both directions using private services (Macrogen, Inc., South Korea).

Sequences were trimmed, refined and assembled; consensus sequence was generated and used for BLAST search using Geneious (V10) software. Pairwise and multiple alignments of the partial *CYP* gene with the BLAST results was performed using Geneious aligned. Polymorphism and diversity were estimated using DNASP (Librado and Rozas, 2009). Phylogenetic analysis was performed using the implemented tool FASTTREE 2.1.5 adjusted by Generalized time reversible (GTR) model and pseudo-counts. Species delimitation was defined based on the taxonomical data for each species included in the analysis, inter and intra distances to the closest species were measured using SpDelim tool in Geneious, where each family was coloured differently.

## RESULTS AND DISCUSSION

### *Oreochromis niloticus cytochrome P450*

Around 1300 bp were amplified successfully, and was found 100% match with the *Oreochromis niloticus cytochrome P450 3A30* (LOC100700506; GenBank ID: XM_019360123). Codon usage was measured, as the highest was UUU-Phenylalanine (17 times) and the lowest was stop codon (no occurrence), succeeded by CGG- Arginine (unique occurrence) in the sequenced region of the gene. The partial sequence coding 449 amino acids started with ATG for methionine and followed by:

"YWGTIARHNRVHYLDDYECAQKYGRIW
GVYEFRKPMLAVMDPDMLKTILVKECFT
YFTNRRNFRLNGDLYDAVSLAEDDQWRR
IRNILSPSFTSGHLKEMFSIMKHHSRKL
TDSLQSKAHNDEVIIVKDFFGPYSIDVM
ASCIFSVDFDSIKNPSNPFITHANQMFK
FPLLLYIFQACFPIFLPLLERLGVSLFP
KSSTAFLKSVAEKVKAERNSSSQTSDDM
LEHLIKCQTASENDKEKKHKGLTEHEII
SQVTMFVFAGYETSAIALVFLAYSLARN
PEIMKRLQREIDSTFPNKGPVEYEALMQ
MEYLDSVVSECLRLYPSIPRLERVAKET
VKISEITIPKGMLVMVPVYALHRDPELW
PEPEEFKPDRFSKENKQNINPYTYLPFG
AGPRSCMGTRFALMMVKLAM-
VEVLQNFSFSVCKETEIPLEMDVAGFVS
PVRPIKLK". No stop codon was detected, as the sequence was complete at the 5' terminal and partially sequenced in the 3' terminal. *CYP* genes play key roles in many crucial biological processes including oxidative transformation of xenobiotics and metabolism of endogenous substrates. They belong to one of the most widespread and diverse gene families that consist of more than 18,500 members (Nelson, 2013) among various species. Zhang *et al*. (2014) systematically identified and characterized a full set of 61 *CYP* genes in channel catfish alone. While others were isolated from several fishes (e.g. Barber *et al*., 2007).

***Comparative analysis among fresh water fish species***

***Polymorphism and diversity among Oreochromis niloticus sequences***

All sequences length was equal (1346 bp), only four sites (at base pair numbers 41, 43, 54, 74) were found polymorphic, the sequence of the current study was identical to haplotype X3 of the *CYP450* gene (Accession: XM_019360124). Site no. 41 T>A changed the amino acid from Phenylalanine (F) to Tyrosine (Y), while sites no. 43 and 54 caused no effect (silent mutations), and site no. 74 A>G changes the amino acid from Glutamic Acid (E) to Glycine (G). Testing the mutation effect on the secondary structure of the protein that formed a major change in its structure. The effect is due to the site no. 74, as the change in the amino acid has affected the formation of the alpha helix as it disappeared in sequences with G amino acid instead of E, subsequently, it was compensated with turn and complemented with a single turn in the amino acid Histidine (H) on the left side while followed by a beta strand on the right side of the amino acid change (Fig. 1).

***Polymorphism and diversity among species***

Overall alignment length among 23 accessions (sequenced sample + 22 Blast result accessions) was 1393 bp, including coding regions sequence (CDS). Sites with gaps 18 and 577 were monomorphic and 754 were polymorphic (total number of mutations was 1071). Total of 531 of the polymorphic sites were parsimony informative sites (304 two- variant sites, 174 three-variant sites and 53 four-variant sites), while 223 were singleton variable sites (189 two-variant sites, 31 three-variant sites and 3 four-variant sites). Number of indels (insertion and deletions sites) events was 4, with average length 3.35; number of Indel haplotypes is 5 with haplotype diversity 0.747. Tajima D was insignificant for indels, which indicate neutrality in genetic diversity. Number of synonymous substitution sites was 298.86 with nucleotide diversity (Pi) of 0.39, while non-synonymous substitution sites was 1030.14 with Pi = 0.12.

The comparative analysis includ nine different fresh family species, namely; *Centrarchidae* (1), *Cichlidae* (8), *Labridae* (1), *Latidae* (2), *Nototheniidae* (1), *Paralichthyida* (1), *Poecilidae* (6), *Pomacentridae* (1), *Sciaenidae* (1). Sequence conservation among species (C) was 0.429, three conserved regions were detected (started from 46 to 221 bp, from 128 to 224 bp and from 999 to 1320 bp) with a minimum length of 95 bp and conservation threshold (CT) = 0.52. Average homozygosity in conserved regions was 0.85, with p-value below 0.5%. Families with more than one species were described in Table (1), where the family *Cichlidae* scored the highest number of species (8) but with a lower number of both polymorphic sites and mutations (50) than *Poecilidae*.

By comparing the sequence of *Cichlidae* family with the other two fami-

lies, *Poecilidae* vs *Cichlidae*, the number of fixed differences was 285 of 1300; number of mutations that were polymorphic in *Poecilidae* but monomorphic in *Cichlidae* 62 and vice versa was 49 and the shared mutation was one. On the other hand, *Latidae* vs *Cichlidae*, the number of fixed differences was 260, number of mutations that were polymorphic in *Latidae* but monomorphic in *Cichlidae* was zero and vice versa was 50, no shared mutations were found.

***Phylogeny and evolution***

Based on the maximum likelihood tree, seven families were defined by 23 species in 45 nodes based on the taxonomy of each species included in the phylogenetic analysis. The family *Cichlidae* was clustered by the maximum bootstrap value (bsv) of 1.00, represented by 8 sequences: four *O. niloticus* samples (current samples + 3 samples of the GenBank) highly clustered with 0.941 bsv, *Pundamilia nyererei*, *Haplochromis brutoni* grouped together by 0.752 bsv and monophyletically clustered with *Maylandia zebra* by 0.06 bsv all clustered with *Neoprologus brichar* by 0.992 bsv. The family *Poecilidae* was clustered by 1.00 bsv, represented by 7 sequences of 5 different species (*Poecilia formosa, P. mexicana, P. latipanna, P. reticulata* and *Xiphophorus maculatus*). Both families were highly clustered together by 0.87 bsv. All other families and the outgroup (*Paralichthys olivaceus*) were group together as subdivided cluster (Fig. 2).

Phylogenetic analysis provided strong support for the identity of the majority of CYPs in fresh water fishes. Zhang *et al*. (2014) performed syntenic analyses which showed that genes in CYP subfamilies existed as tandem duplication arrays which shared synteny with corresponding gene clusters in zebrafish, indicating that members of these subfamilies could be derived from recent lineage-specific gene duplication events. In the current study, the conservative regions among the studied families were found, as such, these tandem arranged *CYP* genes are paralogous one another within conserved regions. The high level of lineage-specific multiplication of these *CYP* genes may suggest that the involved organisms were under evolutionary selection for the rapid expansion of such *CYP* genes, perhaps in the face of heavy environmental pollution (Zhang *et al*., 2014).

## SUMMARY

The cytochrome P450 (CYP) superfamily of heme-containing enzymes catalyzes Phase I biotransformation of endogenous and xenobiotic compounds, including fatty acids, steroids, drugs, and environmental contaminants. Common variations (polymorphisms) in *cytochrome P450* genes can affect the function of the enzymes. The effects of polymorphisms are most prominently seen in the breakdown of medications. In fish, members of the CYP2 and CYP3A families play a major role in the metabolism of xenobiotics and endogenous compounds. The current study aimed to isolate and

compare *cytochrome P450 3A40* gene from *Oreochromis niloticus* (Nile tilapia) to understand its diversity and evolution in comparison with fresh water fish species available in the GenBank database. Total length of 1300 bp was obtained and its polymorphism with similar samples from *O. niloticus* in the GenBank was determined. Site no. 41 T>A changed the amino acid from Phenylalanine (F) to Tyrosine (Y), while sites no. 43 and 54 caused no effect (silent mutations), and site no. 74 A>G changes the amino acid from Glutamic Acid (E) to Glycine (G), in which the two later mutations formed a different protein isoform in its conformational structure. Phylogenetic analysis reflected a clear divergence of fresh water families (*Cichlidae* and *Poecilidae*) from other families, while fixed and shared mutations between families were found. Phylogenetic analysis provided strong support for the identity of the majority of CYPs in fresh water fishes. In the current study, conservative regions among the studied families were found.

## REFERENCES

Barber, D. S., A. J. McNally, N. Garcia-Reyero and N. D. Denslow (2007). Exposure to p,p′-DDE or dieldrin during the reproductive season alters hepatic CYP expression in largemouth bass (*Micropterus salmoides*). Aquatic Toxicology (Amsterdam, Netherlands), 81: 27-35.

Brian, W. R., M. A. Sari, M. Iwasaki, T. Shimada, L. S. Kaminsky and F. P. Guengerich (1990). Catalytic activities of human liver cytochrome P-450 IIIA4 expressed in *Saccharomyces cerevisiae*. Biochemistry, 29: 11280-11292.

Guengerich, F. P. (2007). *Cytochrome p450* and chemical toxicology. Chemical Research in Toxicology, 21: 70-83.

Hannemann, F., A. Bichet, K. M. Ewen and R. Bernhardt (2007). Cytochrome P450 systems-biological variations of electron transport chains. Biochimica et Biophysica Acta (BBA)-General Subjects, 1770: 330-344.

Librado, P. and J. Rozas (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451-1452.

Lynch, T. and A. Price (2007). The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am. Fam. Physician, 76: 391-396.

Miranda, C. L., M. C. Henderson and D. R. Buhler (1998). Evaluation of chemicals as inhibitors of trout cytochrome P450s. Toxicology and Applied Pharmacology, 148: 237-244.

Miranda, C. L., J. L. Wang, M. C. Henderson and D. R. Buhler (1989). Purification and characterization of

hepatic steroid hydroxylases from untreated rainbow trout. Archives of Biochemistry and Biophysics, 268: 227-238.

Nelson, D. R. (2013). A world of cytochrome P450s. Philosophical Transactions of The Royal Society, B: 368.

Thibaut, R., L. Debrauwer, E. Perdu, A. Goksøyr, J. P. Cravedi and A. Arukwe (2002). Regio-specific hydroxylation of nonylphenol and the involvement of CYP2K-and CYP2M-like iso-enzymes in Atlantic salmon (*Salmo salar*). Aquatic Toxicology, 56: 177-190.

Zhang, J., J. Yao, R. Wang, Y. Zhang, S. Liu, L. Sun and Z. Liu (2014). The *cytochrome P450* genes of channel catfish: their involvement in disease defense responses as revealed by meta-analysis of RNA-Seq datasets. Biochimica et Biophysica Acta, 1840: 2813-2828.

Table (1): Families with more than one species are described by number of species (n), number of polymorphic sites within family (PS), total number of detected mutations (mu), nucleotide diversity (Pi) with standard deviation (PiSdt).

| Family | n | PS | mu | Pi | PiSdt |
|---|---|---|---|---|---|
| *Cichlidae* | 8 | 50 | 50 | 0.017 | 0.002 |
| *Poecilidae* | 6 | 62 | 63 | 0.017 | 0.002 |
| *Latidae* | 2 | 0 | 0 | 0 | 0 |

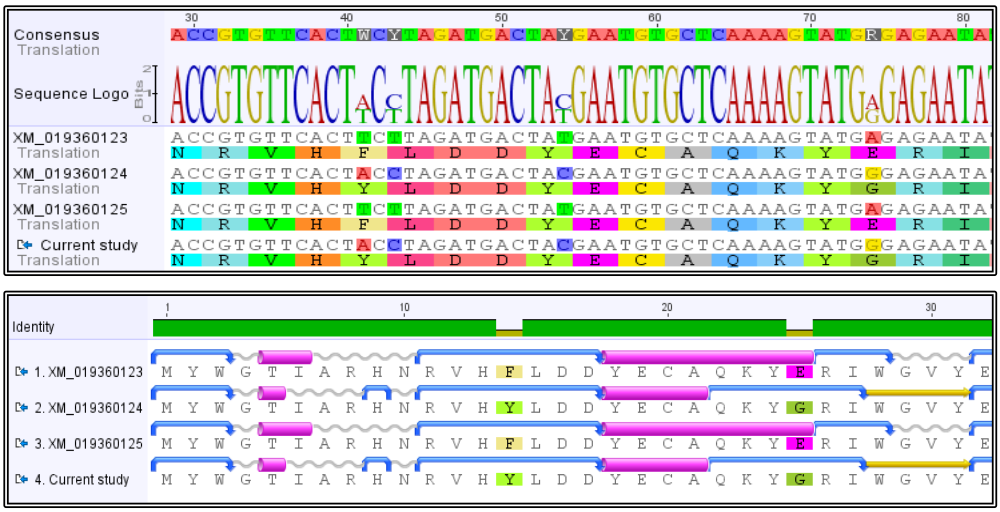


Fig. (1): Windows of 52 nucleotides of a pairwise alignment between *O. niloticus* CYP450 partial sequence. Four sites were polymorphic, while only two showed changes in amino acid sequence (Above). Prediction of protein secondary structure is indicated as: blue turns, grey coils, Pink alpha helix and yellow beta strands (below).

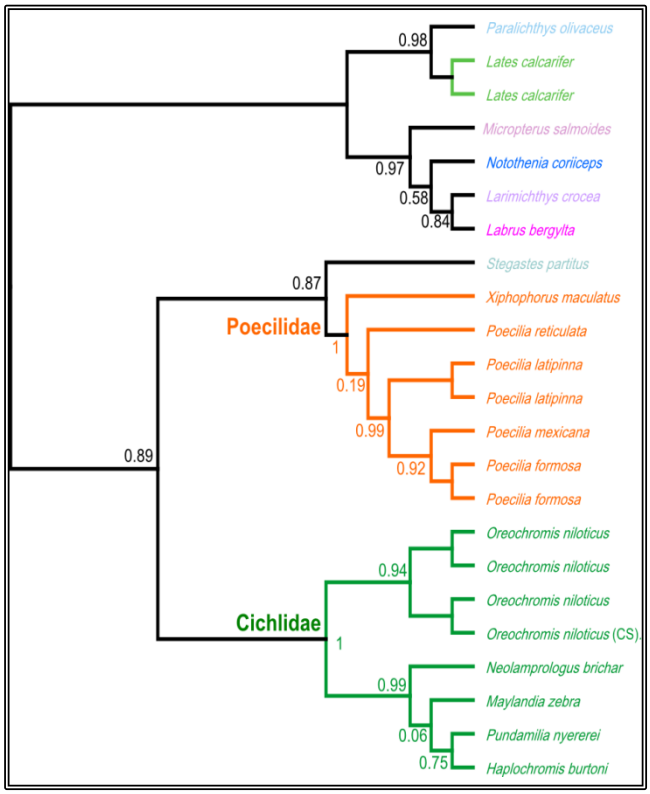


Fig. (2): Maximum likelihood phylogenetic tree based on the partial sequence of *CYP450*, where *Cichlidae* (green) and *Poecilidae* (orange) families are well defined. Each other species were coloured according to its family assignment.