

ADAPTIVE EVOLUTION IN THE *SIGANUS SPINUS* (SCRIBBLED RABBITFISH)
LIVER TRANSCRIPTOME

BY

Jude Anthony Flores Martinez

A thesis submitted in partial fulfillment of the
Requirements for the degree of

MASTER OF SCIENCE
IN
BIOLOGY

SUPERVISORY COMMITTEE

Dr. Jason S. Biggs
Dr. Bastian Bentlage
Dr. David Combosch
Dr. Melissa M. Taitano

UNIVERSITY OF GUAM

December 2019

Abstract

One overarching goal in the field of marine chemical ecology is to understand how chemicals produced by one organism affect the biology and/or population dynamics of others within its environment. A multitude of studies conducted over the last five decades demonstrate that many marine organisms produce chemical substances designed in such a way as not to alter their own internal biochemistry, but rather those within others, (*i.e.*, secondary metabolites). These chemical compounds thereby act as mediators of ecological interactions between organisms and have evolved to produce specific outcomes (*e.g.*, antipredator adaptations). Central to this theme is the evolutionary arms race between marine plants and herbivores, which to date has largely centered on identifying bioactive chemicals produced by algae and testing how their presence alters algal susceptibility to predation. Much less work has been conducted to understand the compensatory mechanisms evolved within marine consumers driving prey preference and overall dietary breadth. To maintain physiological homeostasis while experiencing chemical insult, marine herbivores have adapted integrated control mechanisms. Protection from harmful chemical effects is orchestrated by a group of enzymes known as the xenobiotic metabolizing enzymes (XMEs), which play a key role in the breakdown and elimination of foreign compounds. In this thesis I described putative XME transcripts expressed within *Siganus spinus* liver, compared them with those of model organisms, and evaluated the existence of selection markers, which may underlie the dietary breadth of this generalist herbivore as compared to other teleosts. My major results include the verification of 64 annotated *S. spinus* XME coding sequences, identification of XME genes that exhibit signatures of positive selection, and

identification of specific domains within XME-enzymes that exhibit signatures of positive selection. This study provided evidence that putative *S. spinus* liver XMEs may be used to establish orthologous relationships with model organisms and other fish to potentially strengthen functional inferences of genes across the teleost evolutionary tree. Finally, positive selection in XMEs suggests that XME evolution could have allowed *S. spinus* to exploit ecological niches that remain inaccessible to other herbivores.

Keywords: *marine chemical ecology, siganus spinus, marine herbivores, xenobiotic metabolizing enzymes, positive selection*

Acknowledgements

I wish to express my sincerest gratitude to all those who have helped me get to this point today. To my thesis committee, Dr. Jason S. Biggs, Dr. Bastian Bentlage, Dr. David Combosch, and Dr. Melissa M. Taitano, for their time, patience, guidance, encouragement, and constructive criticism throughout my master's career. To the Utah Sequencing Core, for sequencing *S. spinus*' liver transcriptome. To Sheila Podell and Tessa Pierce at the Scripps Institute of Oceanography, for annotating *S. spinus*' liver transcriptome. To the University of Guam staff, for providing the infrastructure to conduct my thesis and their guidance through the graduate process. To Val Paul and Carmen Emborski, for providing the foundations that this thesis was built upon. To Jordan Gault, for helping me through R and data analyses. To all the Marine Lab students and faculty, for your friendship and the great times. To Jason Miller and Joe Cummings, for their invaluable knowledge of boating, mechanics, and fishing. To AJ Reyes, for the countless hours spent in the ocean, collecting samples, and working in the lab. To Angie Duenas and Christine Camacho, for guiding me through the administrative process, providing endless motivation and encouragement, and sharing their delicious home-cooking! To all my friends near and far, who have spent countless hours with me in and out of the lab encouraging me to complete my thesis. To my family, for their motivation, prayers, support, and imparting local fishing knowledge throughout my career. To my girlfriend, Aubrienne Barretto, and her son, Theo, for their patience and strength this past year. To my parents, Peter Roy and Elizabeth Martinez, for their endless prayers, sacrifices, guidance, patience, and encouragement throughout my master's career. *Un dangkulo na Si Yu'os Ma'ase!*

Table of Contents

Abstract	2
Acknowledgements	4
Table of Contents	5
List of Figures	7
List of Tables	8
Introduction	9
<i>Xenobiotic metabolism from a predator’s perspective</i>	9
<i>Xenobiotic metabolizing enzymes in fish and their potential to access new resources</i>	14
Materials and Methods	16
<i>Study organism</i>	17
<i>Sample collection</i>	18
<i>Total RNA extraction</i>	20
<i>cDNA library preparation</i>	21
<i>Identification of mRNAs encoding putative XMEs</i>	23
<i>Phylogenetic gene annotation</i>	23
<i>Studying positive selection in protein coding genes</i>	26
<i>Synonymous to nonsynonymous substitution rates</i>	27
Results	31
<i>Transcriptome sequencing, assembly, and putative annotations</i>	31
<i>Gene tree</i>	33
<i>Gene Tree – Cytochrome P450</i>	33
<i>Gene Tree – Glutathione S-Transferase</i>	39

<i>Gene Tree – ATP-Binding Cassette Transporter</i>	42
<i>Positive selection</i>	47
Discussion	59
<i>Gene trees</i>	59
<i>Positive selection</i>	60
Literature Cited	67

List of Figures

Figure 1. Xenobiotic metabolism can be categorically divided into three phases: phase I (oxidation), phase II (conjugation), and phase III (transport/excretion) reactions.....	13
Figure 2. Location of sampling sites around Guam.	19
Figure 3. RNA integrity (RIN) scores from total-RNA extraction.	22
Figure 4. Maximum likelihood-based phylogenetic analysis of putative <i>S. spinus</i> liver cytochrome P450 protein sequences.	38
Figure 5. Maximum likelihood-based phylogenetic analysis of putative <i>S. spinus</i> liver glutathione S-transferase protein sequences.	41
Figure 6. Maximum likelihood-based phylogenetic analysis of putative <i>S. spinus</i> liver ATP-binding cassette transporter protein sequences.	46
Figure 7. Box-plot graphs reflect the results from the full-length (whole protein) analysis using the gMYN method.	49
Figure 8. Box-plot graphs reflect the results from the full-length (whole protein) analysis using the gMYN method.	50
Figure 9. The sliding window analysis detected domains with Ka/Ks ratios greater than or equal to 1.0 within 20 putative <i>S. spinus</i> XMEs using the gMYN method.	53
Figure 10. ABCC9 from the sliding window analysis.	54
Figure 11. CYP3A40 from the sliding window analysis.	55
Figure 12. MDR1 from the sliding window analysis.	56
Figure 13. GSTA4 from the sliding window analysis.....	57
Figure 14. GSTM1 from the sliding window analysis.	58

List of Tables

Table 1. Major putative <i>S. spinus</i> liver XMEs selected and verified for maximum likelihood-based phylogenetic analysis.....	25
Table 2. Putative <i>S. spinus</i> XMEs selected for positive selection analysis.....	29
Table 3. Results from University of Utah Sequencing Core (sequencing) and Scripps Institution of Oceanography (assembly and annotation).....	32
Table 4. Summary of <i>S. spinus</i> XMEs used in whole protein and sliding window positive selection analyses.....	51
Table 5. Summary of <i>S. spinus</i> miscellaneous response genes used in whole protein and sliding window positive selection analyses.	52

Introduction

Studies on algal-herbivore interactions have unearthed a plethora of algal defense strategies (Hay and Fenical, 1988; Paul, 1992; Steinberg, 1992; Paul et al., 2001; Paul and Puglisi, 2004; Paul et al., 2006). Many of the most common algal species produced chemical compounds which resemble chemical defenses discovered within terrestrial plants, including terpenes, acetogenins, alkaloids, and polyphenolics; most of which are lipid soluble and anti-nutritive (Paul and Ritson-Williams, 2008; Sotka and Whalen, 2008). Extensive literature regarding the diversity, concentration, and distribution of these chemicals (secondary metabolites) exists for marine plants (Paul and Hay, 1986; Hay and Fenical, 1988; Paul, 1992; Paul and Puglisi, 2004; Paul et al., 2006; Paul et al., 2007; Sotka and Whalen, 2008). Their effectiveness as defenses against particular herbivore guilds, including mollusks, urchins, crustaceans, and fishes have also been extensively studied (Hay et al., 1987; Paul et al., 1988; Paul et al., 1990; Paul et al., 1992; Cronin, 2001; Van Alstyne et al., 2001; Amsler and Fairhead, 2006).

In contrast, much less is understood about the strategies used by predators to overcome prey chemical defenses (Sotka and Whalen, 2008; Sotka et al., 2009). More thorough understanding of how consumers cope with prey chemical defenses are needed not only to understand how biochemistry structures ecosystems, but also to guide efforts to turn these bioactive compounds into effective pharmacological agents (Sotka et al., 2009). This knowledge might also prove critical to the maintenance of biodiversity in the face of climate change and the global redistribution of species.

Xenobiotic metabolism from a predator's perspective

For all intents and purposes, ingestion is a commitment to the absorption of all constituents within a meal, harmful or otherwise. Bioavailable substances, such as

proteins, sugars, fatty acids, metals, and lipophilic chemicals are liberated from food through the digestive process, absorbed into the blood stream, and eventually distributed throughout the body (McLean and Duncan, 2006). Biologically active compounds (*i.e.*, substances that alter physiological mechanisms occurring within living cells) that are not fundamental building blocks of life also enter the body through these processes (Casarett et al., 2003; Katzung et al., 2012). Without mechanisms for their removal, bioactive compounds would eventually saturate the body, perturb normal physiology, and elicit toxicity (Casarett et al., 2003, Karban and Agrawal, 2002; Sotka and Whalen, 2008; Torregrossa and Dearing, 2009; Iason, 2005). Chemical insults coming from the environment are common enough that every living cell has some capacity to express detoxifying enzymes (Feyereisen, 1999; Katzung et al., 2012; Sotka and Whalen, 2008); and predators are particularly adapted to such challenges with specific body plans and biological mechanisms to mitigate risks associated with feeding on their preferred prey (Casarett et al., 2003; Katzung et al., 2012).

In vertebrates, where groups of tissues have evolved to perform specific functions, the metabolism of diet-derived chemicals is carefully orchestrated. Everything that enters the blood stream via the digestive tract must first travel through the portal vein and into the liver, the predominant organ for detoxifying and eliminating foreign compounds (henceforth, xenobiotics) from the body (Casarett et al., 2003; Katzung et al., 2012). This universal path of blood flow gives vertebrates a chance to mitigate the availability of bioactive substances before they enter systemic circulation (Casarett et al., 2003). This front line of defense against xenobiotics is so important to pharmacology that it has been termed ‘first-pass metabolism,’ which can, for example, be extensive enough

in humans to limit the bioavailability and efficacy of many pharmaceutical and personal care products (Pond and Tozer, 1984). Hepatocytes are well equipped for this task, as they have the ability to express a variety of inducible enzyme superfamilies collectively referred to as ‘Xenobiotic Metabolizing Enzymes (XMEs)’ dedicated to these processes (Casarett et al., 2003; Katzung et al., 2012, Sotka and Whalen, 2008; Arellano-Aguilar et al., 2009; Ferreira et al., 2014; Luckenbach et al., 2014; Uno et al., 2012). However, it is important to note that some XMEs also perform endogenous functions by altering the pharmacokinetics of fatty acids, steroids, prostaglandins, retinoids, bile acids, sterols, and biogenic amines (Nelson et al., 1996; Mansuy, 1998; Feyereisen, 1999; Ingelman-Sundberg, 2002; Luckenbach et al., 2014).

To ease discussion, xenobiotic metabolism is often divided into three phases: I) oxidation, II) conjugation, and III) transport/excretion (Casarett et al., 2003; Sotka and Whalen, 2008). Together, they catalyze a suite of highly complex reactions that either convert lipophilic compounds into water-soluble and often less toxic products, and/or actively remove them from cells and the body (Sotka and Whalen, 2008). Phase I enzymes catalyze the unmasking or addition of a polar functional group into the xenobiotic, whereas phase II enzymes cause relatively large polar molecules to covalently bond to functional groups on the xenobiotic (Chang and Kam, 1999; Doi et al., 2004; Sotka and Whalen, 2008; Jancova et al., 2010; Katzung et al., 2012). Phase III enzymes transport phase I and II metabolites, as well as unmodified xenobiotics, out of cells against concentration gradients by using the energy liberated from ATP (Casarett et al., 2003; Karban and Agrawal, 2002; Sotka and Whalen, 2008; Luckenbach et al., 2014). Although phase I, II, and III XMEs are extremely diverse, the majority of enzyme

superfamilies involved include the cytochromes P450 (CYPs; phase I), glutathione s-transferases (GSTs; phase II), and ATP-binding cassette transporters (ABCs; phase III) (Figure 1) (Karban and Agrawal, 2002; Sotka and Whalen, 2008).

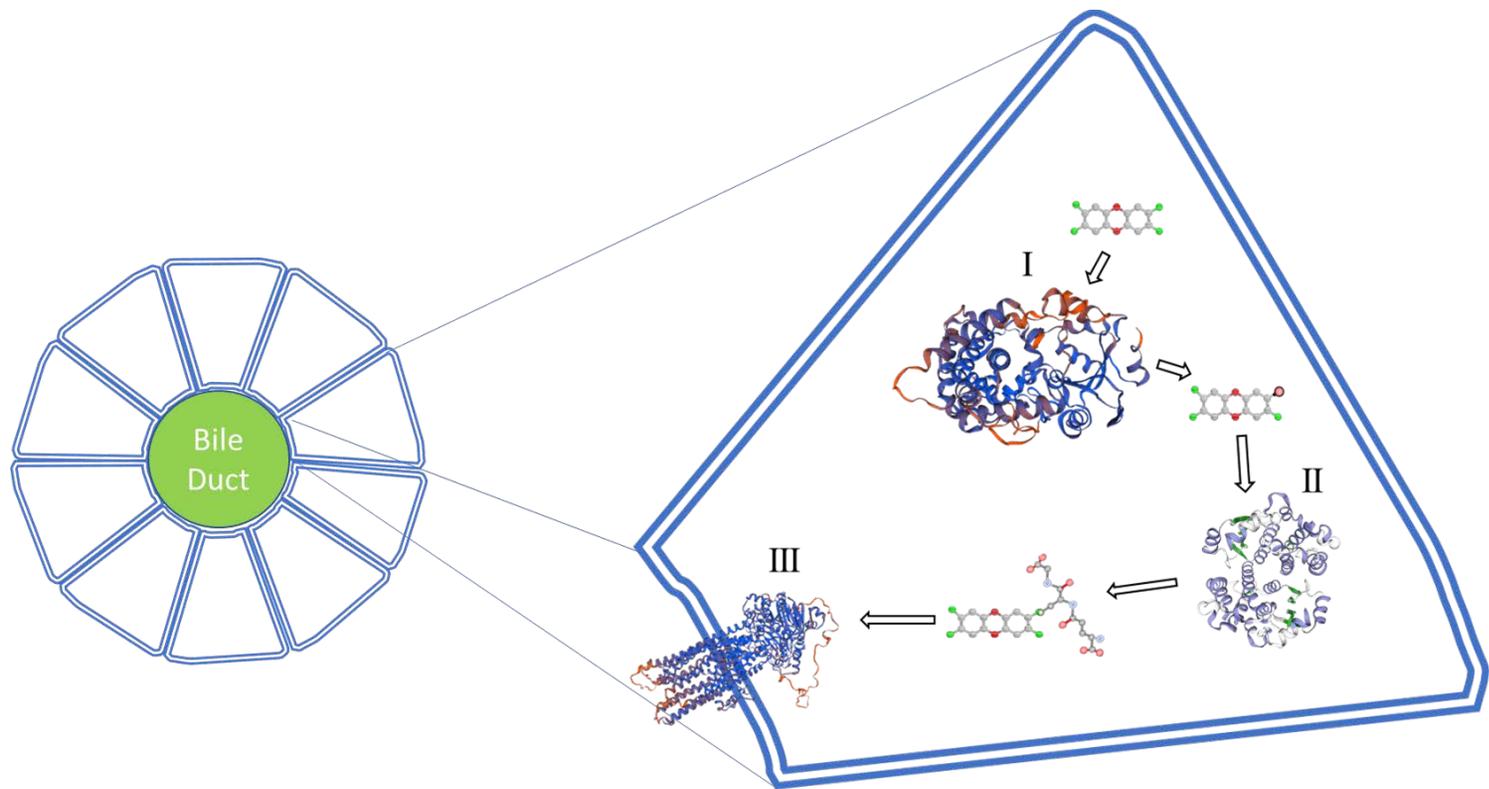


Figure 1. Xenobiotic metabolism can be categorically divided into three phases: phase I (oxidation), phase II (conjugation), and phase III (transport/excretion) reactions. Phase I enzymes (CYPs) catalyze the addition of a polar functional group into lipophilic substrates, while Phase II enzymes (GSTs) use these groups as a “handle” for conjugation with moieties such as glutathione, glucuronic acid, or glycine. Finally, unmodified xenobiotics and their Phase I and II metabolites can be excreted from the cell to the extracellular space or compartmentalized into subcellular organelles by Phase III transmembrane ATP-dependent efflux pumps (ABC transporters; drug pumps).

These main XMEs are classified within their respective superfamilies based on sequence identities, structural similarities, substrate/inhibitor specificities, and phylogenetic/syntenic relationships (Nelson et al., 1996; Nelson et al., 2003; Trute et al., 2007; Liu et al., 2013; Ferreira et al., 2014; Zhang et al., 2014; Jeong et al., 2015). CYPs are categorized into families (>40% sequence homology), subfamilies (>55% sequence homology), and individual genes are named based on the order of their discovery (*e.g.*, CYP1A1) (Nelson et al., 1996; Arellano-Aguilar et al., 2009; Zhang et al., 2014). GSTs are categorized into 8 classes: alpha, kappa, mu, omega, pi, sigma, theta, and zeta; wherein GSTs are assigned to the same class if they share greater than 40% sequence homology, substrate and inhibitor specificity, primary and tertiary structure similarity, and immunological identity, along with a number that denotes the protein (*e.g.*, GSTM1) (Hayes and Pulford, 1995; Sheehan et al., 2001; Konishi et al., 2005; Trute et al., 2007). ABCs are categorized into 7-8 subfamilies (A-G in humans and A-H in eukaryotes) with a number that denotes the protein (*e.g.*, ABCC9) on the basis of sequence homology and the organization of the ATP-binding domains, also known as nucleotide binding domains (NBDs) (Dean and Annilo, 2005; Liu et al., 2013; Ferreira et al., 2014; Jeong et al., 2015).

Xenobiotic metabolizing enzymes in fish and their potential to access new resources

Several sequencing and immunochemical projects conducted over the past 30 years have tremendously improved our understanding of the biochemical roles CYPs, GSTs, and ABC transporters play in fish (Arellano-Aguilar et al., 2009; Buhler and Wang-Buhler, 1998; Sotka and Whalen, 2008; Sotka et al., 2009; Uno et al., 2012; Ferreira et al., 2014; Luckenbach et al., 2014; Qian, 2014). These studies have established model organisms such as the zebrafish (*Danio rerio*), rainbow trout

(*Oncorhynchus mykiss*), and Japanese pufferfish (*Takifugu rubripes*), all of which have been used for a number of genomic and pharmacologic studies (Arellano-Aguilar et al., 2009; Carvan III et al., 2008; Dong et al., 2013; Galus et al., 2013; Kubota et al., 2013; Perez et al., 2013; Ulloa et al., 2014), aquatic pollution studies (Brammell et al., 2010; Buhler and Wang-Buhler, 1998; Connon et al., 2012; Pacitti et al., 2013; Smith et al., 2010; Smith et al., 2012), and multiple genome assemblies and evolutionary studies (Nordberg et al., 2014), respectively. These seminal works, as well as the countless others like them, have documented XME distribution among tissues, conserved domains/motifs, physiological roles, and suggest that the complexity of fish XMEs rival those documented in terrestrial organisms, including humans (Sotka and Whalen, 2008; Uno et al., 2012; Ferreira et al., 2014; Luckenbach et al., 2014).

This complexity of XMEs in teleost fish may enable them to exploit potentially new resources/niches because of their high diversity and copy number due to historic genome duplication events (Robinson-Rechavi et al., 2001; Hoegg et al., 2004; Trute et al., 2007; Uno et al., 2012; Liu et al., 2013; Luckenbach et al., 2014; Liu et al., 2016). Whole genome duplication (WGD) is one of the main driving forces in the evolution of many teleost genes because it produces an enormous number of novel genes with the potential for new functions (Liu et al., 2016). The high diversity of teleost fish is thought to correlate with three rounds of teleost-specific WGD that took place in the common ancestor of all extant teleosts (Robinson-Rechavi et al., 2001; Hoegg et al., 2004; Liu et al., 2013; Liu et al., 2016). As a result, teleost fish frequently possess paralogous copies for many genes (Liu et al., 2013; Liu et al., 2016). Additional lineage-specific genome duplications and gene losses observed within particular groups during evolution have

produced an enormous number of novel genes with the potential for partitioned or new functions (Robinson-Rechavi et al., 2001; Brunet et al., 2006; Liu et al., 2013; Liu et al., 2016), including XMEs. This mode of genome diversification is likely to allow herbivorous fish to adapt to and exploit new resources, especially on coral reefs (Steneck et al., 2017). However, due to these gene duplication events, the functions of XMEs in teleost fish remain difficult to infer based solely on transitive annotation (Trute et al., 2007; Uno et al., 2012; Luckenbach et al., 2014) and few studies have investigated the links between their expression and an ecological and evolutionary outcome (Vrolijk and Targett, 1992; Keller et al., 2006; DeBusk et al., 2008; Sotka and Whalen, 2008). Establishment of orthologous relationships across model organisms and addition of further model fish have the potential to strengthen functional inferences of genes across the tree of teleosts (Trute et al., 2007; Uno et al., 2012; Luckenbach et al., 2014). Additionally, investigating adaptive evolution may provide insight on how teleost fish exploit ecological niches inaccessible to their competitors. For this study, I aim to contribute to our understanding of XME evolution in tropical marine herbivores, specifically the scribbled rabbitfish – *Siganus spinus*.

Materials and Methods

The goal of this thesis was to study adaptive evolution of putative XMEs expressed in the *S. spinus* liver. These transcripts serve as potential references for understanding marine herbivore ecology, physiology, and toxicology. To achieve these goals, *S. spinus* were collected from around Guam, mature mRNA was extracted from their livers, sent to collaborators to be sequenced and annotated, then I identified putative XMEs from the *S. spinus* liver transcriptome, verified their annotations by inferring gene

family phylogenies using homologs from model organisms, and performed pairwise comparisons of synonymous to nonsynonymous substitutions rates between *S. spinus* and other fish to identify selection.

Study organism

Siganus spinus (Siganidae) was chosen as the focus of my study because it has attributes amenable to becoming a model organism for studying marine molecular ecology, especially in terms of algal-herbivore interactions (*e.g.*, the evolution of genes involved in the metabolism of diet-derived xenobiotics). Past studies showed that *S. spinus* consumes various chemically-defended algae that others do not, as well as artificial diets laced with algal secondary metabolites produced by these prey (Nagle et al., 1996; Paul et al., 1988; Paul et al., 1992; Thacker et al., 1997). Building on this background, I aim to understand XME evolution within this species to lay the foundation for a mechanistic understanding of *S. spinus*' ability to sustain a diet of algae rich in secondary metabolites. The most relevant study to date was conducted by Emborski et al. (2012) that demonstrated that both CYP1A1 protein and catalytic activity increased in the presence of beta-naphthaflavone (BNF), a classical inducer of the polyaromatic hydrocarbon removal system. This enzymatic response was both time- and dose-dependent, suggesting that induction occurred at the level of gene transcription. As CYP1A is well documented within the literature for its predominant role in the metabolism of combustion products and fossil fuels, phenotypic plasticity in CYP1A expression provides an opportunity for it to be used as a biomarker for organo-pollution in the tropical Indo-Pacific (Emborski et al., 2012). Finally, sequence analyses of their protein-coding genes may provide insight about their evolution as well as the versatility of their XMEs.

Sample collection

The goal of this thesis was to sequence a comprehensive cDNA library of all possible transcripts expressed within *S. spinus* livers. In order to make sure to capture the diversity of XMEs expressed within *S. spinus*, individuals varying in size (proxy for age) and a number of fish that comprised male and females as well as juveniles and adults were collected from multiple habitats on Guam. At collection times (20:00-01:00), on the nights of March (19, 21, 29, 30) and April (5, 10, 20, 24, 26, 27, 28) 2012, five *S. spinus* were captured from fifteen different locations within Guam coastal waters (Figure 2).

Fish were euthanized at each site by pithing and their livers dissected – following our institution’s approved method – washed by immersing in 70% ethanol (500 ml total volume) and blotting dry with a fresh chem-wipe, and then stabilized within 2.0 mL of RNAlater (Sigma-Aldrich Corp.; St. Louis, MO, USA) at ambient temperature for 5 minutes prior to rapid-freezing in an insulated container of 2 kg dry ice (-78.5°C). Upon returning to the lab, all liver samples were stored in a -80°C freezer until needed.

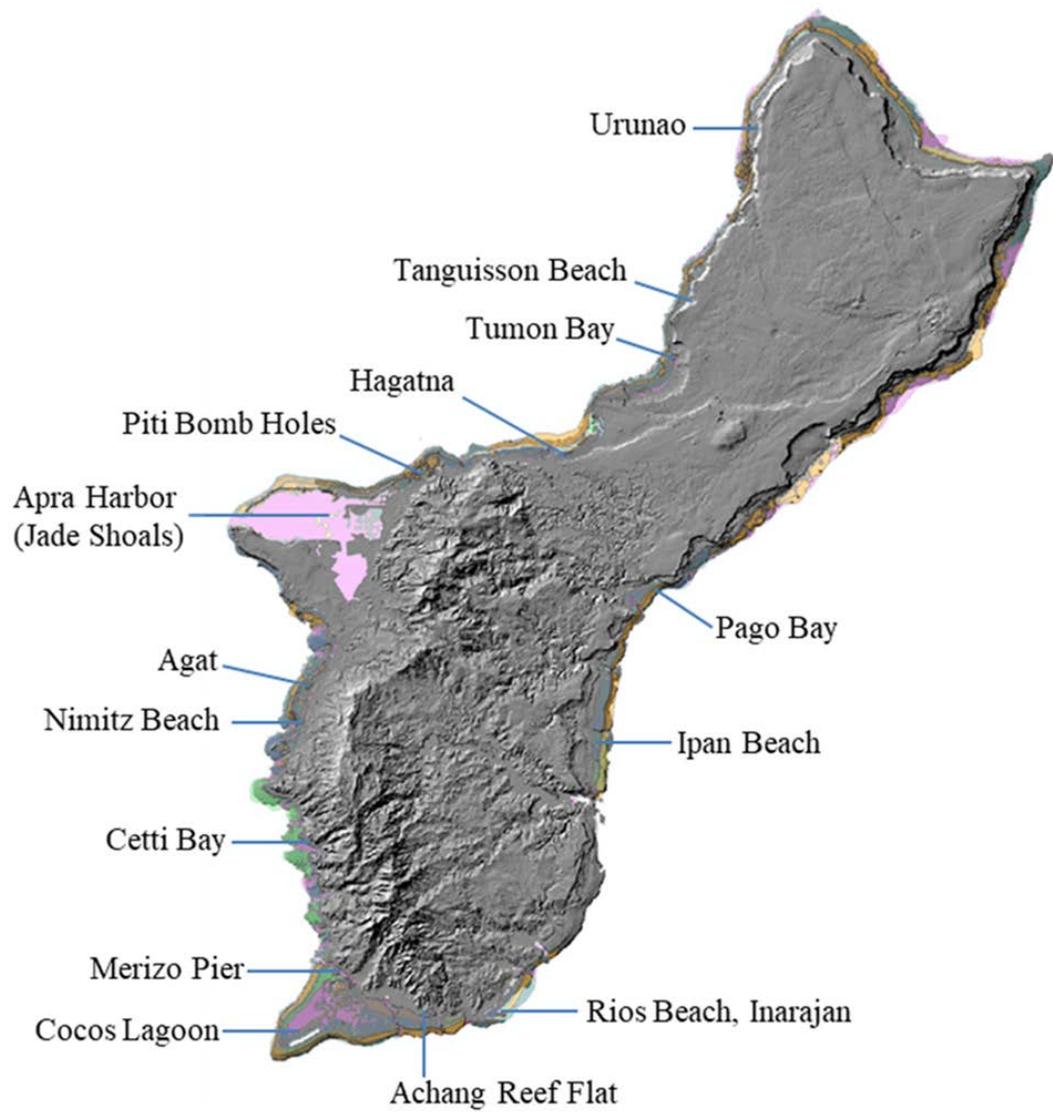


Figure 2. Location of sampling sites around Guam. Following night collections from each site, wild-caught *S. spinus* livers were extracted and preserved in RNAlater.

Total RNA extraction

To initiate the extraction protocol, each liver sample tube was placed in an ice bath and allowed to thaw for ten minutes. Once thawed, liver tissues were blotted dry, placed onto a sheet of aluminum foil, and sectioned using a fresh, sterile razor blade. Approximately, 100 mg of this tissue was transferred into a 50 ml conical tube containing 2 ml of TRIzol Reagent (Thermo Fisher Scientific; Waltham, MA, USA). The remaining tissue was cataloged within our lab's database, and returned to long-term storage. TRIzol immersions were allowed to stand for five minutes at ambient temperature, and then homogenized for ten seconds with a handheld rotor–stator homogenizer.

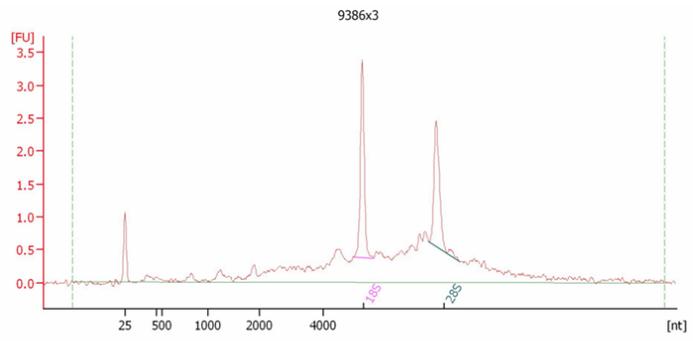
Downstream total-RNA isolation was completed for the TRIzol Plus RNA Purification Kit (Thermo Fisher Scientific; Waltham, MA, USA) and PureLink Micro-to-Mid RNA Total-RNA Purification System (Thermo Fisher Scientific; Waltham, MA, USA) following the manufacturer suggested protocols with the addition of 2-Mercaptoethanol to our ethanol washing solution to denature ribonucleases released during cell lysis. Finally, to further increase XME diversity, mRNA extracts prepared this same way from three *S. spinus*, each of which received an intraperitoneal (IP) injection of BNF, rifampicin, or dexamethasone at a concentration of 50 mg/kg (Emborski et al., 2012), were added to the total-mRNA pool.

Resulting mRNA-enriched, total-RNA sample concentrations were determined (ng/uL) using a Nano-Drop 1000 (Thermo Fisher Scientific; Waltham, MA, USA). Aliquots containing equal amounts of nucleic acid were transferred from each extract and combined within two other RNase/DNase-free microcentrifuge tubes to create two *S. spinus* total-mRNA pools, each amounting to 1 mg of total-RNA. One of these tubes was

subjected to lyophilization, whereas the other was allowed to remain solubilized in the kit's elution buffer.

cDNA library preparation

Both tubes were then sent via dry-shipper (-196°C) to the University of Utah Health Sciences Sequencing Core Facility (Salt Lake City, Utah, USA), and assessed for RNA integrity (RIN) using an Agilent 2100 Bioanalyzer (capillary gel electrophoresis; Santa Clara, CA, USA). RIN scores for these two samples were 6.70 and 8.20, respectively (Figure 3), suggesting that lyophilization in elution buffer degrades *S. spinus* total-RNA samples. The most intact sample was then prepped using the TruSeq RNA Library Prep Kit (Illumina; San Diego, CA, USA) and Illumina HiSeq 2000 platform (Illumina; San Diego, CA, USA) was used for 101 paired-end sequencing.

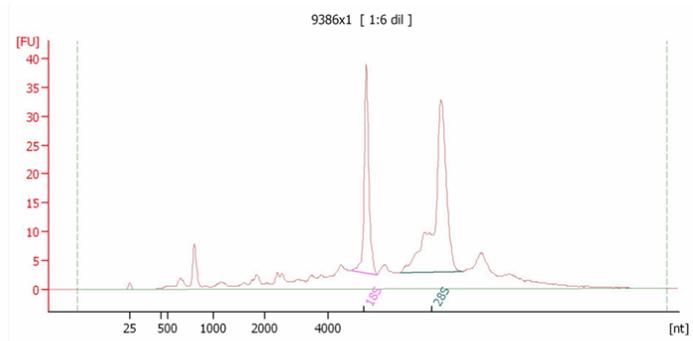


Overall Results for sample 11 : 9386x3

RNA Area:	23.5	RNA Integrity Number (RIN):	6.7 (B.02.08)
RNA Concentration:	57 ng/μl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.8	Result Flagging Label:	RIN: 6.70

Fragment table for sample 11 : 9386x3

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	4,978	5,598	2.7	11.4
28S	7,325	8,283	2.1	9.1



Overall Results for sample 9 : 9386x1

RNA Area:	279.5	RNA Integrity Number (RIN):	8.2 (B.02.08)
RNA Concentration:	684 ng/μl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	1.9	Result Flagging Label:	RIN: 8.20

Fragment table for sample 9 : 9386x1

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	4,700	5,544	38.3	13.7
28S	6,274	8,242	72.1	25.8

Figure 3. RNA integrity (RIN) scores from total-RNA extraction.

De novo assembly and annotation of putative mature mRNA

Raw reads were sent to our collaborators at the Scripps Institution of Oceanography (San Diego, CA, USA) where they used Trinity (Haas et al., 2013) to assemble and annotate the *S. spinus* liver transcriptome.

Identification of mRNAs encoding putative XMEs

From my Trinity-assembled coding sequence (CDS) list for the *S. spinus* liver transcriptome comprising 59,412 annotated genes, 64 putative XME-coding genes were identified for further study. These 64 genes encode proteins that are known to be involved in detoxification, metabolism, elimination, and stress responses based on the results of previous studies (Nelson et al., 1996; Mansuy, 1998; Feyereisen, 1999; Karban and Agrawal, 2002; Casarett et al., 2003; Despres et al., 2007; Li et al., 2007; Szakács et al., 2008; Sotka and Whalen, 2008; Arellano-Aguilar et al., 2009; Katzung et al., 2012; Uno et al., 2012; Ferreira et al., 2014; Luckenbach et al., 2014). Homologs were identified as the top hit from a BLASTx (GenBank's web interface) search using default settings (E value: 1e-5). These homologs served as an initial screening of the identities for the 64 putative *S. spinus* XMEs.

Phylogenetic gene annotation

Gene family trees for the three major XME superfamilies – CYPs, GSTs, and ABCs – were reconstructed using a maximum likelihood (ML) framework to verify and refine BLAST-based annotations of putative *S. spinus* homologs (Table 1). A comprehensive dataset of CYP, GST, and ABC transporter gene coding sequences was assembled for the model vertebrates human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), and Japanese pufferfish (*Takifugu rubripes*) from the Genome Reference Consortium (www.ncbi.nlm.nih.gov/grc).

TransDecoder (Haas et al., 2013) was used to identify the longest open reading frames for the putative *S. spinus* XME transcripts.

Table 1. Major putative *S. spinus* liver XMEs selected and verified for maximum likelihood-based phylogenetic analysis.

<i>S. spinus</i> XMEs used in Phylogenetic Analysis	
Gene	Sequence Length (AA)
ATP-Binding Cassette Transporter	
A1	1763
A3	1404
A5	1662
A12	110
B1	145
C9	618
C10	1536
D1	583
E1	599
G20	280
Cytochrome P450	
1A1	500
1B1	530
2C16	400
2D10	212
2D20	150
2D28	103
2F2	503
2F3	502
2F5	99
2J1	185
2J2	370
2J5	113
2J6	489
2K1 Internal	313
2K1 5' partial	142
2K3	132
3A4	224
3A9	258
3A13	171
3A17	127
3A27	501
3A30	515
3A40	492
4B1	352
20A1	462
26A1	533
26B1	371
Glutathione S-Transferase	
Alpha	226
Alpha 4	133
Mu1	227
Mu3	148
Mu4	116
Omega 1	271
Theta 1	280

The amino acid sequences from model vertebrates and *S. spinus* were aligned using MAFFT v7.271 (Kato and Standley, 2013) with default settings, yielding three alignments for phylogenetic analyses. Alignments were trimmed by trimAl 1.2rev59 to remove poorly aligned regions (Capella-Gutierrez et al., 2009). Gene family trees were inferred from CYP, GST, and ABC transporter alignments using RAxML v8.2X (Stamatakis, 2014). To evaluate robustness of relationships in the phylogeny, 100 non-parametric bootstrap replicates were calculated. Putative *S. spinus* XME transcripts whose annotations clustered with their respective homologs with bootstrap scores ~60-100 were considered as robust indicators of proper annotation. Putative *S. spinus* XME transcripts whose annotations did not cluster with their respective homologs were relabeled according to their original assembled contig descriptor (e.g., c34579_g1_i1) to indicate potentially erroneous annotation.

Studying positive selection in protein coding genes

Amino acids have multiple spellings within the genetic code. Mutations that code for the same amino acid are called synonymous mutations. Mutations that alter the amino acids being expressed are called non-synonymous mutations. Comparing synonymous (silent; K_s) and non-synonymous (amino acid changing; K_a) substitution rates in protein-coding genes provides a means for understanding molecular evolution and evaluating sequence variation for orthologs across species or lineages (Wang et al., 2010; Yang et al., 2000 Yang and Bielawski, 2000). The ratio of synonymous to non-synonymous substitutions (K_a/K_s) can provide evidence for selection. A ratio less than 1 (synonymous mutations dominate) is indicative of purifying selection while a ratio greater than 1 (non-synonymous substitutions dominate) indicates positive selection

(Yang et al., 2000; Yang and Bielawski, 2000; Zhang and Yu, 2006; Zhang et al., 2006; Wang et al., 2009).

Positive selection, also known as adaptive evolution, is the process by which beneficial alleles increase in frequency as a result of their presence imparting increases in fitness (Swanson, 2003). This mechanism of selection is difficult to detect and analyze by summing across the entire length of a coding region because neutral and deleterious mutations occur more frequently (Nei and Kumar, 2000). Furthermore, a vast majority of sites within the coding region of a gene tend to be invariable due to functional constraints (purifying selection). Thus, Ka/Ks calculations based on the substitutions across an entire gene can mask signals of positive selection and adaptation (Wang et al., 2010). Whereas, specific motifs or domains within a gene may be under positive selection, forming selective hotspots that are associated with aspects of a protein's function (Wang et al., 2010).

To identify XMEs under positive selection in *S. spinus*, I thus used two approaches that rely on Ka/Ks ratios: 1) I calculated Ka/Ks ratios for *S. spinus* genes that encode XMEs across the full length (whole protein analysis) of genes; and 2) I employed a sliding window approach to calculate Ka/Ks ratios for substrings of each gene with the aim of identifying potential selective hotspots within genes.

Synonymous to nonsynonymous substitution rates

To identify signatures of selection in *S. spinus* XMEs compared to other teleost fish, putative *S. spinus* XME transcripts were compared to orthologs from marine teleosts (Table 2). Marine teleost orthologs were obtained from the Fish-T1k transcriptome database (db.cngb.org/fisht1k/; Sun et al., 2016) using putative *S. spinus* XME transcripts as BLAST queries (BLASTn; E value: 1e-5). TransDecoder (Haas et al., 2013) was used

to identify and translate the longest open reading frame encoded within these BLAST-hits. Primary amino acid sequences inferred from Fish-T1k transcripts were aligned with their respective *S. spinus* sequence using MAFFT (Kato and Standley, 2013) under default parameters.

Table 2. Putative *S. spinus* XMEs selected for positive selection analysis. Sequence lengths (bp) and number of sequences within the alignment are provided.

<i>S. spinus</i> XMEs used in Positive Selection Analysis		
Gene	Sequence Length (bp)	# of Sequences in Alignment
ATP-Binding Cassette Transporter		
A1	5211	13
A1.2	330	57
A3	4212	58
A5	4986	17
C9	1854	20
G20	441	32
MDR1	4533	12
MDR4	2055	14
Cytochrome P450		
1A1	1500	48
1B1	750	62
2D28	279	18
2C16	1080	27
2F2	1377	27
2F3	1377	27
2J5	336	34
2J6	1461	16
2K1 Internal	804	10
2K1 6' partial	417	48
2K3	396	14
3A4	671	20
3A13	498	44
3A27	1503	41
3A40	1467	29
4B1	1053	21
20A1	1308	37
26A1	1464	11
26B1	1077	55
Thromboxane-A Synthase	792	32
Glutathione S-Transferase		
Alpha 4	396	17
Mu1	618	70
Mu3	444	65
Mu4	339	68
Omega1	717	19
Theta1	681	17
Miscellaneous <i>S. spinus</i> Liver XMEs		
AARF domain containing protein kinase 4	894	38
Acetylcholine receptor subunit alpha	1299	22
Actin alpha skeletal muscle	315	26
Apoptosis-inducing factor 3	297	24
Aryl hydrocarbon receptor	1854	36
Aryl hydrocarbon receptor nuclear translocator-like protein 1	756	83
Aryl hydrocarbon receptor nuclear translocator	1056	94
Aryl hydrocarbon receptor repressor	525	47
Breast cancer metastasis-suppressor 1 like protein-A	834	56
cAMP-dependent protein kinase catalytic subunit alpha	681	95
Carbonyl reductase NADPH 1	303	26
Catalase	1110	93
Cell cycle control protein 50A	984	87
Cell death activator CIDE-3	717	15
Cell death regulator AVEN	624	15
Estrogen-related receptor gamma	666	41
Extracellular superoxide dismutase CuZn	546	16
Growth-regulated alpha protein	297	15
Growth arrest and DNA damage-inducible protein GADD45 beta	435	22
Heat shock 70 kDa protein 14	1497	43
Heat shock protein beta-11	654	10
Heat shock protein beta-7	468	38
Nucleodioxin	1254	23
Protein disulfide-isomerase A5	1470	74
Protein disulfide-isomerase A6	579	49
Protein disulfide-isomerase TM3	1125	54
Ras GTPase-activating-like protein (IQGAP1)	699	30
Stress-induced-phosphoprotein 1	1632	84
Transcription factor 7-like 2	801	64
Tumor protein p53-inducible nuclear protein 1	885	48
Tumor protein p53-inducible nuclear protein 2	528	57

PAL2NAL version 14.0 (Suyama et al. 2006) was used to align transcript DNA sequences into codon-based alignments using the amino acid alignments as a guide. These alignments were reviewed and manually edited with Geneious v8.1.5 (www.geneious.com, San Diego, CA; Kearse et al., 2012) to minimize gaps, delete areas with low coverage across taxa on the 5' and/or 3' ends of alignment, and to ensure each alignment was in reading frame 1 to facilitate downstream processing. Alignments that had less than 10 taxa and whose lengths were less than 200bp were also removed due to a lack of information for downstream analyses.

To determine if putative *S. spinus* XME transcripts showed signs of selection compared to other marine teleosts (FishT1K), Ka/Ks ratios were calculated for '*S. spinus* versus FishT1K orthologs' and 'FishT1K orthologs versus FishT1K orthologs', the latter dataset estimating the baseline distribution of Ka/Ks ratios for significance testing. Ka/Ks ratios were estimated across full-length sequences and subsections of each sequence using a sliding window approach (window length 60, corresponding to 3 DNA sequence motifs of 20bp lengths; step length 6) using KaKs Calculator 2.0 (Zhang et al., 2006; Wang et al., 2009a) using the gamma-series modified Yang-Nielson (gMYN) method (Wang et al., 2009b). The gMYN algorithm is an approximation method that considers four major dynamic features of sequence evolution such as biases in transition/transversion rate, nucleotide frequency, unequal transitional substitution, and incorporates unequal substitution rates among different sites based on the assumption that the evolutionary rate at each site follows a mode of gamma distribution (Wang et al., 2009a; Wang et al., 2009b). For each XME in the full-length analysis, the Wilcoxon Rank Sum Test was used to determine if Ka/Ks ratio distribution observed in the '*S.*

spinus versus FishT1K orthologs' dataset were significantly different from the distribution of Ka/Ks ratios observed in the 'FishT1K versus FishT1K' dataset. XMEs found to have significantly different distributions were considered to be potential positively selected genes. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the '*S. spinus* versus FishT1K ortholog' dataset and in the 'FishT1K orthologs versus FishT1k orthologs' dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*.

Results

Transcriptome sequencing, assembly, and putative annotations

I obtained a raw sequence file of 101bp, paired-end reads totaling 202,309,724 reads and achieving a theoretical 6.8X coverage based on the size of the human genome from the University of Utah Sequencing Core. My collaborators at Scripps Institution of Oceanography assembled 143,060 Trinity-assembled scaffolds > 200 nt ($n_{50} = 2,212$), containing 59,412 predicted protein sequences > 99 aa (Table 3), wherein TransDecoder (Haas et al., 2013) classified 33,587 of these predicted proteins as "complete," with the remainder identified as partial5', partial3', or internal. My collaborators from Scripps provided an annotation file that includes predicted peptide sizes, CDS coordinates, and matches to the SwissProt, Pfam, EggNog, and KEGG databases. A quality assessment program called CEGMA found complete versions of 241/248 core, conserved eukaryotic genes (97%). The remaining seven core genes were found in partial form. This assembly had 260 matches, 167 of which were classified as complete by the gene-calling software.

Table 3. Results from University of Utah Sequencing Core (sequencing) and Scripps Institution of Oceanography (assembly and annotation).

<i>S. Spinus File</i>	# of Sequences	Sequence Length		Notes
Raw reads from Utah Sequencing Core	202,309,724	101	-	101bp paired-end
Trinity contigs from Scripps	59,412	297	15,120	Coding sequence (CDS)
	59,412	99	5,040	AA peptide sequence (pep)
	143,060	201	16,054	Scaffold (scf)

Gene tree

The tree results included three XME ML trees: CYPs, GSTs, and ABC transporters. A total of 44 individual putative *S. spinus* XME-genes were phylogenetically analyzed: 27 CYPs, 7 GSTs, and 10 ABC transporters (Table 1). These putative XME annotations were verified and refined with the tree analysis. The CYPs were divided into 6 families: 2 CYP1, 14 CYP2, 7 CYP3, 1 CYP4, 1 CYP20, and 2 CYP26. The GSTs were divided into 4 classes: 2 GSTA, 3 GSTM, 1 GSTO, and 1 GSTT. The ABC transporters were divided into 5 subfamilies: 4 ABCA, 1 ABCB, 2 ABCC, 1 ABCD, and 1 ABCE.

Gene Tree – Cytochrome P450

CYP1. Two mRNA sequences were identified as putative CYP1 genes, including homologs of CYP1A1 and CYP1B1. The phylogenetic analysis strongly supported the annotation of both putative *S. spinus* CYP1 genes. The putative *S. spinus* CYP1A1 and CYP1B1 clustered with its respective homologs from the other model species with perfect bootstrap values of 100.

CYP2. Fourteen mRNA sequences were identified as putative CYP2 genes, including homologs of CYP2C16, CYP2D10, CYP2D20, CYP2D28, CYP2F2, CYP2F3, CYP2F5, CYP2J1, CYP2J2, CYP2J5, CYP2J6, CYP2K1, and CYP2K3. The phylogenetic analysis supported the annotation of half of the putative *S. spinus* CYP2 genes. The putative *S. spinus* CYP2F5 clustered within the CYP2F/2X clade with a perfect bootstrap of 100. The putative *S. spinus* CYP2J1 and 2J2 are in a larger clade that contains paraphyletic groups of CYP2V/2AD/2P subfamilies with a bootstrap value of 97. The putative *S. spinus* CYP2K1internal and CYP2K3 formed their own cluster

within the CYP2K clade with a high bootstrap value of 96. The putative *S. spinus* CYP2F2 and 2F3 clustered within the CYP2F/ 2Y clade with a high bootstrap value of 97.

I found tremendous inconsistencies throughout the CYP2 clade indicating multiple potentially erroneous annotations. The *S. spinus* CYP2C16 gene was found in the CYP2F/2Y clade. The *S. spinus* CYP2K1 5' partial gene was found in the CYP2AD clade. The *S. spinus* CYP2J5, 2J6, and 2D10 genes were found in the CYP2F/2X clade. The *S. spinus* CYP2D20 gene was found in the CYP2J/2P clade. The *S. spinus* CYP2D28 gene was found in the CYP2J clade. Finally, there were multiple CYP2F clades found on the tree.

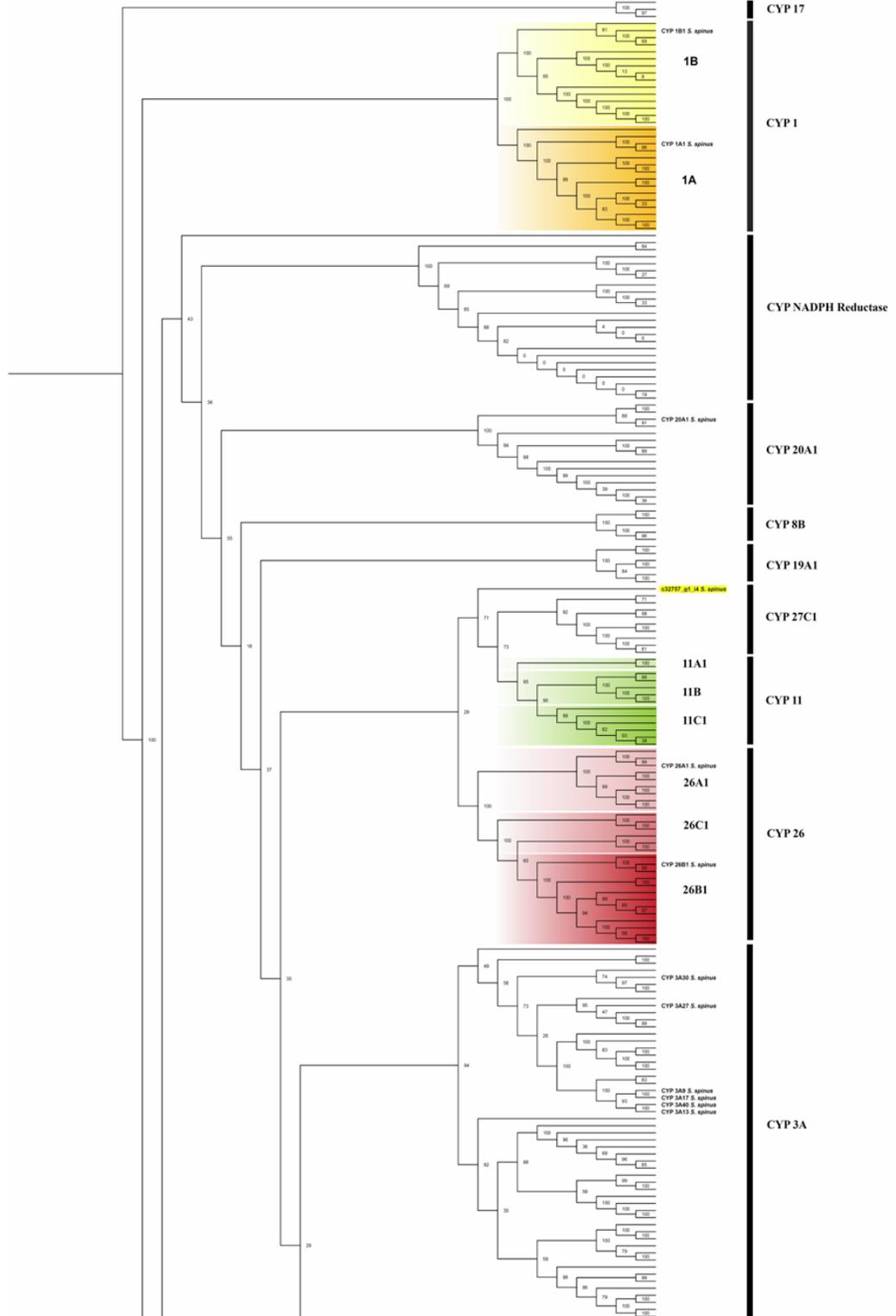
CYP3. Seven mRNA sequences were identified as putative CYP3A genes, including homologs of CYP3A4, CYP3A9, CYP3A13, CYP3A17, CYP3A27, CYP3A30, and CYP3A40. The phylogenetic analysis supported all annotations of *S. spinus* CYP3A genes as it clusters within the CYP3A clade with an overall bootstrap value of 94 with the exception of *S. spinus* CYP3A4. The *S. spinus* CYP3A4 gene may be an erroneous annotation because it clusters within the CYP27C1 and CYP11A/B/C1 clade with a bootstrap value of 71.

CYP4. One mRNA sequence was identified as a putative CYP4B1 gene. The phylogenetic analysis strongly supported the annotation of *S. spinus* CYP4B1, which clustered with its respective homologs from the other model species with a bootstrap value of 100.

CYP20. One mRNA sequence was identified as a putative CYP20A1 gene. The phylogenetic analysis strongly supported the annotation of *S. spinus* CYP20A1 gene.

The *S. spinus* CYP20A1 clustered with its respective homologs from the other model species with a bootstrap value of 88.

CYP26. Two mRNA sequences were identified as putative CYP26 genes, including homologs of CYP26A1 and CYP26B1. The phylogenetic analysis strongly supported the annotation of *S. spinus* CYP26 genes. The *S. spinus* CYP26A1 and CYP26B1 clustered with its respective homologs from the other model species with bootstrap values of 100 (Figure 4).



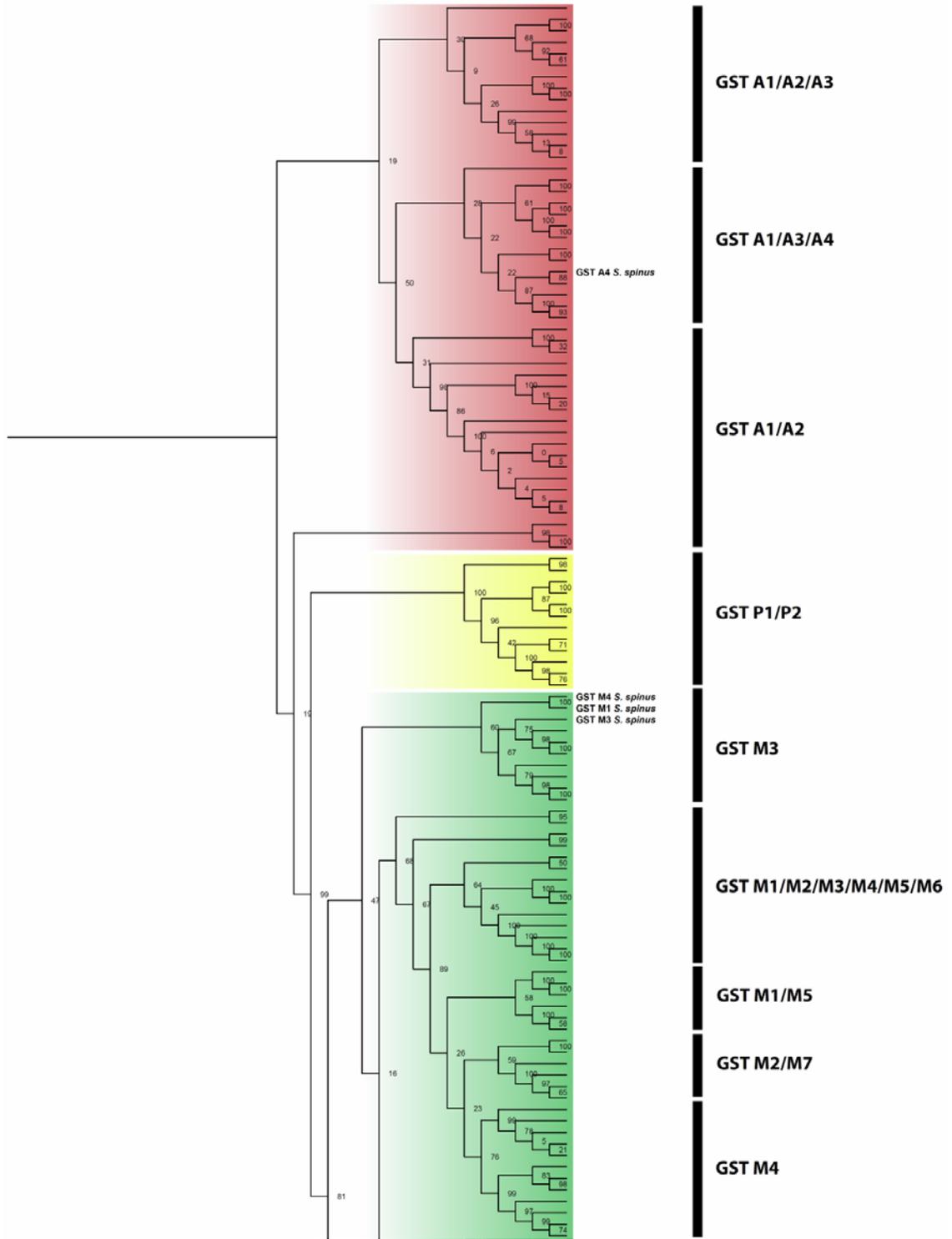
Gene Tree – Glutathione S-Transferase

GSTA. Two mRNA sequences were identified as putative *GSTA* genes, including homologs of *GSTA* and *GSTA4*. The putative *S. spinus* *GSTA4* clustered with its respective homologs in the *GSTA4* clade. The putative *S. spinus* *GSTA* clustered with the rest of the *GSTA* homologs with a highly supported bootstrap value of 99.

GSTM. Three mRNA sequences were identified as putative *GSTM* genes, including homologs of *GSTM1*, *GSTM3*, and *GSTM4*. The phylogenetic analysis supported the annotation of the *S. spinus* *GSTM* genes, which clustered with the model species in the overall *GSTM* clade.

GSTO. One mRNA sequence was identified as a putative *GSTO* gene, *GSTO1*. The phylogenetic analysis strongly supported the annotation of *S. spinus* *GSTO1*, which clustered with its respective homologs from the other model species with a bootstrap value of 100.

GSTT. One mRNA sequence was identified as a putative *GSTT* gene, *GSTT1*. The phylogenetic analysis strongly supported the annotation of *S. spinus* *GSTT1*, which clustered with their respective homologs from the other model species in the *GSTT1/T2* clade with a bootstrap value of 98 (Figure 5).



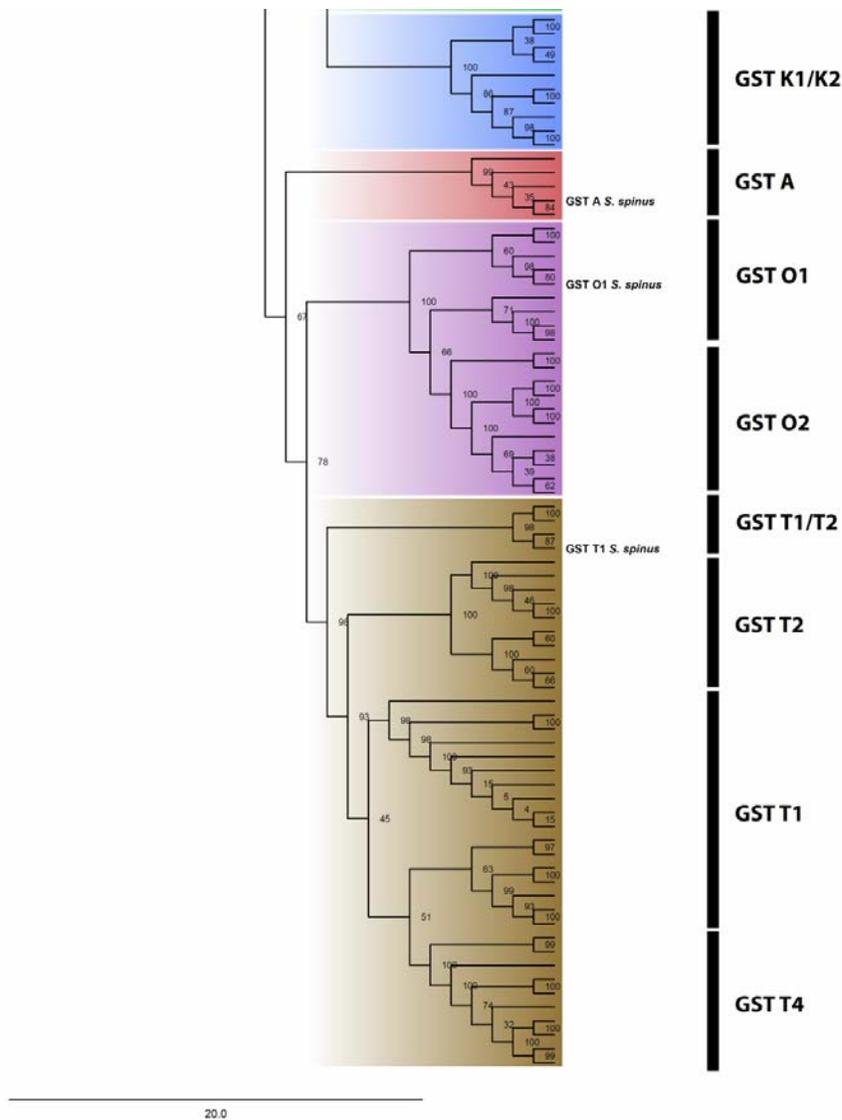


Figure 5. Maximum likelihood-based phylogenetic analysis of putative *S. spinus* liver glutathione S-transferase protein sequences. Numbers at nodes represent bootstrap support values. Different colors represent the various GST families: red – GSTA, yellow – GSTP, green – GSTM, blue – GSTK, purple – GSTO, and brown – GSTT. Only *S. spinus* GST names are listed. The GST names not shown are from the model organisms: *H. sapiens*, *M. musculus*, *G. gallus*, *D. rerio*, and *T. rubripes*.

Gene Tree – ATP-Binding Cassette Transporter

ABCA. Four mRNA sequences were identified as putative ABCA genes, including ABCA1, ABCA3, ABCA5, and ABCA12. The phylogenetic analysis supported the annotation of all four putative *S. spinus* ABCA genes, which clustered with their respective homologs from the other model species in the overall ABCA clade. Within their respective clades, the *S. spinus* ABCA12 had the strongest support (bootstrap value: 99), followed by ABCA5 (bootstrap value: 79), and ABCA3 (bootstrap value: 67). The *S. spinus* ABCA1 clustered with its respective homologs from the ABCA1 clade.

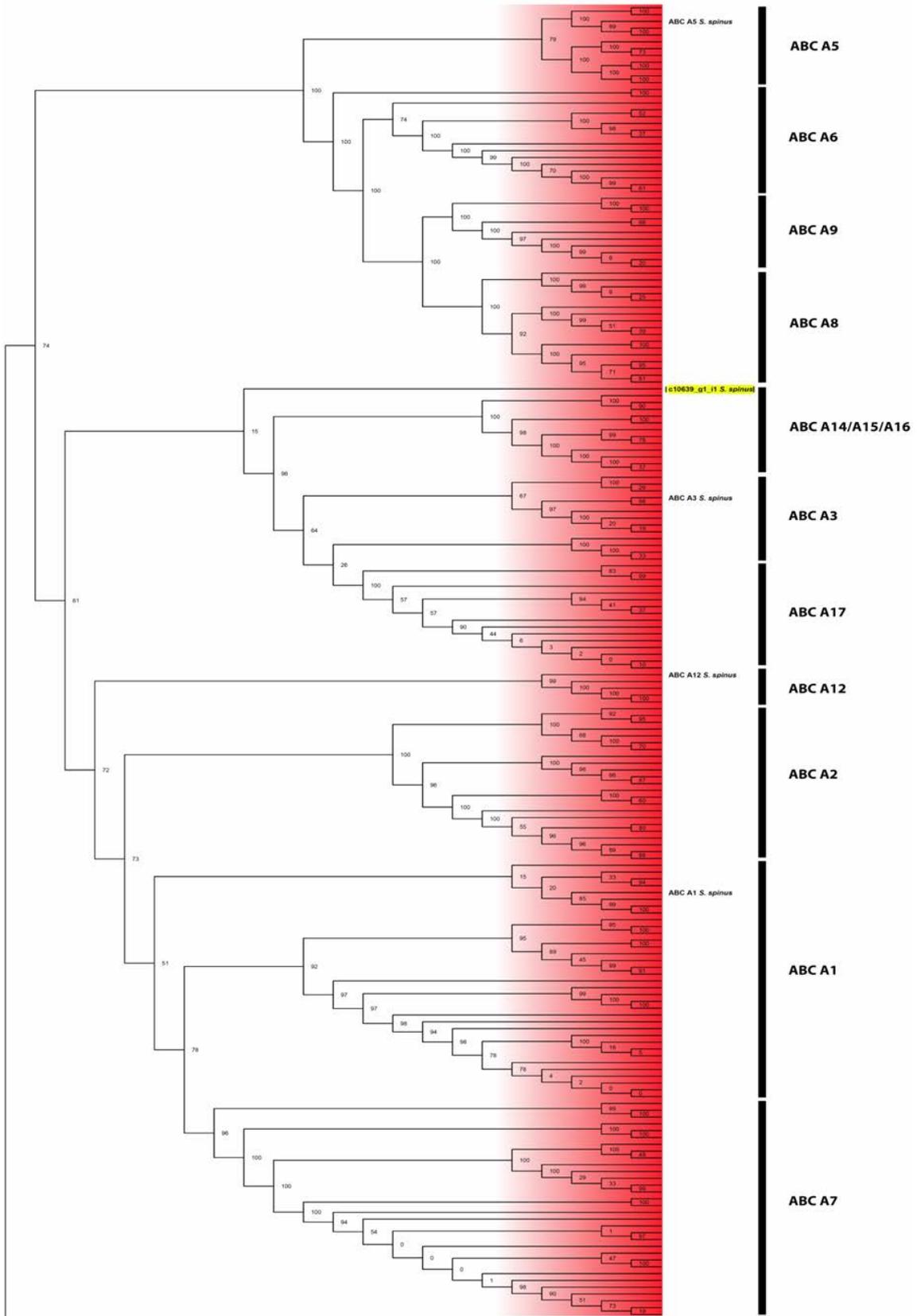
ABCB. One mRNA sequence was identified as a putative ABCB1 gene. The *S. spinus* ABCB1 gene clustered with the model species in the ABCB9 clade with a bootstrap value of 95.

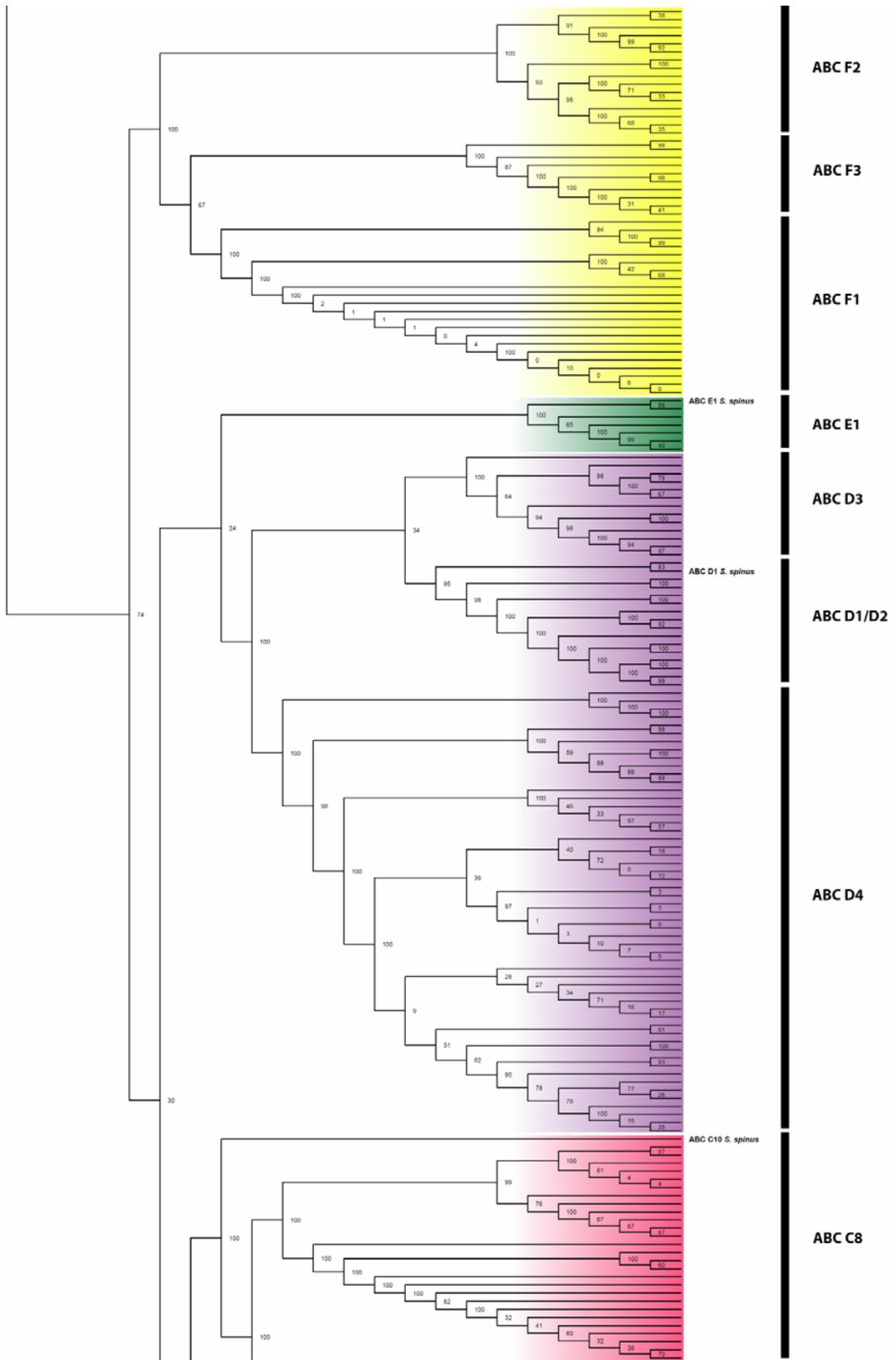
ABCC. Two mRNA sequences were identified as putative ABCC genes, including homologs ABCC9 and ABCC10. The phylogenetic analysis strongly supported the annotation of putative *S. spinus* ABCC genes. The putative *S. spinus* ABCC9 gene clustered with its respective homologs from the other model species with a bootstrap value of 100. The putative *S. spinus* ABCC10 gene clustered with the entire ABCC clade with a bootstrap value of 100.

ABCD. One mRNA sequence was identified as a putative ABCD1 gene. The phylogenetic analysis strongly supported the annotation of putative *S. spinus* ABCD1, which clustered with their respective homologs from the other model species in the ABCD1/2 clade with a bootstrap value of 95.

ABCE. One mRNA sequence was identified as a putative ABCE1 gene. The phylogenetic analysis strongly supported the annotation of putative *S. spinus* ABCE1, which clusters with its respective homologs with a bootstrap value of 100.

ABCG. One mRNA sequence was identified as a putative ABCG20 gene and is potentially erroneous based on the phylogenetic analysis. The *S. spinus* ABCG20 gene clustered with the ABCA14/15/16 (Figure 6).





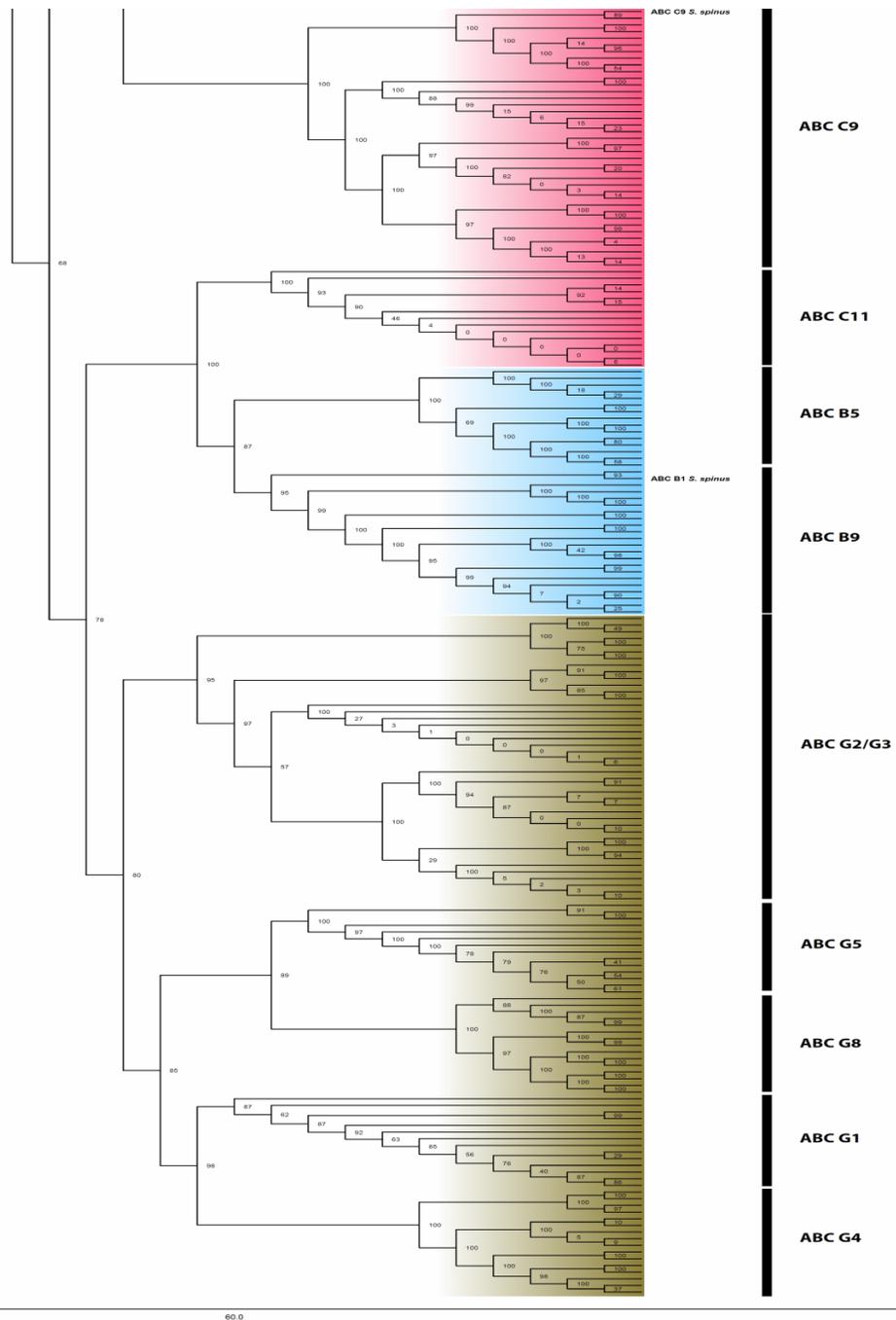


Figure 6. Maximum likelihood-based phylogenetic analysis of putative *S. spinus* liver ATP-binding cassette transporter protein sequences. Numbers at nodes represent bootstrap support values. Different colors represent the various ABC families: red – ABCA, yellow – ABCF, green – ABCE, purple – ABCD, pink – ABCC, blue – ABCB, and brown – ABCG. Only *S. spinus* ABC names are listed. The ABC names not shown are from the model organisms: *H. sapiens*, *M. musculus*, *G. gallus*, *D. rerio*, and *T. rubripes*. The highlighted annotation was considered to be an erroneous annotation.

Positive selection

Two separate analyses to identify signatures of positive selection in XMEs were conducted across the full-length of sequence and from substrings created from the sliding window. Positive selection across the full-length (whole protein) sequences for each XME was estimated from pairwise comparisons within each of two different datasets – ‘*S. spinus* versus FishT1K orthologs’ and ‘FishT1K orthologs versus FishT1K orthologs’ (Figure 7 & 8). The Ka/Ks ratios calculated from each pairwise comparison within each dataset were illustrated using box plot graphs (Figure 7 & 8), and then the distributions of each XME were compared using the Wilcoxon Rank Sum Test. Eleven XMEs showed their respective distributions to be significantly different from each other based on the Wilcoxon Rank Sum Test. None of the 64 XMEs from the ‘*S. spinus* versus FishT1K orthologs’ dataset had Ka/Ks ratios greater than or equal to 1 using this full-length analysis technique (Figure 7 & 8; Table 4 & 5).

In contrast, positive selection estimations from the sliding window analysis for both datasets found that 20 of the 64 XMEs exhibited signs of positive selection based on median Ka/Ks values ($Ka/Ks \geq 1$) (Figure 9; Table 4 & 5). Moreover, sliding window Ka/Ks ratios for 4 XMEs, CYP3A40 (Phase I), GSTA4 (Phase II), MDR1 (Phase III), and ABCC9 (Phase III) were greater than or equal to 2 (Figures 10, 11, 12, 13; Table 4) in both datasets (‘*S. spinus* versus FishT1K orthologs’ and ‘FishT1K orthologs versus FishT1K orthologs’). Furthermore, I was able to identify regions of selection within putative *S. spinus* GSTA4 and GSTM1 that showed signs of positive selection significantly (95% confidence) different from those of the FishT1k orthologs (Figures 13, 14; Table 3). Glutathione S-transferase A4 and M1 both had signatures of positive selection in the C-terminal alpha helical domain. For GSTA4, this region also comprised

of a dimer interface, substrate binding pocket (H-site), and parts of the N-terminal domain interface.

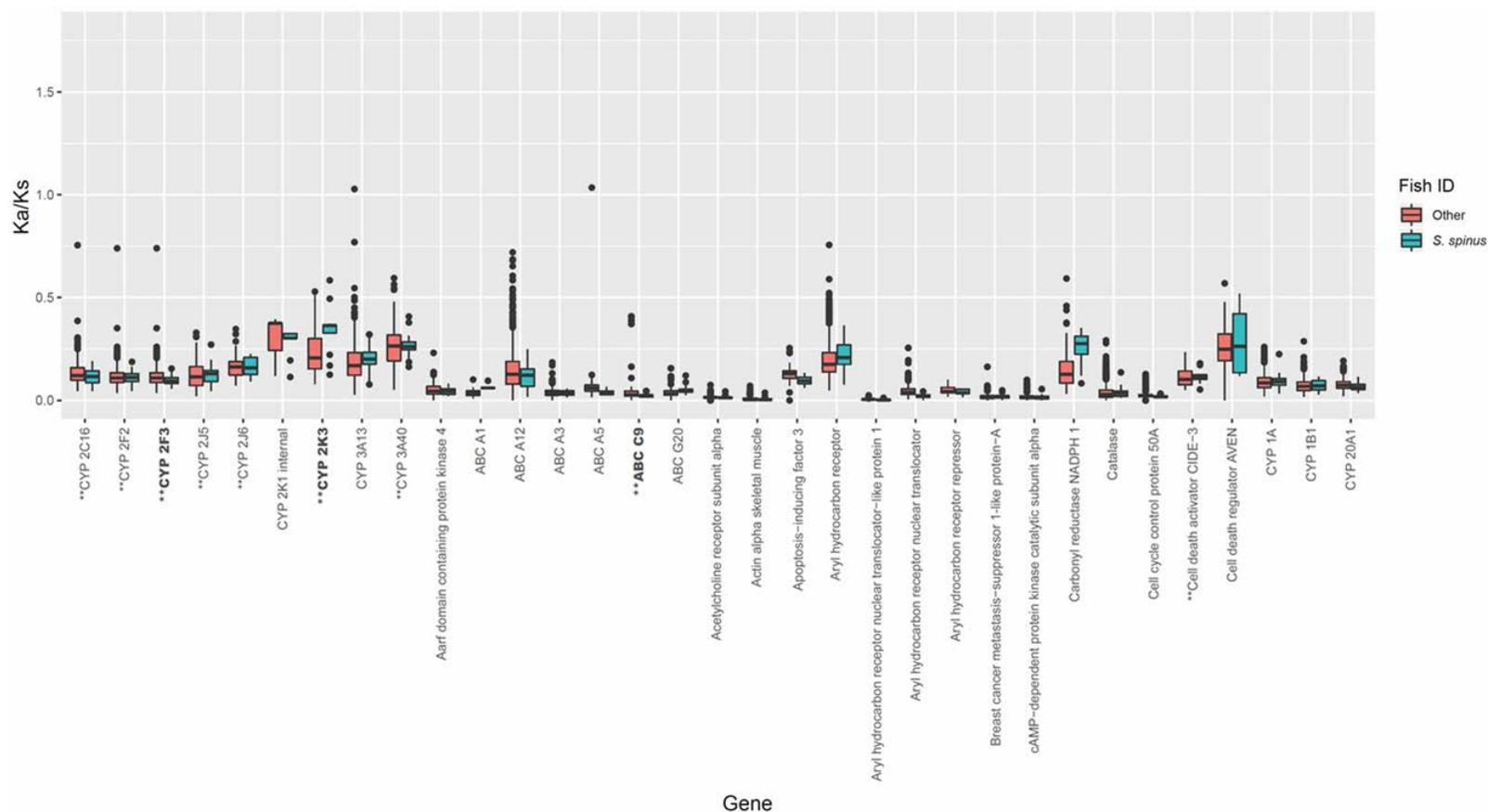


Figure 7. Box-plot graphs reflect the results from the full-length (whole protein) analysis using the gMYN method. In this analysis, 64 putative *S. spinus* XMEs were compared with other fish from the Fish-T1k database in a pairwise fashion for each respective gene. The distributions for each gene were tested by the Wilcoxon Rank Sum Test to determine if they were significantly different. Genes in bold reflect that the distributions had significant differences. Genes with a double asterisk (**) were found to have Ka/Ks ratios greater than or equal to 1 in the sliding window analysis. The dots represent the Ka/Ks ratio for each pairwise comparison.

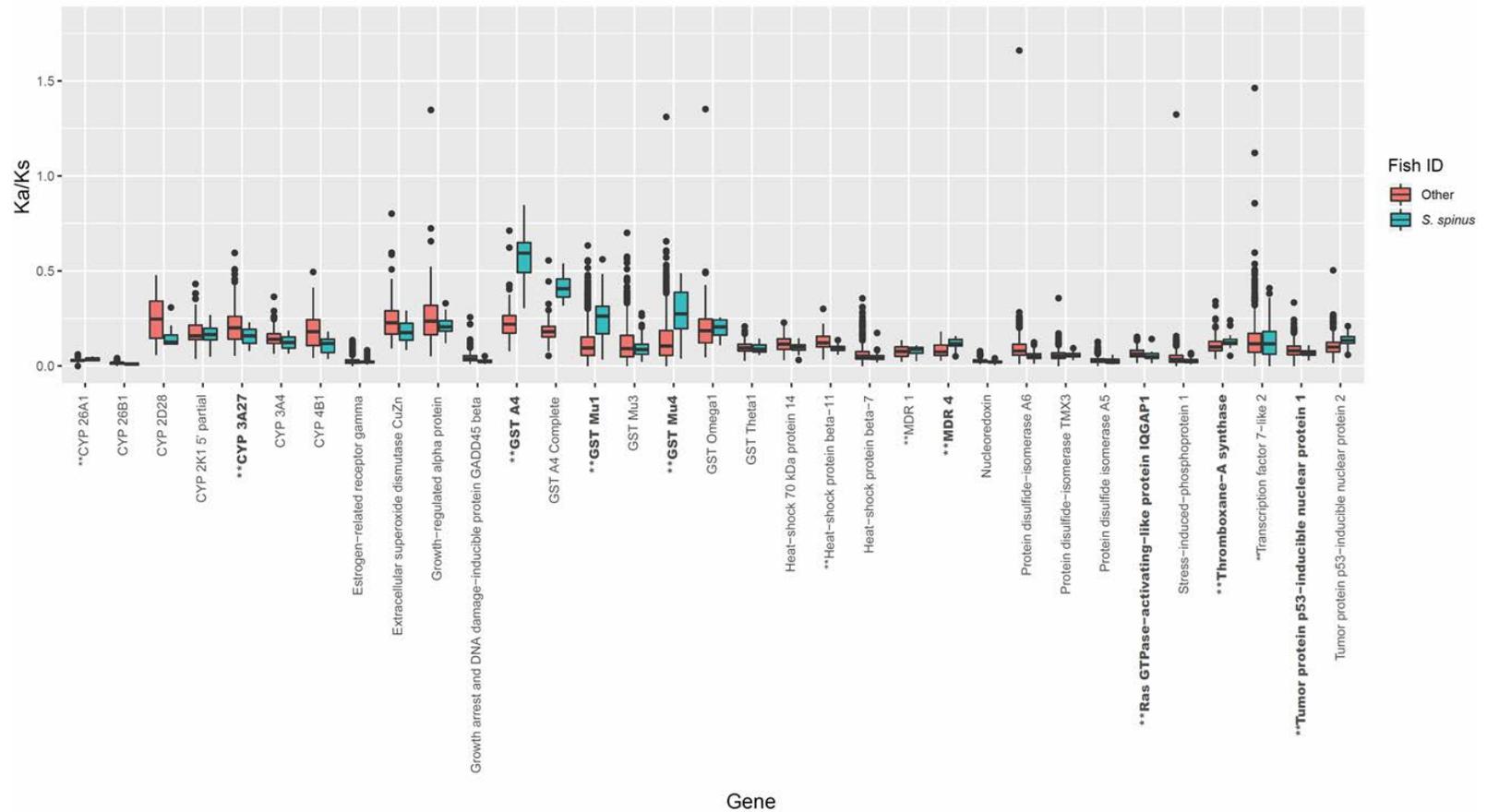


Figure 8. Box-plot graphs reflect the results from the full-length (whole protein) analysis using the gMYN method. In this analysis, 64 putative *S. spinus* XMEs were compared with other fish from the Fish-T1k database in a pairwise fashion for each respective gene. The distributions for each gene were tested by the Wilcoxon Rank Sum Test to determine if they were significantly different. Genes in bold reflect that the distributions had significant differences. Genes with a double asterisk (**) were found to have Ka/Ks ratios greater than or equal to 1 in the sliding window analysis. The dots represent the Ka/Ks ratio for each pairwise comparison.

Table 4. Summary of *S. spinus* XMEs used in whole protein and sliding window positive selection analyses. XMEs highlighted in red were removed from the positive selection analyses due to erroneous annotation results found in the phylogenetic analysis.

<i>S. spinus</i> XMEs used in Whole Protein and Sliding Window Positive Selection Analyses					
	Gene	Significantly different from fish (Wilcoxon Rank Sum Test)	Ka/Ks \geq 1 (Sliding Window Analysis)	Ka/Ks \geq 2 Across All Fish (Sliding Window Analysis)	Ka/Ks \geq 1 Specific to <i>S. spinus</i> (Sliding Window Analysis)
CYPs	1A1				
	1B1				
	2D28				
	2C16		X		
	2F2		X		
	2F3	X	X		
	2J5		X		
	2J6		X		
	2K1 Internal				
	2K1 5' partial				
	2K3	X	X		
	3A4				
	3A13				
	3A27	X	X		
	3A40		X	X	
	4B1				
	20A1				
	26A1			X	
	26B1				
	Thromboxane-A Synthase	X	X		
GSTs	A4	X	X		X
	M1	X	X		X
	M3				
	M4	X	X		
	O1				
	T1				
ABCs	A1				
	A12				
	A3				
	A5				
	C9	X	X	X	
	G20				
	MDR1		X	X	
MDR4	X	X			

Table 5. Summary of *S. spinus* miscellaneous response genes used in whole protein and sliding window positive selection analyses.

<i>S. spinus</i> XMEs used in Whole Protein and Sliding Window Positive Selection Analyses					
	Gene	Significantly different from fish (Wilcoxon Rank Sum Test)	Ka/Ks \geq 1 (Sliding Window Analysis)	Ka/Ks \geq 2 Across All Fish (Sliding Window Analysis)	Ka/Ks \geq 1 Specific to <i>S. spinus</i> (Sliding Window Analysis)
Miscellaneous Response Genes	AARF domain containing protein kinase 4				
	Acetylcholine receptor subunit alpha				
	Actin alpha skeletal muscle				
	Apoptosis-inducing factor 3				
	Aryl hydrocarbon receptor				
	Aryl hydrocarbon receptor nuclear translocator-like protein 1				
	Aryl hydrocarbon receptor nuclear translocator				
	Aryl hydrocarbon receptor repressor				
	Breast cancer metastasis-suppressor 1 like protein -A				
	cAMP-dependent protein kinase catalytic subunit alpha				
	Carbonyl reductase NADPH 1				
	Catalase				
	Cell cycle control protein 50A				
	Cell death activator CIDE-3			X	
	Cell death regulator AVEN				
	Estrogen-related receptor gamma				
	Extracellular superoxide dismutase CuZn				
	Growth-regulated alpha protein				
	Growth arrest and DNA damage-inducible protein GADD45 beta				
	Heat-shock 70 kDa protein 14				
	Heat-shock protein beta-11			X	
	Heat-shock protein beta-7				
	Nucleoredoxin				
	Protein disulfide-isomerase A5				
	Protein disulfide-isomerase A6				
	Protein disulfide-isomerase TMX3				
	Ras GTPase-activating-like protein (IQGAP1)	X		X	
	Stress-induced-phosphoprotein 1				
Transcription factor 7-like 2			X		
Tumor protein p53-inducible nuclear protein 1	X		X		
Tumor protein p53-inducible nuclear protein 2					

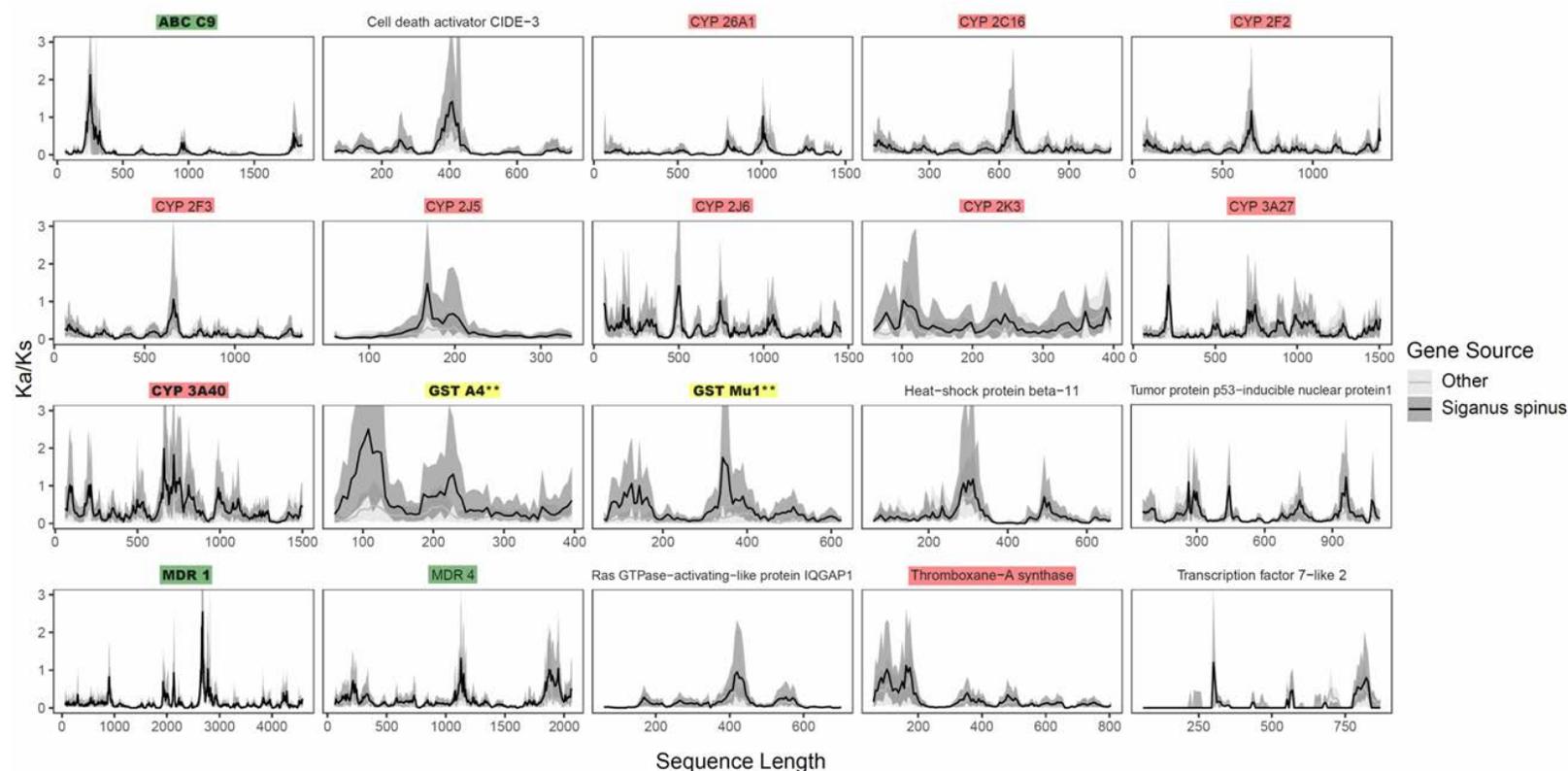


Figure 9. The sliding window analysis detected domains with Ka/Ks ratios greater than or equal to 1.0 within 20 putative *S. spinus* XMEs using the gMYN method. The XMEs in bold – CYP3A40, ABCC9, and MDR1 – showed Ka/Ks ratios greater than or equal to 2.0 across all fish. The XMEs in bold and with double asterisks (**) – GSTA4 and GSTM1 – showed Ka/Ks ratios that were specific to *S. spinus* with high confidence. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the ‘*S. spinus* versus FishT1K ortholog’ dataset and in the ‘FishT1K orthologs versus FishT1k orthologs’ dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The line and shaded areas represents the 95% confidence that the true median lies within that range/distribution. Different colors represent the XMEs from the three phases of xenobiotic metabolism: red – CYPs/Phase I, yellow – GSTs/Phase II, green – ABCs/Phase III.

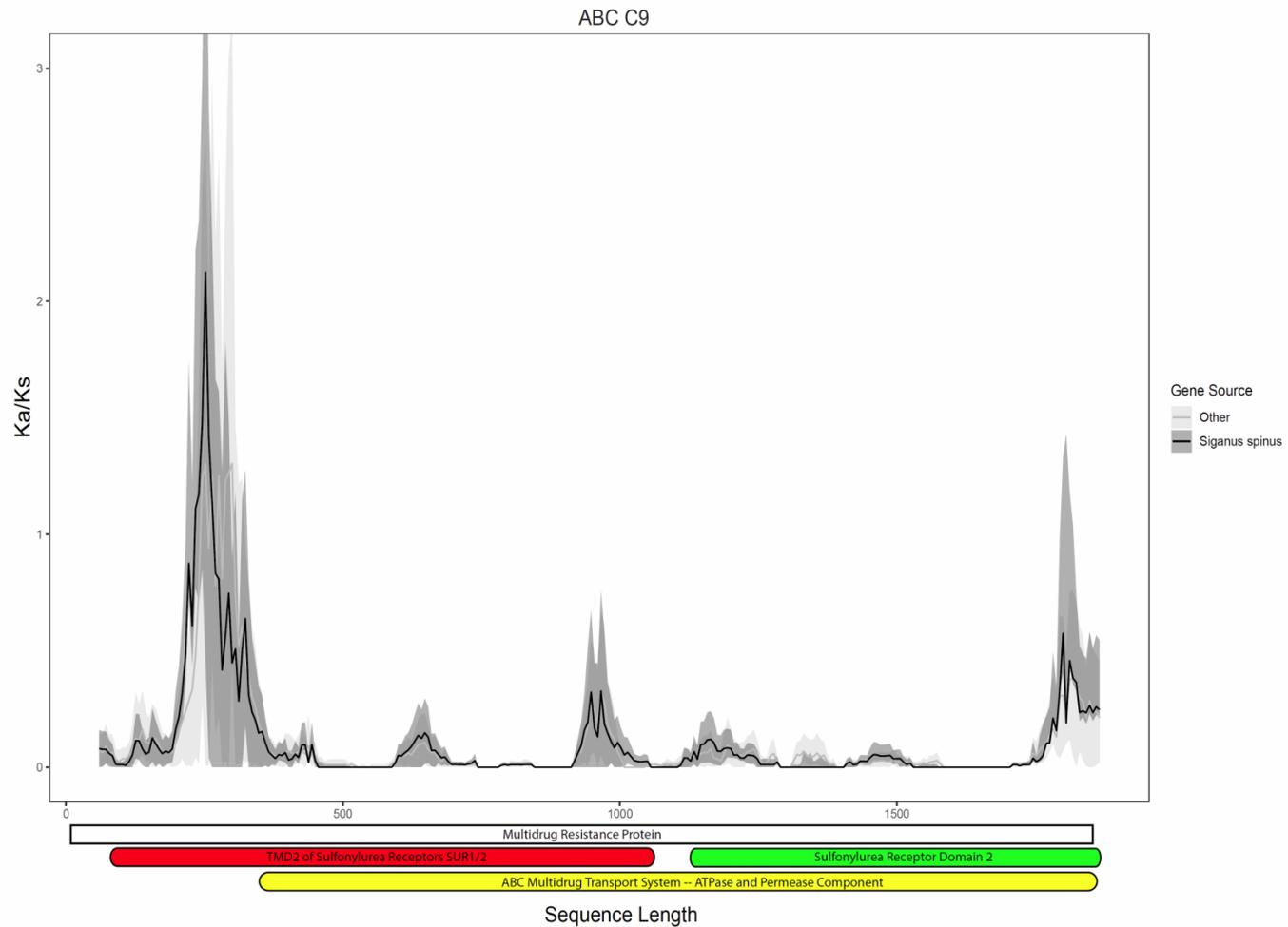


Figure 10. ABCC9 from the sliding window analysis. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the ‘*S. spinus* versus FishT1K ortholog’ dataset and in the ‘FishT1K orthologs versus FishT1k orthologs’ dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The shaded areas represents the 95% confidence that the true median lies within that range/distribution. Features below the graph illustrate conserved domains within the XME.

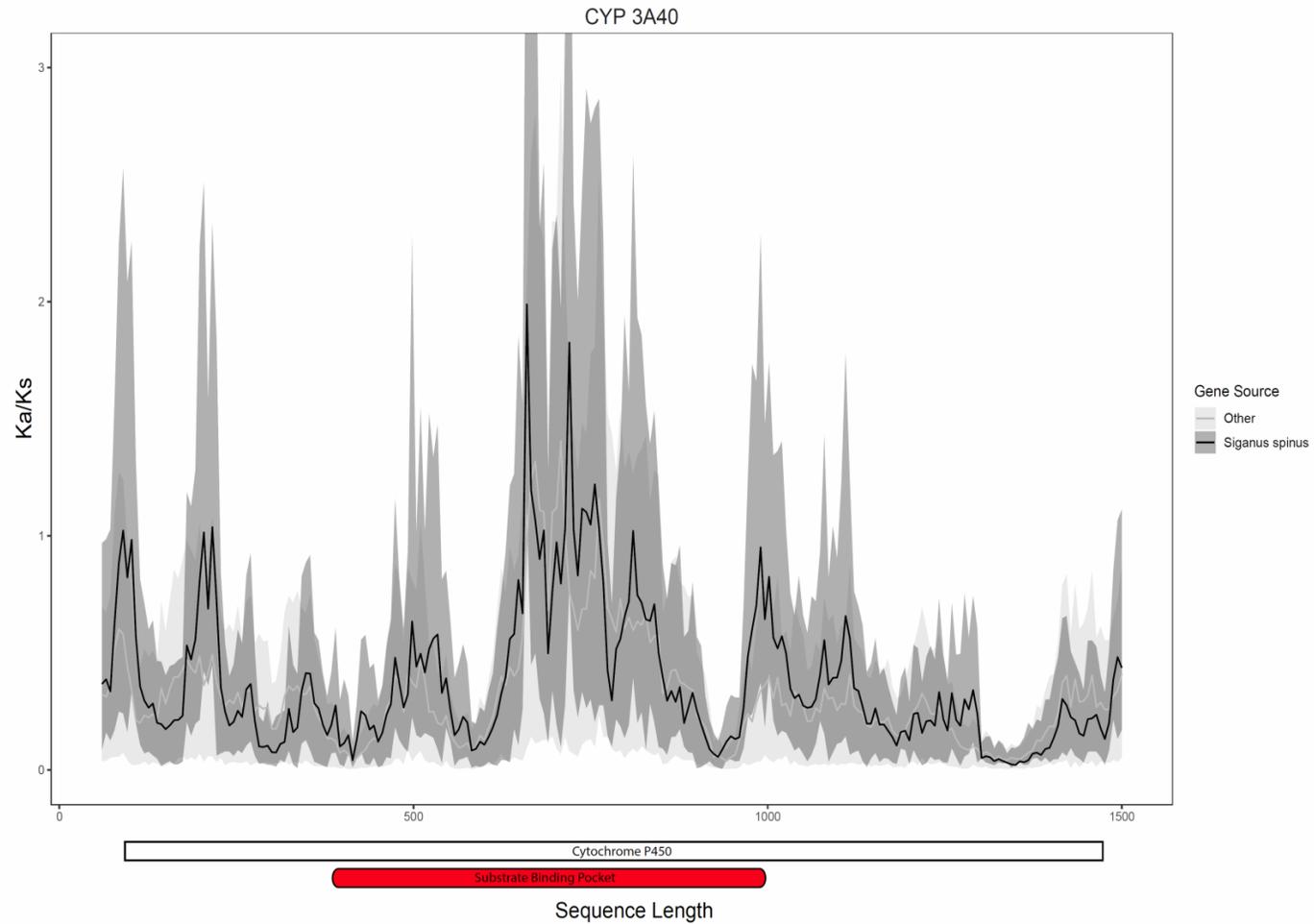


Figure 11. CYP3A40 from the sliding window analysis. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the ‘*S. spinus* versus FishT1K ortholog’ dataset and in the ‘FishT1K orthologs versus FishT1k orthologs’ dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The shaded areas represents the 95% confidence that the true median lies within that range/distribution. Features below the graph illustrate conserved domains within the XME.

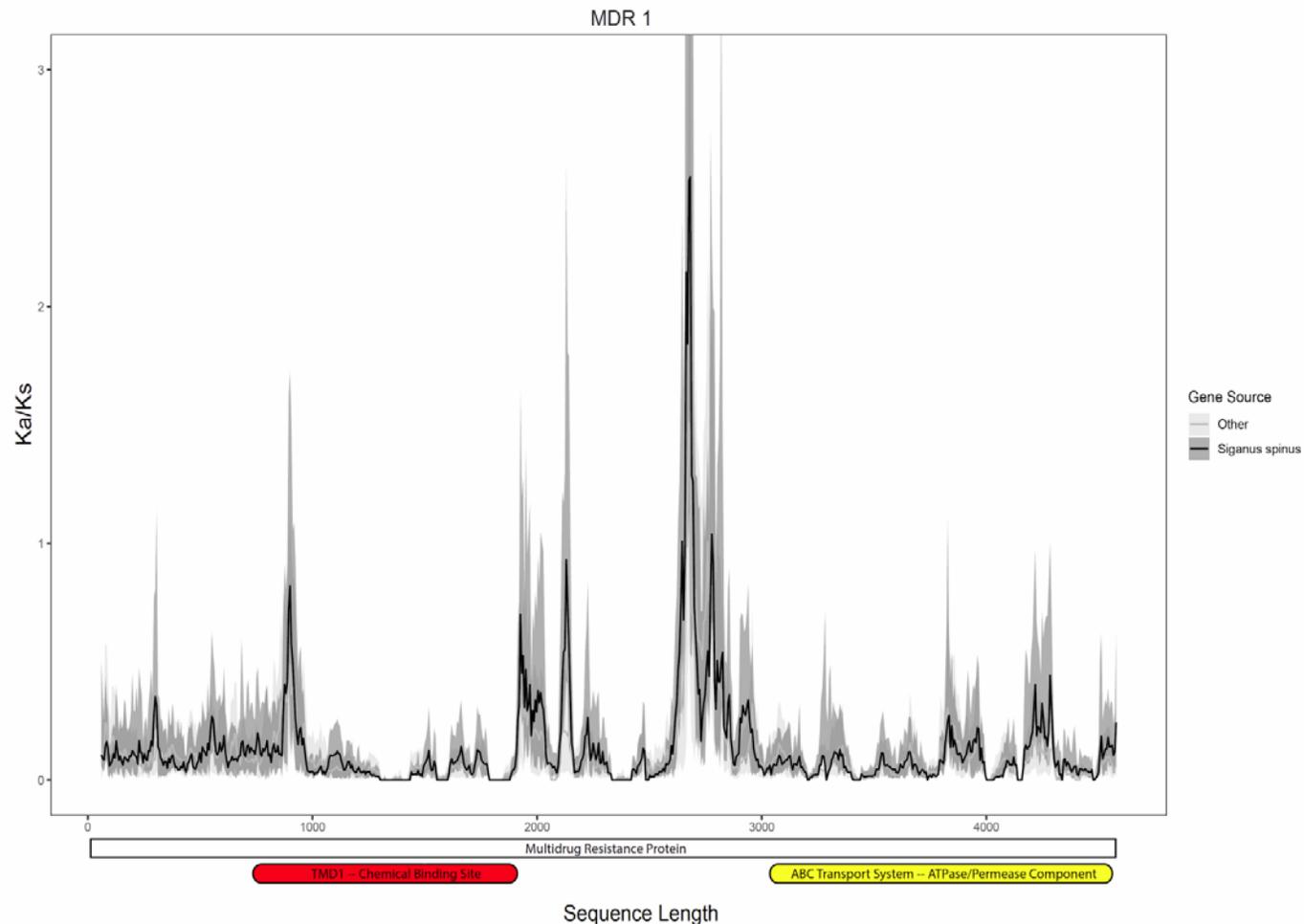


Figure 12. MDR1 from the sliding window analysis. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the ‘*S. spinus* versus FishT1K ortholog’ dataset and in the ‘FishT1K orthologs versus FishT1k orthologs’ dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The shaded areas represents the 95% confidence that the true median lies within that range/distribution. Features below the graph illustrate conserved domains within the XME.

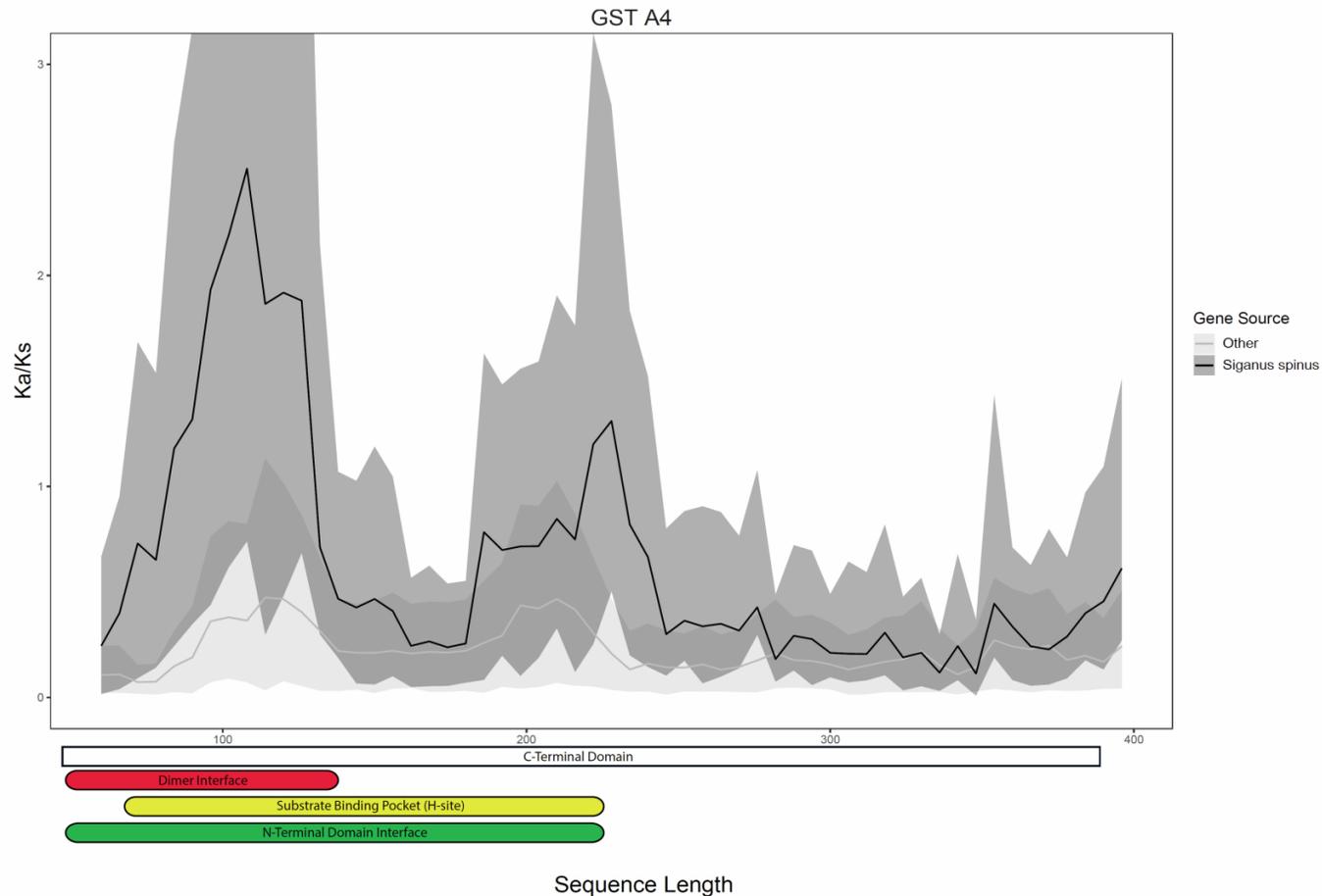


Figure 13. GSTA4 from the sliding window analysis. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the '*S. spinus* versus FishT1K ortholog' dataset and in the 'FishT1K orthologs versus FishT1k orthologs' dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The shaded areas represents the 95% confidence that the true median lies within that range/distribution. Features below the graph illustrate conserved domains within the XME.

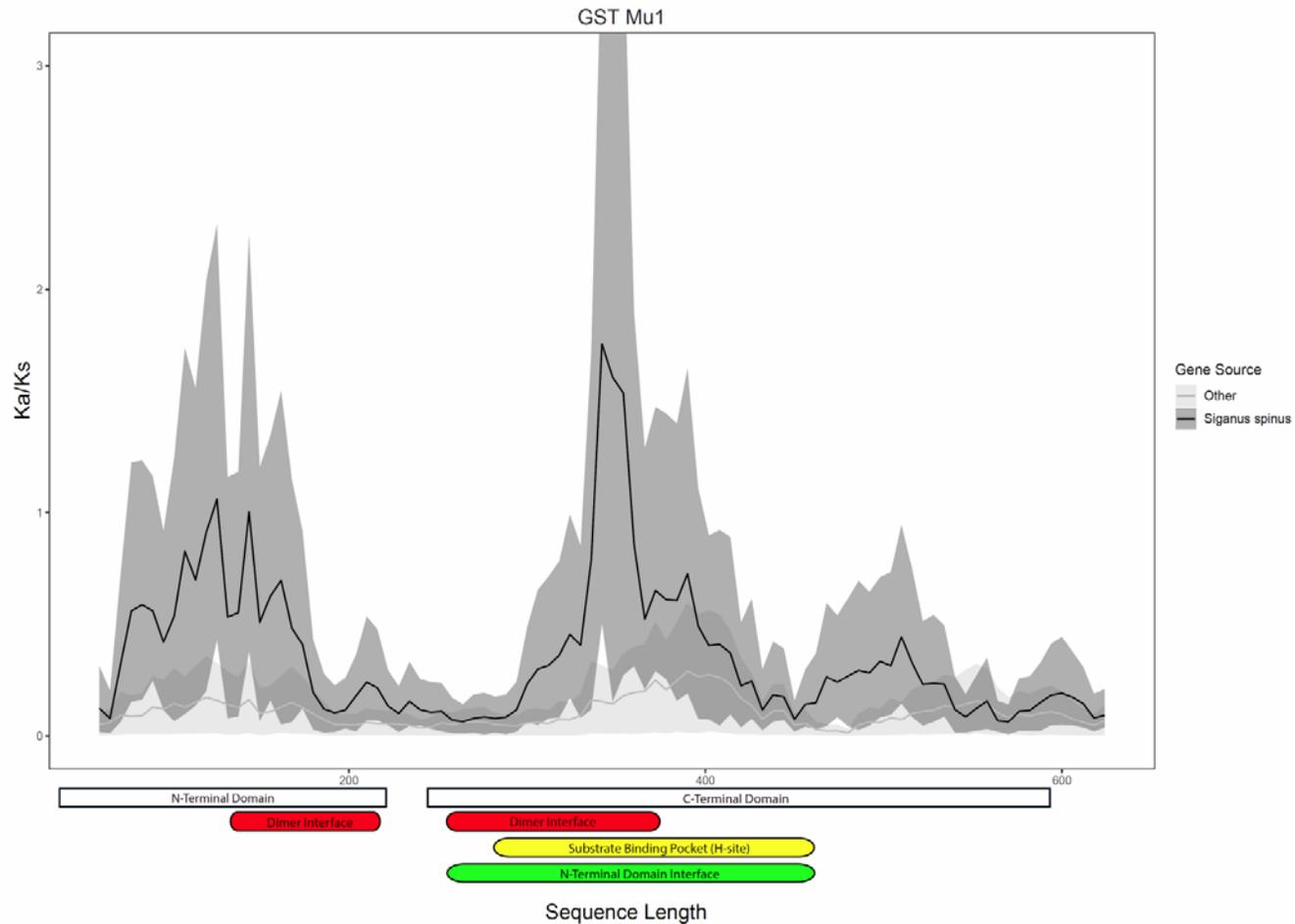


Figure 14. GSTM1 from the sliding window analysis. For each XME in the sliding window analysis, the 95% confidence intervals around median Ka/Ks ratios observed in each window from the ‘*S. spinus* versus FishT1K ortholog’ dataset and in the ‘FishT1K orthologs versus FishT1k orthologs’ dataset were estimated using bootstrapping. Ka/Ks ratios were resampled 10,000 times, yielding confidence intervals around the medians to determine if there were signatures of positive selection that occurred across all fish or specifically within *S. spinus*. The lines (black and white) represent the mean median values of each Ka/Ks ratio at each sliding window. The shaded area (gray and white) represent the medians at the upper and lower confidence intervals (the tails). The shaded areas represents the 95% confidence that the true median lies within that range/distribution. Features below the graph illustrate conserved domains within the XME.

Discussion

This study investigated the identity of putative *S. spinus* liver XMEs and potential selection signatures within their mRNA transcripts. My major results included the verification of most annotated *S. spinus* XMEs, identification of XMEs that show signatures of positive selection, and identification of specific domains within a subset of XMEs that show signatures of positive selection. This study provided evidence that putative *S. spinus* liver XME genes may be used to establish orthologous relationships with model organisms and other fish to potentially strengthen functional inferences of genes across the tree of teleosts. Finally, this study provided evidence for positive selection in XMEs that potentially underlie the ability of *S. spinus* to exploit ecological niches inaccessible to other marine herbivores.

Gene trees

Based on the gene family trees, 44 *S. spinus* XME annotations were verified. Forty-two formed highly supported clades within their respective gene superfamilies alongside homologs from model organisms, implying that putative *S. spinus* XMEs have conserved motifs and domains indicative of each gene superfamily, as well as similar activity and function to that of known isoforms. Two additional sequences (*i.e.*, GSTM1 and GSTM4) clustered with model fish sequences of the Glutathione-S-transferase Mu-superfamily, but not specifically within the isoform designations populated with model organisms. Such grouping suggests that labeling them as specific GSTM-isoforms could be a lesser error likened to those caused by cryptic diversity in species. However, clustering of the two sequences entirely among teleost-specific GSTM-genes warranted their inclusion in further comparisons. Finally, two others (*i.e.*, CYP3A4 and ABCG20) did not cluster with their respective homologs nor their respective XME-gene

superfamilies, suggesting that their annotations were completely erroneous and thus, these sequences were stripped of their annotations and removed prior to further analyses.

XME research in bony fishes has proven interesting due to gene duplication events early on in their separation from terrestrial vertebrates wherein fish frequently possess paralogous copies for many genes (Liu et al., 2013; Liu et al., 2016; Lee et al., 2018). Among XMEs, CYPs are among the most difficult to establish homology and function for (Yan and Cai, 2010; Goldstone et al., 2010; Uno et al., 2012; Zhang et al., 2014; Lee et al., 2018). For example, CYP2s are some of the most diverse CYPs that can metabolize both xenobiotics and endogenous molecules, yet information regarding their catalytic functions, biological roles, and gene-regulation pathways is virtually non-existent (Goldstone et al., 2010; Lee et al., 2018). Furthermore, CYP2s show the largest degree of divergence across fish species (Uno et al., 2012) and homology with other vertebrates is difficult to establish due to these high levels of sequence diversity and divergence (Goldstone et al., 2010). This CYP2 diversity was evident in my *S. spinus* CYP tree, echoing the complexities involved in studying teleost XME-gene evolution (Figure 4). Nonetheless, these results provide an opportunity to establish orthologous relationships among *S. spinus*, model organisms, and other fish species, thus enabling future gene characterization studies and functional inferences.

Positive selection

Twenty of the sixty-four putative XME genes expressed within the *S. spinus* transcriptome had Ka/Ks ratios greater than or equal to 1 across the length of their coding regions, suggesting that positive selection has acted broadly upon a variety of XME-gene families in teleost fishes (Figure 9; Table 4 & 5). Of these, all three phases of xenobiotic metabolism were represented, including sequences encoding CYPs (10), GSTs (2), and

ABCs (3), along with miscellaneous anabolic XMEs (5). Each of these gene families are discussed – both in terms of the selection markers discovered and their possible ecological significance – in greater detail below.

Half of the XMEs exhibiting K_a/K_s ratios greater than or equal to 1 across the length of their coding regions belong to CYPs. More specifically, 6 of the 10 CYP sequences encode CYP2 isoforms, and another two were CYP3A sequences, all eight of which localize to the endoplasmic reticulum (ER) and primarily act upon xenobiotics (Arellano-Aguilar, et al., 2009; Anzenbacher and Anzenbacherová, 2001). The remaining two encode mitochondrial CYP isoforms, which are generally involved in the creation and breakdown of endogenous molecules. Recall, that CYPs 1-4 are known to metabolize xenobiotics in terrestrial vertebrates and thus, they are highly diverse, multi-gene families in part due to the need to account for the pharmacopeia of exogenous compounds that have evolved within these environments (Uno et al., 2012; Zhang et al., 2014). In terms of human health, CYP3 isoforms are of paramount importance because they are the most abundant liver isoforms and metabolize a wide variety of compound classes (Uno et al., 2012). In fact, CYP3-catalyzed oxidation reactions are involved in the removal pathways of ca. 75% of all drugs used today (Guengerich, 2008; Yan and Cai, 2010). Their biochemical functions in medaka (*O. latipes*) include aromatase reactions, reduction of methylene groups, and steroid hydroxylase activity (Marchler-Bauer et al., 2017). Given this breadth of substrates, it comes as no surprise to discover signs of positive selection in the *S. spinus* CYP2 and CYP3 genes which are indistinguishable from other fish sequences used in these analyses. Furthermore, the three-dimensional folding of these enzymes creates substrate binding interactions among

residues that are spread throughout their primary amino acid sequences rather than being concentrated in distinct motifs, which could cause CYP sequences to be more difficult to dissect in terms of attributing mechanistic outcome to specific amino acid substitutions. However, given their importance to drug metabolism in humans, selection signatures in *S. spinus* CYP2 and CYP3 gene sequences warrant further investigation.

Sliding window analysis increased the sensitivity of positive selection detected within five *S. spinus* XME-genes, all of which had one or more regions with $Ka/Ks \gg 1$, suggesting that selective pressures act upon specific sites of protein structure rather than evenly across the full length of these genes (Figures 10, 11, 12, 13, 14; Table 3) (Wang et al., 2009b; Wang et al., 2010). Three of these XMEs – CYP3A40, ABCC9, and MDR1 – showed signs of positive selection across all fish (Figure 10, 11, & 12), suggesting that they may play pivotal, yet more general roles in marine chemical ecology. Finally, there were two XMEs – GSTA4 and GSTM1 – that showed signs of positive selection specific to *S. spinus*. Most intriguingly, the specific GST domains that showed signs of selection were the substrate binding site in both cases (Figure 13 & 14).

Three members of the ABC superfamily – ABCC9, MDR1, and MDR4 – had Ka/Ks ratios greater than or equal to 1 across the length of their coding regions. It is important to note that of these three, ABCC9 homologs are not known for transporting xenobiotics. Instead, they act as sulfonyleurea receptors which regulate the activity of potassium-channels (Riordan et al., 1989; Hibino and Kurachi, 2006; Aleksandrov et al., 2007; Bryan et al., 2007; Luckenbach et al., 2014). According to UniProtKB, ABCC9's molecular functions in *D. rerio* include ATPase activity, ATPase-coupled transmembrane transporter activity, ATP binding, and sulfonyleurea-receptor activity. Sulfonyleurea-

receptor processes are often linked to the secretion of insulin, a peptide hormone which regulates blood sugar levels in vertebrates (Marchler-Bauer et al., 2017). Positive selection within the ABCC9-gene could possibly involve desensitization to xenobiotic action, thus preventing blood-sugar imbalance.

MDR1 and MDR4 were first recognized for their roles in impeding chemotherapeutic success of drugs used to cure cancer, which inadvertently has made them two of the most well characterized isoforms within what is arguably one of the largest, and possibly one of the oldest gene families characterized to date (Ferreira et al., 2014; Dean and Annilo, 2005). In fish, MDR1 has broad substrate specificity with one of its various roles being to remove toxic metabolites and xenobiotics from cells into urine, bile, and the intestinal lumen (Dean and Annilo, 2005; Liu et al., 2013). Yet, like CYPs, the three-dimensional folding of transmembrane transporters creates substrate binding interactions among residues that are spread relatively throughout their primary amino acid sequences. Both isoforms are known to transport therapeutic drugs and sometimes even serve as first lines of defense in cellular detoxification by preventing xenobiotics entering the cell from reaching biochemically relevant concentrations (Ferreira et al., 2014; Liu et al., 2016). Therefore, the existence of selection markers in these two phase III XMEs provides further evidence for the role of XME evolution in determining dietary breadth of *S. spinus*.

GST isoforms A4 and M1 were the two XMEs which showed signs of selection specific to *S. spinus* (Table 4). GSTs aid in the excretion of xenobiotics, phase I metabolites, and other electrophilic compounds by catalyzing their conjugation to glutathione (GSH) rendering them more hydrophilic, thus preparing them for excretion

from the cell and disabling their ability to cross cellular membranes through passive diffusion (Hayes et al., 2005; Donham et al., 2005a; Donham et al., 2005b; Sotka and Whalen, 2008; Kolawole, 2016). GSTs are relatively small XMEs (200-300 residues) that share common folds – an N-terminal thioredoxin-fold (GSH binding domain) and a C-terminal alpha helical domain (xenobiotic/hydrophobic substrate binding domain) – with an active site located in a cleft between the two domains (Reinemer et al., 1991; Blanchette et al., 2007; Kolawole 2016). GSTs can form homo- and heterodimers wherein a single isozyme can catalyze reactions among multiple types of substrates that are not isoform specific, but rather overlap the specificities of each isoform (Kolawole 2016). Known substrates of GSTs include chemotherapeutic agents, insecticides, carcinogens, environmental pollutants, oxidative stress by-products, and marine natural products (Sheehan et al., 2001; Enayati et al., 2005; Hayes et al., 2005; Li et al., 2007; Sotka and Whalen, 2008). My major finding is that both, GSTA4 and GSTM1, have relatively high sliding-window K_a/K_s values (≥ 2), within their respective substrate binding domains (Figure 13 & 14), wherein the hydrophobic substrate (*i.e.*, the xenobiotic) is positioned for further modification (Reinemer et al., 1991; Blanchette et al., 2007; Kolawole, 2016). Selection markers within in the binding domain suggests that *S. spinus* GSTs have great potential for conjugating secondary metabolites within their diets, including reactive intermediates created by CYPs during phase I reactions (Guengerich, 2003). These findings suggest that these GSTs may have evolved in ways that protect against compounds which become more reactive following CYP-catalyzed reactions, as could be the case if *S. spinus* were to expand its prey preference to include novel algal secondary metabolites.

This study builds upon the work of Emborski et al. (2012), who characterized CYP1A activity in *S. spinus* hepatic microsomes, and demonstrated the conservation of classical CYP1A induction mechanisms. My analyses showed no signs of positive selection in either CYP1A1 or CYP1B1, providing even more evidence that the ecological role of these two CYP enzymes also remains conserved. Furthermore, the discovery of only one CYP1A gene within *S. spinus* is consistent with the notion that fish generally possess only one CYP1A isoform, which appeared early in the evolution of vertebrates (Morrison et al., 1995; Morrison et al., 1998). This study also builds upon the work of Reyes (2017), who demonstrated CYP3A activity in *S. spinus* hepatic microsomes, and demonstrated the upregulation of both CYP1A1 and CYP3A4 in response to estrogenic compounds. Given that I have found selection markers distributed throughout the *S. spinus* CYP3A sequence, it is promising to note that this enzyme remains active.

In conclusion, this study further establishes *S. spinus* as a model species for studying the chemical ecology of marine plant-herbivore interactions by characterizing xenobiotic metabolizing enzyme expression using phylogenetic and positive selection analyses. Despite the major limitation of not conducting functional assays to validate *S. spinus* XME functions, each putative XME identified within our transcriptome is a potential resource for future gene characterization studies. More importantly, however, these sequences now provide the foundation for future ecological studies based on gene expression analyses. A hallmark of many XMEs is that they are induced by the very substrates which they metabolize, thus feeding experiments using crude algal extracts and their specific algal secondary metabolites have the potential to rapidly expand our

understanding of the specific functional roles each of these enzymes play, and the evolutionary relationships they share with model vertebrates and teleosts. This can also be expanded to include exposure to common organic pollutants found in the marine environment. Intracellular receptors (*e.g.* aryl hydrocarbon receptor-AHR and pregnane X receptor-PXR) from mammalian models have been shown to be key mediators in regulating the activity of phase I, II, and III XMEs (Xu et al., 2005; Ferreira et al., 2014). Despite the lack of knowledge for transcriptional regulation and functional inference in fish, studies conducted on zebrafish (*D. rerio*) showed an association between PXR, CYP3A, and MDR1 suggesting that this receptor triggers the coordinated efforts of CYP3A and MDR1 to detoxify and excrete foreign compounds rapidly and efficiently (Bresolin et al., 2005; Ferreira et al., 2014). Ultimately, it is quite interesting to find that positive selection seems to be acting upon these two *S. spinus* XMEs. Could it be that these two enzymes work synergistically to overcome the diversity of chemical defenses within this herbivore's diet? This study is the first step in answering that question.

Literature Cited

- Agrawal, A.A. (2004). Plant defense and density dependence in the population growth of herbivores. *The American Naturalist*, 164(1), 113-120.
- Aleksandrov, A.A., Aleksandrov, L.A., & Riordan, J.R. (2007). CFTR (ABCC7) is a hydrolysable ligand-gated channel. *Pflugers Archiv - European Journal Physiology*, 453(5), 693–702.
- Amsler, C.D., & Fairhead, V.A. (2006). Defensive and sensory chemical ecology of brown algae. *Advances in Botanical Research*, 43, 1-91.
- Anzenbacher, P., & Anzenbacherová, E. (2001). Cytochromes P450 and metabolism of xenobiotics. *Cellular and Molecular Life Sciences*, 58(5-6), 737-747.
- Arellano-Aguilar, O., Montero-Montoya, R., & Macías-García, C. (2009). Endogenous functions and expression of cytochrome P450 enzymes in teleost fish: A review. *Reviews in Fisheries Science*, 17(4), 541–556.
doi.org/10.1080/10641260903243487
- Blanchette, B., Feng, X., & Singh, B.R. (2007). Marine glutathione S-transferases. *Marine Biotechnology*, 9(5), 513-542.
- Brammell, B. F., McClain, J. S., Oris, J. T., Price, D. J., Birge, W. J., & Elskus, A. A. (2010). CYP1A expression in caged rainbow trout discriminates among sites with various degrees of polychlorinated biphenyl contamination. *Archives of Environmental Contamination and Toxicology*, 58(3), 772–782.
doi.org/10.1007/s00244-009-9368-x
- Bresolin, T., Rebelo, M.D., & Bainy, A.C.D. (2005). Expression of PXR, CYP3A and MDR1 genes in liver of zebrafish. *Comparative Biochemistry and Physiology – Part C: Toxicology*, 140(3-4), 403–407. doi: 10.1016/j.cca.2005.04.003
- Brunet, F.G., Crollius, H.R., Paris, M., Aury, J.M., & Gibert, P. (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Molecular Biology and Evolution*, 23(9), 1808–1816.
- Bryan, J., Muñoz, A., Zhang, X., Düfer, M., Drews, G., Krippeit-Drews, P., & Aguilar-Bryan, L. (2007). ABCC8 and ABCC9: ABC transporters that regulate K⁺ channels. *Pflugers Archiv - European Journal Physiology*, 453(5), 703–718.
- Buhler, D. R., & Wang-Buhler, J.L. (1998). Rainbow trout cytochrome P450s: purification, molecular aspects, metabolic activity, induction and role in environmental monitoring. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, 121(1–3), 107–137.
doi.org/10.1016/S0742-8413(98)10033-6

- Capella-Gutierrez, S., Silla-Martinez, J.M., & Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetics analyses (Version 1.2b).
- Carvan III, M. J., Incardona, J. P., & Rise, M. L. (2008). Meeting the challenges of aquatic vertebrate ecotoxicology. *BioScience*, 58(11), 1015–1025.
doi.org/doi:10.1641/B581105
- Casarett, L. J., Klaassen, C. D., & Watkins, J. B. (2003). Casarett and Doull's essentials of toxicology. New York: *McGraw-Hill*.
- Chang, G. W. M., & Kam, P. C. A. (1999). The physiological and pharmacological roles of cytochrome P450 isoenzymes. *Anaesthesia*, 54(1), 42–50.
doi.org/10.1046/j.1365-2044.1999.00602.x
- Connon, R. E., D'Abronzio, L. S., Hostetter, N. J., Javidmehr, A., Roby, D. D., Evans, A. F., Loge, F. J., & Werner, I. (2012). Transcription profiling in environmental diagnostics: Health assessments in Columbia River basin steelhead (*Oncorhynchus mykiss*). *Environmental Science and Technology*, 46(11), 6081–6087.
doi.org/10.1021/es3005128
- Cronin, G. (2001). Resource allocation in seaweeds and marine invertebrates: chemical defense patterns in relation to defense theories. In: McClintock, I.B., Baker, B.I. (eds) *Marine Chemical Ecology*. CRC, Boca Raton, FL. pp. 325-354.
- Dean, M., & Annilo, T. (2005). Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annual Review of Genomics and Human Genetics*, 6, 123–142. doi:10.1146/annurev.genom.6.080604.162122
- DeBusk, B.C., Slattery, M., Ki, J., Lee, J., Aparicio-Fabre, R., & Schlenk, D. (2008). Species differences and effects of soft coral extracts from *Sinularia maximus* on the expression of cytochrome P4501A and 2N in butterflyfishes (*Chaetodon* spp.). *Fish Physiology and Biochemistry*, 34(4), 483–492.
- Despres, L., David, J.-P., & Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology and Evolution*, 22(8), 298–307.
- Doi, A.M., Pham, R.T., Hughes, E.M., Barber, D.S., & Gallagher, E.P. (2004). Molecular cloning and characterization of a glutathione S-transferase from largemouth bass (*Micropterus salmoides*) liver that is involved in the detoxification of 4-hydroxynonenal. *Biochemical Pharmacology*, 67(11), 2129–2139.
[PubMed:15135309]
- Dong, M., Zhu, L., Shao, B., Zhu, S., Wang, J., Xie, H., Wang, J., & Wang, F. (2013). The effects of endosulfan on cytochrome P450 enzymes and glutathione S-transferases in zebrafish (*Danio rerio*) livers. *Ecotoxicology and Environmental*

Safety, 92, 1–9. doi.org/10.1016/j.ecoenv.2012.10.019.

Donham, R.T., Morin, D., Jewell, W.T., Burns, S.A, Mitchell, A.E., Lame, M.W., Segall, H.J., & Tjeerdema, R.S. (2005a). Characterization of glutathione S-transferases in juvenile white sturgeon. *Aquatic Toxicology*, 71(3), 203–214. [PubMed: 15670627]

Donham, R.T., Morin, D., Jewell, W.T., Lame, M.W., Segall, H.J., & Tjeerdema, R.S. (2005b). Characterization of cytosolic glutathione S-transferases in juvenile Chinook salmon (*Oncorhynchus tshawytscha*). *Aquatic Toxicology*, 73(3), 221–229. [PubMed: 15935862]

Emborski, C., Reyes, A., & Biggs, J. S. (2012). Effect of β -naphthoflavone on hepatic cytochrome P4501A activity in the scribbled rabbitfish (*Siganus spinus*) from tropical Indo-Pacific coral reefs. *Ecotoxicology*, 21(8), 2153–2162. doi.org/10.1007/s10646-012-0969-1

Enayati, A., Ranson, H., & Hemingway, I. (2005). Insect glutathione transferases and insecticide resistance. *Insect Molecular Biology*, 14(1), 3-8.

Ferreira, M., Costa, J., & Reis-Henriques, M. A. (2014). ABC Transporters in fish species: A Review. *Frontiers in Physiology*, 5, 266. doi:10.3389/fphys.2014.00266

Feyereisen, R. (1999). Insect P450 Enzymes. *Annual Review of Entomology*, 44, 507–33. doi.org/10.1146/annurev.ento.44.1.507.

Forbey, J. S., Dearing, M. D., Gross, E. M., Orians, C. M., Sotka, E. E., & Foley, W. J. (2013). A pharm-ecological perspective of terrestrial and aquatic plant-herbivore interactions. *Journal of Chemical Ecology*, 39(4), 465–80. doi.org/doi: 10.1007/s10886-013-0267-2

Galus, M., Jeyaranjaan, J., Smith, E., Li, H., Metcalfe, C., & Wilson, J. Y. (2013). Chronic effects of exposure to a pharmaceutical mixture and municipal wastewater in zebrafish. *Aquatic Toxicology*, 132–133, 212–222. doi.org/10.1016/j.aquatox.2012.12.016

Goldstone, J.V., McArthur, A.G., Kubota, A., Zanette, J., Parente, T., Jönsson, M.E., Nelson, D.R., & Stegeman, J.R. (2010). Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish. *BMC Genomics*, 11:643. www.biomedcentral.com/1471-2164/11/643

Guengerich, F.P. (2008). Cytochrome P450 and Chemical Toxicology. *Chemical Research in Toxicology*, 21(1), 70–83.

Guengerich, F.P. (2003). Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions. *Archives of Biochemistry and Biophysics*, 409(1), 59-71.

- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., & Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-Seq: Reference generation and analysis with Trinity. *Nature Protocols*, 8, 1–43. doi.org/10.1038/nprot.2013.084.De
- Hayes, J.D. & Pulford, D.J. (1995). The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Critical Reviews in Biochemistry and Molecular Biology*, 30(6), 445–600. [PubMed: 8770536]
- Hayes, I.D., Flanagan, J.U., & Jowsey, I.R. (2005). Glutathione transferases. *Annual Review of Pharmacology and Toxicology*, 45, 51-88.
- Hay, M.E., Duffy, J.E., Pfister, C.A., & Fenical, W. (1987). Chemical defense against different marine herbivores: are amphipods insect equivalents? *Ecology*, 68(6), 1567-1580.
- Hay, M. E., & Fenical, W. (1988). Marine plant-herbivore interactions: The ecology of chemical defense. *Annual Review of Ecology and Systematics*, 19, 111–145. doi.org/10.1146/annurev.ecolsys.19.1.111
- Hay, M. E. (1996). Marine chemical ecology: What's known and what's next? *Journal of Experimental Marine Biology and Ecology*, 200(1–2), 103–134. doi.org/10.1016/S0022-0981(96)02659-7
- Hay, M. E., & Kubanek, J. (2002). Community and ecosystem level consequences of chemical cues in the plankton. *Journal of Chemical Ecology*, 28(10), 2001–2016. doi.org/10.1023/A:1020797827806
- Hibino, H., & Kurachi, Y. (2006). A new insight into the pathogenesis of coronary vasospasm. *Circulation Research*, 98, 579–581.
- Hoegg, S., Brinkmann, H., Taylor, J.S., & Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *Journal of Molecular Evolution*, 59(2), 190–203. doi: 10.1007/s00239-004-2613-z WOS:000223424800004. PMID: 15486693
- Iason, G.R. (2005). The Role of Plant Secondary Metabolites in Mammalian Herbivory: Ecological Perspectives. *The Proceedings of the Nutrition Society*, 64(1), 123-31.
- Ingelman-Sundberg, M. (2002). Polymorphism of cytochrome P450 and xenobiotic toxicity. *Toxicology*, 181–182, 447–452. doi.org/10.1016/S0300-483X(02)00492-4

- Jancova, P., Anzenbacher, P., & Anzenbacherova, E. (2010). Phase II drug metabolizing enzymes. *Biomedical Papers of the Medical Faculty of the University of Palacky, Olomouc, Czech Republic*, 154(2), 103–116.
- Jeong, C.B., Kim, B.M., Kang, H.M., Choi, I.Y., Rhee, J.S., & Lee, J.S. (2015). Marine medaka ATP-binding cassette (ABC) superfamily and new insight into teleost Abch nomenclature. *Scientific Reports*, 5, 15409. doi: 10.1038/srep15409
- Karban, R., & Agrawal, A. A. (2002). Herbivore offense. *Annual Review of Ecology and Systematics*, 33, 641–664. doi.org/10.1146/annurev.ecolsys.33.010802.150443
- Katzung, B. G., Masters, S. B., & Trevor, A. J. (2012). Basic and Clinical Pharmacology. *The McGraw-Hill Companies, Inc.* (12th ed.).
- Katoh, K., & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Keller, J.M., McClellan-Green, P.D., Kucklick, J.R., Keil, D.E. & Peden-Adams, M.M. (2006). Effects of organochlorine contaminants on loggerhead sea turtle immunity: Comparison of a correlative field study and *in vitro* exposure experiments. *Environmental Health Perspectives*, 114(1), 70-76.
- Konishi, T., Kato, K., Araki, T., Shiraki, K., Takagi, M., & Tamaru, Y. (2005). A new class of glutathione S-transferase from the hepatopancreas of the red sea bream *Pagrus major*. *Biochemical Journal*, 388, 299-307.
- Kubota, A., Bainy, A. C. D., Woodin, B. R., Goldstone, J. V., & Stegeman, J. J. (2013). The cytochrome P450 2AA gene cluster in zebrafish (*Danio rerio*): Expression of CYP2AA1 and CYP2AA2 and response to phenobarbital-type inducers. *Toxicology and Applied Pharmacology*, 272(1), 172–179. doi.org/10.1016/j.taap.2013.05.017
- Lee, B., Kim, D., Kim, H., Kim, B., Han, J., & Lee, J. (2018). Identification of 74 cytochrome P450 genes and co-localized cytochrome P450 genes of the CYP2K, CYP5A, and CYP46A subfamilies in the mangrove killifish *Kryptolebias marmoratus*. *BMC Genomics*, 19(1), 7.
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, 52, 231–253. Retrieved from

www.annualreviews.org/doi/abs/10.1146/annurev.ento.51.110104.151104?rfr_dat=c
r_pub%3Dpubmedandurl_ver=Z39.88-
2003andrfr_id=ori%3Arid%3Acrossref.organdjournalCode=ento

- Liu, S., Li, Q., & Liu, Z. (2013). Genome-wide identification, characterization and phylogenetic analysis of 50 catfish ATP-binding cassette (ABC) transporter genes. *PLoS ONE*, 8(5), e63895. doi:10.1371/journal.pone.0063895
- Liu X., Li S., Peng W., Feng S., Feng J., & Mahboob S. (2016). Genome-wide identification, characterization and phylogenetic analysis of ATP-binding cassette (ABC) transporter genes in common carp (*Cyprinus carpio*). *PLoS ONE*, 11(4), e0153246. doi:10.1371/journal.pone.0153246
- Luckenbach, T., Fischer, S., & Sturm, A. (2014). Current advances on ABC drug transporters in fish. *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology*, 165, 28–52. doi.org/10.1016/j.cbpc.2014.05.002
- Mansuy, D. (1998). The great diversity of reactions catalyzed by cytochromes P450. *Comparative Biochemistry and Physiology, Part C*, 121(1-3), 5–14. doi.org/10.1016/S0742-8413(98)10026-9
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., & Bryant, S.H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(DI), D200-D203.
- McLean, S., & Duncan, A.J. (2006). Pharmacological perspectives on the detoxification of secondary plant metabolites: Implications for ingestive behaviour of herbivores. *Journal of Chemical Ecology*, 32(6), 1213-1228.
- Morrison, H.G., Oleksiak, M.F., Cornell, N.W., Sogin, M.L., & Stegeman, J.J. (1995). Identification of cytochrome P-450 1A (CYP1A) genes from two teleost fish, toadfish (*Opsanus tau*) and scup (*Stenotomus chrysops*), and phylogenetic analysis of CYP1A genes. *Biochemical Journal*, 308(Pt 1), 97–104.
- Morrison, H.G., Weil, E.J., Karchner, S.I., Sogin, M.L., & Stegeman, J.J. (1998). Molecular cloning of CYP1A from the estuarine fish *Fundulus heteroclitus* and phylogenetic analysis of CYP1A genes: Update with new sequences. *Comparative Biochemistry Physiology*, 121(1-3), 231–240.
- Nagle, D. G., Paul, V. J., & Roberts, M. A. (1996). Ypaoamide, a new broadly acting feeding deterrent from the marine cyanobacterium *Lyngbya majuscula*. *Tetrahedron Letters*, 37(35), 6263–6266. doi.org/10.1016/0040-4039(96)01391-3

- Nagle, D.G., & Paul, V.J. (1998). Chemical defense of a marine cyanobacterial bloom. *Journal of Experimental Marine Biology and Ecology*, 225(1), 29–38.
- Nei, M., & Kumar, S. (2000). Molecular evolution and phylogenetics. *New York; Oxford: Oxford University Press*.
- Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., & Nebert, D. W. (1996). P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, 6(1), 1–42. Retrieved from www.ncbi.nlm.nih.gov/pubmed/8845856
- Nelson, D.R. (2003). Comparison of P450s from human and fugu: 420 million years of vertebrate P450 evolution. *Archives of Biochemistry Biophysics*, 409(1), 18–24.
- Noga, E.J., Khoo, L., Stevens, J.B., Fan, Z. & Burkholder, J.M. (1996). Novel toxic dinoflagellate causes epidemic disease in estuarine fish. *Marine Pollution Bulletin*, 32(2), 219-224.
- Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., & Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research*, 42(1), D26-31.
- Pacitti, D., Wang, T., Page, M. M., Martin, S. A. M., Sweetman, J., Feldmann, J., & Secombes, C. J. (2013). Characterization of cytosolic glutathione peroxidase and phospholipid-hydroperoxide glutathione peroxidase genes in rainbow trout (*Oncorhynchus mykiss*) and their modulation by *in vitro* selenium exposure. *Aquatic Toxicology*, 130–131, 97–111. doi.org/10.1016/j.aquatox.2012.12.020
- Paerl, H.W. (1988). Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. *Limnology and Oceanography*, 33(4), 823-847.
- Paul, V.J., & Hay, M.E. (1986). Seaweed susceptibility to herbivory: Chemical and morphological correlates. *Marine Ecology Progress Series*, 33, 255-264.
- Paul, V. J., Hay, M. E., Duffy, J. E., Fenical, W., & Gustafson, K. (1988). Chemical defense in the seaweed *Ochtodes secundiramea* (Montagne) Howe (Rhodophyta): effects of its monoterpenoid components upon diverse coral-reef herbivores. *Journal of Experimental Marine Biology and Ecology*, 114(2–3), 249–260. doi.org/10.1016/0022-0981(88)90141-4
- Paul, V. J., Nelson, S. G., & Sanger, H. R. (1990). Feeding preferences of adult and juvenile rabbitfish *Siganus argenteus* in relation to chemical defenses of tropical seaweeds. *Marine Ecology Progress Series*, 60(1-2), 23–34.

- Paul, V. J., Meyer, K. D., Nelson, S. G., & Sanger, H. R. (1992). Deterrent effects of seaweed extracts and secondary metabolites on feeding by the rabbitfish *Siganus spinus*. *Proceedings of the Seventh International Coral Reef Symposium*, 2, 1–8.
- Paul, V.J. (1992). Seaweed chemical defenses on coral reefs. In: Paul, V.J. (ed) *Ecological Roles of Marine Natural Products*. Comstock, Ithaca. 24-50.
- Paul, V.J., Cruz-Rivera, E., & Thacker, R.W. (2001). Chemical mediation of macro algal-herbivore interactions: Ecological and evolutionary perspectives. In: McClintock JB, Baker BJ (eds) *Marine Chemical Ecology*. CRC, Boca Raton, FL, 227-266.
- Paul, V. J., & Puglisi, M. P. (2004). Chemical mediation of interactions among marine organisms. *Natural Product Reports*, 21(1), 189–209. doi.org/10.1039/b302334f
- Paul, V.J., Puglisi, M.P., & Ritson-Williams, R. (2006). Marine chemical ecology. *Natural Product Reports*, 23(2), 153-180.
- Paul, V. J., Arthur, K. E., Ritson-Williams, R., Ross, C., & Sharp, K. (2007). Chemical defenses: From compounds to communities. *The Biological Bulletin*, 213(3), 226–251. doi.org/ doi.org/10.2307/25066642
- Paul, V.J., & Ritson-Williams, R. (2008). Marine chemical ecology. *Natural Product Reports*, 25(4), 637–832.
- Pérez, J., Domingues, I., Monteiro, M., Soares, A. M. V. M., & Loureiro, S. (2013). Synergistic effects caused by atrazine and terbuthylazine on chlorpyrifos toxicity to early-life stages of the zebrafish *Danio rerio*. *Environmental Science and Pollution Research*, 20(7), 4671–4680. doi.org/10.1007/s11356-012-1443-6
- Pond, S. M., & Tozer, T. N. (1984). First-pass elimination. Basic concepts and clinical consequences. *Clinical Pharmacokinetics*, 9(1), 1–25. doi.org/10.2165/00003088-198409010-00001
- Qian, X., Ba, Y., Zhuang, Q., & Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *Omics: A Journal of Integrative Biology*, 18(2), 98–110. doi.org/10.1089/omi.2013.0110
- Reinemer, P., Dirr, H.W., Ladenstein, R., Schaffer, J., Gallay, O., & Huber, R. (1991). The three-dimensional structure of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3A resolution. *EMBO Journal*, 10(8), 1997-2005.
- Reyes, A.J. (2017). *Siganus spinus* (scribbled rabbitfish) as a sentinel species for assessing xenobiotic pollution in tropical marine ecosystems (Master's thesis). University of Guam, Mangilao, Guam.

- Riordan, J.R., Rommens, J.M., Kerem, B.S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., Drumm, M.L., Ianuzzi, M.C., Collins, F.S., & Tsui, L.C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245(4922), 1066–1073.
- Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., & Laudet, V. (2001). Euteleost fish genomes are characterized by expansion of gene families. *Genome Research*, 11(5), 781–788.
- Sheehan, D., Meade, G., Foley, V. M. & Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*, 360(Pt.1), 1–16.
- Smith, E. M., Chu, S., Paterson, G., Metcalfe, C. D., & Wilson, J. Y. (2010). Cross-species comparison of fluoxetine metabolism with fish liver microsomes. *Chemosphere*, 79(1), 26–32. doi.org/10.1016/j.chemosphere.2010.01.058
- Smith, E. M., Iftikar, F. I., Higgins, S., Irshad, A., Jandoc, R., Lee, M., & Wilson, J. Y. (2012). *In vitro* inhibition of cytochrome P450-mediated reactions by gemfibrozil, erythromycin, ciprofloxacin and fluoxetine in fish liver microsomes. *Aquatic Toxicology*, 109, 259–266. doi.org/10.1016/j.aquatox.2011.08.022
- Sotka, E. E., & Whalen, K. E. (2008). Herbivore offense in the sea: The detoxification and transport of secondary metabolites. *Algal Chemical Ecology, In: Amsler*, 203–228.
- Sotka, E.E., Forbey, J., Horn, M., Poore, A.G.B., Raubenheimer, D., & Whalen, K.E. (2009). The emerging role of pharmacology in understanding consumer – prey interactions in marine and freshwater systems. *Integrative and Comparative Biology*, 49(3), 291.–313.
- Steinberg, P.D. (1992). Geographical variation in the interaction between marine herbivores and brown algal secondary metabolites. In: Paul, V. (ed) Ecological roles of marine natural products. *Comstock, Ithaca*. pp 51-92.
- Sun, Y., Huang, Y., Li, X., Baldwin, C. C., Zhou, Z., Yan, Z., Crandall, K. A., Zhang, Y., Zhao, X., Wang, M., Wong, A., Fang, C., Zhang, X., Huang, H., Lopez, J. V., Kilfoyle, K., Zhang, Y., Ortí, G., Venkatesh, B., & Shi, Q. (2016). Fish-T1K (Transcriptomes of 1,000 Fishes) project: Large-scale transcriptome data for fish evolution studies. *GigaScience*, 5, 18. doi:10.1186/s13742-016-0124-7.
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server issue), W609-12.

- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Steneck, R.S., Bellwood, D.R., & Hay, M.E. (2017). Herbivory in the marine realm: Shaping ecosystems and colliding with the Anthropocene. *Current Bioogy*, 27(11), R484-R489. doi: 10.1016/j.cub.2017.04021.
- Swanson, W. J. (2003). Adaptive evolution of genes and gene families. *Current Opinion in Genetics and Development*, 13(6), 617-622.
- Szakács, G., Váradi, A., Özvegy-Laczka, C., & Sarkadi, B. (2008). The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discovery Today*, 13(9–10), 379–393. doi.org/10.1016/j.drudis.2007.12.010
- Thacker, R. W., Nagle, D. G., & Paul, V. J. (1997). Effects of repeated exposures to marine cyanobacterial secondary metabolites on feeding by juvenile rabbitfish and parrotfish. *Marine Ecology Progress Series*, 147, 21–29. doi.org/10.3354/meps147021
- Trute, M., Gallis, B., Doneanu, C., Shaffer, S., Goodlett, D., & Gallagher, E. (2007). Characterization of hepatic glutathione S-transferases in coho salmon (*Oncorhynchus kisutch*). *Aquatic Toxicology*, 81(2), 126–136. doi:10.1016/j.aquatox.2006.11.009.
- Torregrossa, A. M., & Dearing, M. D. (2009). Nutritional toxicology of mammals: Regulated intake of plant secondary compounds. *Functional Ecology*, 23(1), 48-56. doi.org/10.1111/j.1365-2435.2008.01523.x.
- Ulloa, P. E., Medrano, J. F., & Feijoo, C. G. (2014). Zebrafish as animal model for aquaculture nutrition research. *Frontiers in Genetics*, 5(313), 1–6. doi.org/10.3389/fgene.2014.00313
- Uno, T., Ishizuka, M., & Itakura, T. (2012). Cytochrome P450 (CYP) in fish. *Environmental Toxicology and Pharmacology*, 34(1), 1–13. doi.org/10.1016/j.etap.2012.02.004
- Van Alstyne, K.L., Dethier, M.N., & Duggins, D.O. (2001). Spatial patterns in macroalgal chemical defenses. In: McClintock, J.B., and Baker, B.I. (Eds.) *Marine Chemical Ecology*. CRC, Boca Raton, FL, 301-324
- Vrolijk, N.H., & Targett, N.M. (1992). Biotransformation enzymes in *Cyphoma gibbosum* (Gastropoda, Ovulidae) – implications for detoxification of gorgonian allelochemicals. *Marine Ecology Progress Series*, 88, 237–246.
- Wang, D., Zhang, Y., Zhang, Z., & Yu, J. (2009a). User's manual KaKs_Calculator

- Toolbox 2.0. *Methods*, 1-19.
- Wang D. P., Wan, H. L., Zhang, S., & Yu, J. (2009b). Gamma-MYN: A new algorithm for estimating Ka and Ks with consideration of variable substitution rates. *Biology Direct*, 4, 20.
- Wang D., Zhang, S., He, F., Zhu, J., Hu, S., & Yu, J. (2009). How do variable substitution rates influence Ka and Ks calculations? *Genomics Proteomics Bioinformatics*, 7(3), 116–127.
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics and Bioinformatics*, 8(1), 77-80.
- Xu, C., Li, Y.C., & Kong, A.T. (2005). Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research*, 28(3), 249–268. doi:10.1007/BF02977789
- Yang, Z., & Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, 15(12), 496–503.
- Yang, Z., Swanson, W.J., & Vacquier, V.D. (2000). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution*, 17(10), 1446–1455.
- Yan, J., & Cai, Z. (2010). Molecular evolution and functional divergence of the cytochrome P450 3 (CYP3) family in Actinopterygii (ray-finned fish). *PLoS ONE*, 5(12), e14276. doi:10.1371/journal.pone.0014276.
- Zhang, Z., & Yu, J. (2006). Evaluation of six methods for estimating synonymous and nonsynonymous rates. *Genomics, Proteomics, and Bioinformatics*, 4(3), 173-181.
- Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., & Yu, J. (2006). KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics, and Bioinformatics*, 4(4), 259-263.
- Zhang, J., Yao, J., Wang, R., Zhang, Y., Liu, S., Sun, L., Jiang, Y., Feng, J., Liu, N., Nelson, D., Waldbieser, G., & Liu, Z. (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(9), 2813–2828. doi:10.1016/j.bbagen.2014.04.016.