Mol Genet Genomics (2010) 283:531–539
DOI 10.1007/s00438-010-0538-3

ORIGINAL PAPER

Alternative splicing in teleost fish genomes: same-species and cross-species analysis and comparisons

Jianguo Lu · Eric Peatman · Wenqi Wang · Qing Yang · Jason Abernathy · Shaolin Wang · Huseyin Kucuktas · Zhanjiang Liu

Received: 4 February 2010 / Accepted: 24 March 2010 / Published online: 10 April 2010

Abstract Alternative splicing (AS) is a mechanism by which the coding diversity of the genome can be greatly increased. Rates of AS are known to vary according to the complexity of eukaryotic species potentially explaining the tremendous phenotypic diversity among species with similar numbers of coding genes. Little is known, however, about the nature or rate of AS in teleost fish. Here, we report the characteristics of AS in teleost fish and classification and frequency of five canonical AS types. We conducted both same-species and cross-species analysis utilizing the Genome Mapping and Alignment Program (GMAP) and an AS pipeline (ASpipe) to study AS in four genome-enabled species (*Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus*, and *Takifugu rubripes*) and one species lacking a complete genome sequence, *Ictalurus punctatus*. AS frequency was lowest in the highly duplicated genome of zebrafish (17% of mapped genes). The compact genome of the pufferfish showed the highest occurrence of AS (~43% of mapped genes). An inverse correlation between AS frequency and genome size was consistent across all analyzed species. Cross-species comparisons utilizing zebrafish as the reference genome allowed the identification of additional putative AS genes not revealed by zebrafish transcripts. Approximately, 50% of AS genes identified by same-species comparisons were shared among two or more species. A searchable website, the Teleost Alternative Splicing Database, was created to allow easy identification and visualization of AS transcripts in the studied teleost genomes. Our results and associated database should further our understanding of alternative splicing as an important functional and evolutionary mechanism in the genomes of teleost fish.

Keywords Alternative splicing · Teleost · Fish · Genome · Duplication

Communicated by S. Hohmann.

J. Lu and E. Peatman contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-010-0538-3) contains supplementary material, which is available to authorized users.

J. Lu · E. Peatman · W. Wang · J. Abernathy · S. Wang · H. Kucuktas · Z. Liu (✉)
The Fish Molecular Genetics and Biotechnology Laboratory, Aquatic Genomics Unit, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, USA
e-mail: zliu@acesag.auburn.edu

W. Wang
College of Animal Sciences, Qingdao Agricultural University, 700 Changcheng Road, Qingdao 266109, China

Q. Yang
Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849, USA

Introduction

Alternative splicing (AS) is a cellular mechanism in eukaryotes that produces multiple mature mRNA isoforms from a pre-mRNA molecule. It is not only one of the major mechanisms for generating diversity of gene products, but also an important mechanism for modulating gene expression and function (Graveley 2001; Maniatis and Tasic 2002; Black 2003). AS is known to impact protein usage by altering signals for phosphorylation, glycosylation, and trafficking (Hiller et al. 2005). Recent research in

mammalian species is drawing attention to the fact that AS and its regulatory pathways can have important physiological consequences for an organism, including the areas of cancer progression and immune repertoire generation (Hofstetter et al. 2010; Sahoo and Im 2010).

The extent of AS has been shown to vary greatly among organisms. In some cases, however, the differences can be attributed to the extent that the genome of the organism has been studied. Approximately, 40% of human genes have two or more AS products (Mironov et al. 1999; Modrek et al. 2001; Brett et al. 2000; Nigumann et al. 2002). Brett et al. (2002) compared seven different eukaryote species with sufficient coverage of expressed sequence tags (ESTs) and mRNA data, and reported rates of approximately 45, 30, 15, 15, 10, 10, and 6% AS in human, mouse, cow, fly, rat, worm, and plants, respectively.

While the highest quality predictions of AS are generated from same-species transcript-genome alignments, these studies have traditionally been hindered by the lack of complete genome sequences, small transcript resources, or both. Cross-species approaches to detection of AS allow the utilization of a related genome sequence or transcript set to aid in analysis. Additionally, cross-species approaches offer an evolutionary assessment of conservation of AS events and mechanisms and may highlight important functional requirements for AS that have been maintained across species (Boue et al. 2003). Alignment of mouse, rat, and human ESTs in cross-species fashion allowed the identification of novel, previously unannotated exons and AS events that were subsequently validated by reverse-transcription PCR (RT-PCR) (Chen et al. 2006). Similarly, cross-species approaches in legumes allowed the identification of novel and conserved AS events (Wang et al. 2008).

No systematic analysis of rates and types of AS has been conducted in teleost fish. Two groups have previously included a teleost species, zebrafish, in their analysis of general vertebrate AS levels, both without further analysis of teleost results (Sammeth et al. 2008; Kim et al. 2007). Teleost species have only recently obtained sufficient levels of transcript sequences (Wang et al. 2010) and assembly of multiple whole genome sequences to allow meaningful analysis of AS. Given the diversity of fish species, the wide variety of fish genome sizes and complexities, and continued research into fish-specific genome duplication (Santini et al. 2009; Navratilova et al. 2010), an investigation of teleost AS is particularly relevant to our understanding of teleost genome evolution.

In this study, all available transcript sequences from 4 genome-enabled fish species fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), zebrafish (*Danio rerio*), and stickleback (*Gasterosteus aculeatus*) were aligned to their genome sequences for same-species analysis of AS using an analysis pipeline modified from Wang et al. 2008. We found that rates of AS vary widely across the studied species, even when normalizing for EST gene coverage. Rates of AS appear to be inversely correlated with genome size. Cross-species alignments of medaka, fugu, stickleback, and channel catfish, a species currently lacking a genome sequence, onto the zebrafish genome allowed the identification of additional putative AS genes. Our results represent the first major genome-wide analysis of AS in teleost fish and a major step toward understanding genome structure and evolution in this large species group.

Methods

Datasets for AS analysis

The genome sequences from zebrafish, medaka, fugu, and stickleback were downloaded from the Ensembl database. The zebrafish genome sequences version is Zv8. Other genome sequence datasets used in this study were current as of January 20, 2010. All EST/cDNA sequences from zebrafish, medaka, fugu, stickleback, and catfish were retrieved from GenBank and Ensembl genome databases.

Alignment of transcripts to genome sequences

The EST/cDNA sequences of the four genome-enabled species were mapped to their respective genome sequences using the GMAP computer program (Wu and Watanabe 2005; http://research-pub.gene.com/gmap/). Default parameters were used for GMAP. Then using the GMAP alignment output as the input for ASpipe 1.0 (http://sourceforge.net/projects/aspipe/), coordinates and scores for high-quality predicted intron/exon/alignments were extracted from the GMAP program outputs and uploaded into a local MySQL5.0 database (http://www.mysql.com/). In order to decrease AS artifacts, highly stringent parameters were used. For same-species EST/cDNA alignment, the parameters were >95% sequence identity and 80% alignment coverage. For cross-species EST/cDNA alignment the parameters were decreased to 80% sequence identity with alignment coverage remaining at 80% (Supplementary Fig. 1).

Alternative splicing types and alternative splicing identification

The coordinate information of predicted introns and exons were compared in pairwise fashion to identify AS

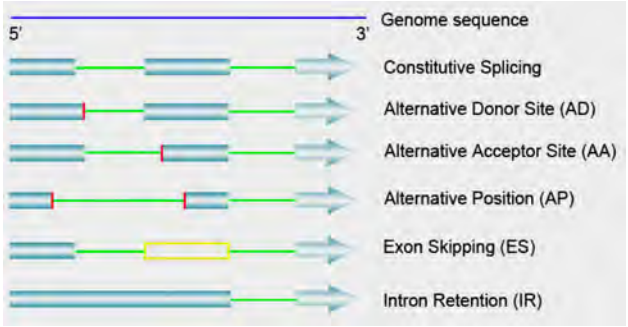

Fig. 1 Visualization of five AS types in teleost species. The *top blue line* represents the genome sequence. *Filled light blue boxes* and *arrows* represent exons, with the *arrow* indicating the direction of transcription. *Thin green lines* connecting the *boxes* indicate introns. The *open yellow box* represents a skipped exon. *Red vertical bars* represent AltD/AltA

candidates and particular types of splicing events following the methodology of Wang et al. (2008). In the case of intron versus intron comparisons, if two exons had the same 5′-end but different 3′-ends, this case was classified as Alternative Donor Site (AltD). If two exons had the same 3′-end but differed only in the 5′-ends, this case was classified as Alternative Acceptor Site (AltA). If both 3′-ends and 5′-ends differed but had overlapping introns, this case was classified as Alternative Position Site (AltP). For intron versus exon comparisons, if the intron was completely replaced by an exon, this case was classified as Intron Retention (IntronR). Alternatively, if the exon was completely replaced by an intron, this case was classified as Exon Skipping (ExonS) (Fig. 1).

AS events were identified by two approaches: (1) same-species AS events identification was based on EST/cDNA alignment with their own genome sequence; (2) cross-species AS events were identified between the zebrafish genome and the transcript sequences of 4 other teleost species. For the first case, all the EST/cDNA of zebrafish, medaka, fugu, and stickleback were aligned with their own genome sequences, respectively. Only those with identical coordinates of an alternatively processed intron/exon were regarded as same-species AS events. For the second case, using the EST/cDNA from channel catfish, fugu, medaka, and stickleback, alignments were conducted on the zebrafish genome sequence before input into ASpipe for detection of AS events.

Based on the AS genes identified using the same-species approach for fugu, medaka, stickleback, and zebrafish, basic local alignment search tool (BLAST) searches were conducted against the Uniprot database. Genes with the same Uniprot top hits were recorded and used to assess potential levels of conservation of AS genes in teleost species (Supplementary Tables 1–4).

Results

Same-species transcript/genome alignments

The EST/cDNA information of four teleost species (zebrafish, medaka, fugu, and stickleback) was extracted from NCBI and Ensembl databases for AS analysis (Supplementary Fig. 1). A total of 1,780,568 EST/cDNA from zebrafish, 638,483 EST/cDNA from medaka, 304,239 EST/cDNA from stickleback, and 73,945 EST/cDNA from fugu were retrieved (Table 1). The Genome Mapping and Alignment Program (GMAP) was then used to align the EST/cDNAs with their respective genome sequences. The percentage of mapped transcripts varied from 61.2 to 93.2%. A majority of transcript information from fugu came from predicted Ensembl gene transcripts, leading to a high rate of successful mapping to its genome sequence. The number of identified transcription units (genes) were similar among medaka, fugu, and stickleback (21,613 to 25,443), but markedly higher in zebrafish (41,365), the genome of which is characterized by high rates of gene duplications (Robinson-Rechavi et al. 2001; Peatman and Liu 2006). Analysis of AS was naturally restricted to those transcription units with multiple ESTs (58.8–71.7%). There were large disparities in average numbers of ESTs per gene among the four species, reflecting the depth of transcript resources that have been generated to-date. Zebrafish, with the largest EST resources, averaged 28.9 ESTs/gene while fugu, with minimal transcripts available, averaged only 3.3 ESTs/gene. Additional information obtained from the GMAP alignments reflected the genome characteristics of the four species. Intron sizes were dramatically larger in zebrafish (average of 6,767.4 bp) than in the compact genome of fugu (average of 687.7 bp), with medaka and stickleback falling in between these two extremes (2,317.5 and 1,216.7 bp, respectively). Similarly, average exon sizes ranged from 257.5 bp in zebrafish to 157.9 bp in fugu (Table 1).

Rates of alternative splicing vary among teleost species

GMAP alignment outputs were fed into ASpipe to identify AS genes and categorize splicing mechanisms (Table 2). Interestingly, the largest number of AS events was identified from fugu (20,676) and the smallest number from zebrafish (12,222). Medaka and stickleback had similar numbers of AS events (13,246 and 12,241, respectively) to that of zebrafish. A similar pattern was observed when the number of unique AS genes was considered. A total of 9,336 unique genes were AS in fugu, compared to 7,036 in zebrafish. Not only was the number of unique AS genes greater in fugu than the other teleost species, but also the number of AS events per gene was larger. On average, 2.21

Table 1 Transcript genome alignments and intron and exon features in 4 fish species

	Zebrafish	Medaka	Fugu	Stickleback
EST/cDNA	1,780,568	638,483	73,945	304,239
Mapped to genome	1,120,795 (62.9%)	522,516 (81.8%)	68,928 (93.2%)	186,324 (61.2%)
Transcription units (Genes)	41,365	25,443	21,613	23,188
Multi EST TU (Genes)	24,305 (58.8%)	18,250 (71.7%)	13,293 (61.5%)	15,889 (68.5%)
Average ESTs/gene	28.9	21.0	3.3	8.5
Number of introns	180,717	198,676	257,697	214,516
Average intron size (bp)	6,767.4	2,317.5	687.7	1,216.7
Long intron (>1,000 nt)	51.0%	28.6%	15.1%	17.3%
Number of internal exons	147,318	174,879	240,429	196,533
Average internal exon size (bp)	257.5	197.5	157.9	200.5

Table 2 Comparison of AS types and frequencies in 4 fish species

	Zebrafish	Medaka	Fugu	Stickleback
AltD	1,525 (12.5%)	1,427 (10.8%)	1,967 (9.5%)	1,203 (9.7%)
AltA	1,966 (16.1%)	2,272 (17.2%)	3,011 (14.6%)	2,046 (16.5%)
AltP	1,767 (14.5%)	2,391 (18.1%)	4,977 (24.1%)	2,427 (19.5%)
ExonS	4,180 (34.2%)	4,062 (30.7%)	7,255 (35.1%)	3,533 (28.4%)
IntronR	2,784 (22.8%)	3,094 (23.4%)	3,466 (16.8%)	3,212 (25.9%)
Total AS events	12,222	13,246	20,676	12,421
Unique AS genes (% of total genes)	7,036 (17.0%)	7,929 (31.2%)	9,336 (43.2%)	7,513 (32.4%)
Events/gene	1.74	1.67	2.21	1.65

AS events per gene were identified from fugu, compared to approximately 1.6–1.7 events/gene in the other three species. Most striking was the comparison of percentages of AS genes among the four species. Greater than 43% of all fugu genes were detected to be AS, while only 17% of zebrafish genes were alternatively spliced. Based on our analysis, medaka and stickleback again fell in between these two extremes, with 31.2 and 32.4% of their genes being AS, respectively (Table 2).

We next asked whether the differing coverage of ESTs/gene (Table 1) affected the detected frequency of AS genes. We, therefore, sought to normalize our analysis by EST levels. We plotted AS frequency against ESTs per gene for each of the four teleost species (Fig. 2).

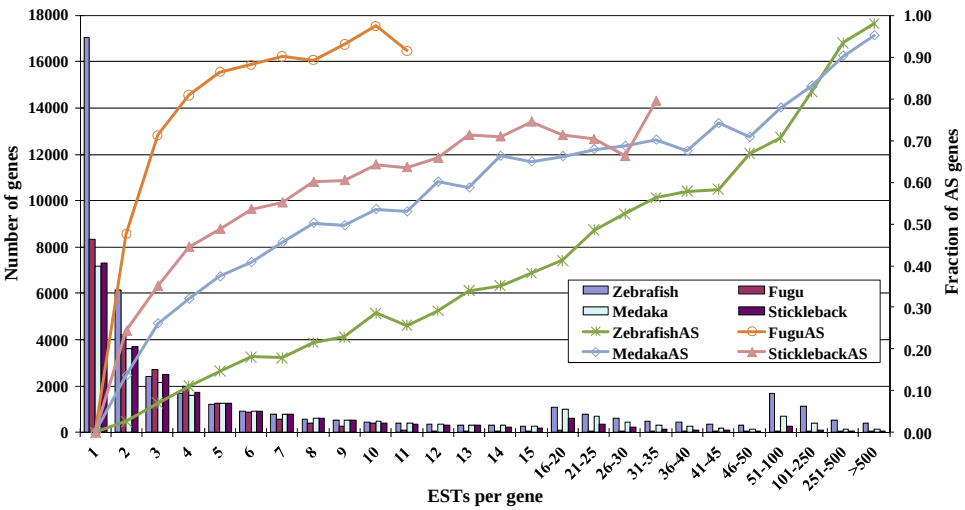

Fig. 2 Correlation between AS frequency and EST coverage. The *x*-axis represents different EST numbers per gene. The primary *y*-axis of *bars* represents the total number of genes with different EST number per gene groups. The secondary *y*-axis of lines represents the fraction of total genes involved in AS with different EST number per gene groups

Our results revealed that although AS frequency did generally increase with depth of EST coverage, the AS frequencies of the four species varied greatly regardless of EST level up to 11 ESTs per gene. Taking, for example, genes with 4 ESTs/gene across all four species, greater than 80% of these genes were AS in fugu, compared to around 10% in zebrafish. The number of fugu genes with EST coverage greater than 11 ESTs/gene and stickleback genes with greater than 35 ESTs/gene were too small for analysis, and, therefore, these categories were removed from consideration. For medaka and zebrafish, for which adequate data were available for genes with large numbers of ESTs, AS frequencies continued to vary until gene coverage was greater than 50 ESTs.

Similar distribution of alternative splicing types among teleost species

AS genes from the four teleost species were additionally characterized as to their type as defined in "Methods" (Fig. 1). Five types were quantified for each species, Alternative Donor (AltD), Alternative Acceptor (AltA), Alternative Position (AltP), Exon Skipping (ExonS), and Intron Retention (IntronR). Exon skipping was the most abundant AS type in all four species, varying from 28.4 to 35.1% of genes (Table 2). Intron retention was the next most prevalent type, comprising 16.8–25.9% of the cases. The least abundant form of AS in the tested species was the alternative donor type, accounting for only 9.5–12.5% of the AS genes in each species (Fig. 3).

Gene ontology of teleost AS genes does not indicate category enrichment

To assess whether AS genes from teleost species are evenly distributed across molecular functions, cellular components, and biological process categories, gene ontology assessments were carried out using BLAST2GO based on the UniProt database (Supplementary Fig. 2). Results were compared to those obtained using non-AS gene sets from each species (data not shown). Distributions of AS genes among the GO categories did not differ greatly from those of non-AS genes, indicating that AS genes, on the whole, are likely not highly enriched in any of the broader categories of cell structure, molecular function, or biological process across teleost species.

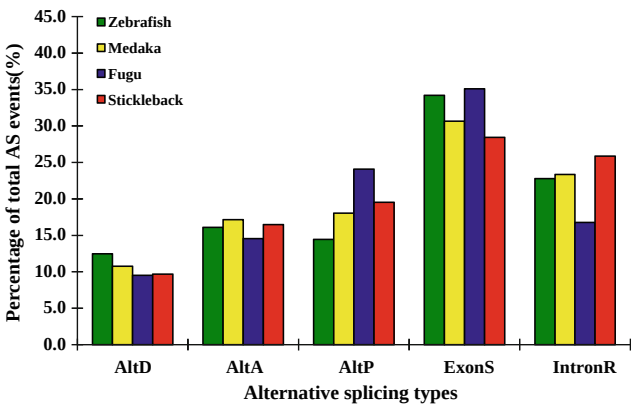

Fig. 3 The distribution of different AS types in teleost species

Conservation of AS genes among four teleost species

We also compared the Uniprot top-hit gene identities of the four species to determine the degree of conservation of the AS gene set across teleost genomes. The proportion of putative AS genes with shared identities with one or more of the other three species varied from 47.2% in zebrafish to approximately 57% in both medaka and stickleback (Fig. 4). Fugu, medaka, and stickleback, which share closer evolutionary relationships with each other than with the more distantly related zebrafish, shared larger numbers of AS gene identities with each other than with zebrafish. A total of 820 AS genes were shared among fugu, medaka, and stickleback. A smaller subset of 182 AS genes were detected in same-species alignments of all four teleost genomes. AS in these genes may play crucial roles that have aided in their maintenance throughout millions of years of evolutionary drift. The shared AS gene set identities and splicing mechanisms are included in Supplementary Tables 1–4.

Cross-species alignments for AS detection

Cross-species alignments of transcripts to target genomes have been used to identify novel AS events as well as study conserved patterns of AS across multiple species. Additionally, cross-species alignments can allow the utilization and analysis of ESTs from species lacking their own genome sequence. All available EST/cDNA sequences from channel catfish (no whole genome sequence), medaka, fugu, and stickleback were aligned against the zebrafish genome. Due to the stringent criteria necessary for accurate mapping of transcript sequences to the exon/intron boundaries of the zebrafish genome, only a small subset of highly conserved genes from each species were mapped (Table 3). The smallest number of genes (597) was mapped utilizing the transcript set from channel catfish which, unlike the other three species, lacks full-length Ensembl cDNA data. The largest number of genes was mapped from medaka (2,097). Similarly, ASpipe analysis revealed 39 AS genes based on the transcripts of channel catfish, and 132 AS genes based on medaka transcripts. A large percentage of the cross-species-identified AS genes in zebrafish had not previously been identified by analysis with zebrafish

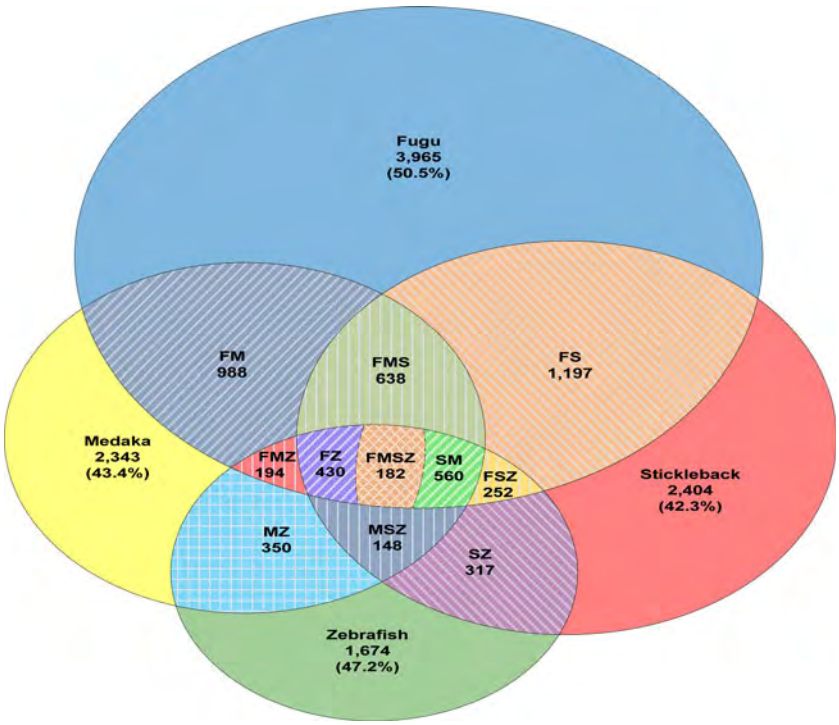

Fig. 4 Venn diagram of shared AS genes among four teleost species. Percentages of non-shared genes are given in parentheses for each species and were calculated based on number of total AS genes in each species with significant (1e-5) hits in the Uniprot database (3,547 in zebrafish, 5,403 in medaka, 7,846 in fugu, and 5,698 in stickleback). For shared regions, *Z* is zebrafish, *F* fugu, *M* medaka, and *S* stickleback. *FZ*, therefore, represents the number of AS genes shared only between fugu and zebrafish and *FMSZ* encompasses those AS genes shared among all four species. Artificial divisions were created to allow visualization of the number of shared genes between fugu and zebrafish, and between stickleback and medaka

ESTs alone. Percentages of novel AS genes identified with the cross-species transcripts varied considerably depending on species, from 31.8% of AS genes supported by medaka sequence evidence to 75.0% of AS genes from fugu (Table 3).

The distribution of AS types identified from cross-species alignments was similar to that found with the same-species approach. On average, exon skipping was the most abundant AS type identified by both cross-species and same-species approaches, and alternative donor and alternative acceptor AS were the rarest in both cases (Table 4).

Teleost alternative splicing database

An online database for teleost alternative splicing was created using the ASviewer format (Wang et al. 2008). The database is publically available and can be found at http://asviewer.acesag.auburn.edu/ASviewer/index.htm. Users of the database can search for the number, type, and location of AS genes in channel catfish, fugu, medaka, zebrafish, and stickleback as well as visualizing the AS event (Fig. 5). AS information will be added for additional species as warranted.

Discussion

Alternative splicing is recognized as one of the mechanisms by which the coding capacity and diversity of the genome can be amplified. Rates of AS have been shown to vary according to the complexity of eukaryotic species, potentially explaining the tremendous phenotypic diversity among species with similar numbers of coding genes. Little is known, however, regarding mechanisms and rates of AS in teleost fish, a major vertebrate group encompassing over 27,000 species and more than half of vertebrate diversity. In this study, we have conducted the first major genome-wide analysis of AS in teleost fish. Based on same-species

Table 3 Cross-species EST alignment results with zebrafish genome sequences

Species	EST/cDNA	Mapped to genome	Genes[a]	AS genes	Novel[b]
Channel catfish	354,377	5,313 (1.5%)	597	39	20 (51.3%)
Medaka	638,483	58,976 (9.2%)	2,097	132	42 (31.8%)
Fugu	73,945	3,423 (4.6%)	1,099	84	63 (75.0%)
Stickleback	304,239	17,916 (5.9%)	1,117	87	54 (62.1%)

[a] Number of genes successfully mapped using cross-species transcripts

[b] Novel AS genes not identified using zebrafish transcripts

Table 4 AS events and types predicted from cross-species EST alignment with zebrafish

Species	Events	AltD	AltA	AltP	ExonS	IntronR
Channel catfish	46	2 (4.3%)	4 (8.7%)	10 (21.7%)	24 (52.3%)	6 (13.0%)
Medaka	177	13 (7.3%)	12 (6.8%)	38 (21.5%)	68 (38.4%)	46 (26.0%)
Fugu	114	4 (3.5%)	12 (10.5%)	32 (28.1%)	34 (29.8%)	32 (28.1%)
Stickleback	115	14 (12.2%)	5 (4.3%)	26 (22.6%)	34 (29.6%)	36 (31.3%)
Average percentage of cross-species[a]		7.6%	7.0%	23.9%	35.4%	26.2%
Average percentage of same-species[a]		10.6%	16.1%	19.1%	32.1%	22.2%

[a] Average percentage of AS events encompassed by the indicated types by cross-species and same-species approaches

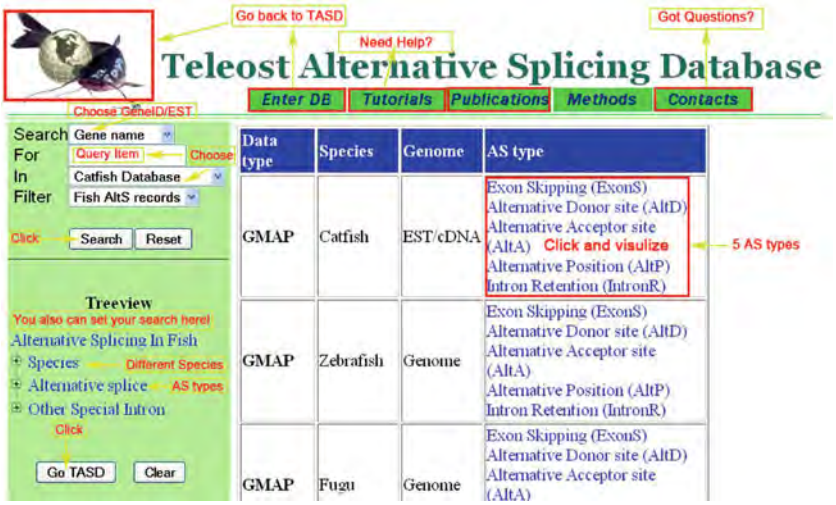

Fig. 5 Teleost Alternative Splicing Database screenshot. The database is publicly available on the web at http://asviewer.acesag.auburn.edu/ASviewer/index.htm

and cross-species alignment approaches, here we report that rates of AS appear to vary widely across teleost species, even when accounting for differing EST gene coverage rates. Our analysis indicates that rates of AS are inversely correlated with genome size, potentially shedding light on coding mechanisms underlying differing genome evolutionary strategies within teleost fish.

Expressed sequence tag (EST) coverage rates varied among the analyzed teleost species, potentially impacting our ability to detect AS events. Wang et al. (2008) reported that apparent differences in AS frequency between legume species could be attributed to differences in numbers of ESTs available for analysis. Legume species with larger EST resources had higher detected AS frequencies. After normalizing for EST number per gene, the observed frequency of AS was similar across all tested legume species. We observed a markedly different phenomenon in teleost species. EST coverage ranged from 3.3 ESTs per gene in fugu to 28.9 ESTs per gene in zebrafish. However, despite low transcript coverage, the highest frequency of AS was detected in fugu. Also, normalizing data for ESTs per gene did not diminish differences in AS frequency among the four species. Pronounced differences in AS frequency were observed at low to medium levels of EST gene coverage, with convergence of AS frequencies appearing to occur only at greater than 25 ESTs per gene, in species with sufficient EST data in this range (Fig. 2). These differences in AS frequency, therefore, appear to reflect genuine differences in genomic architecture and protein coding strategy among teleost species.

Indeed, detected AS frequencies for the four teleost species are inversely correlated with their genome sizes. Fugu (43.2% AS frequency) has a compact genome size of 0.4 Gb. Stickleback (32.4% AS) and medaka (31.2% AS) have intermediate genome sizes of 0.6 Gb and 0.7 Gb, respectively. Zebrafish, with only 17% of genes being detected as AS, has the largest genome by far—1.7 Gb (Aparicio et al. 2002; Vinogradov 1998; Kasahara et al. 2007; Hukriede et al. 1999). Fish species with smaller genomes may rely more heavily on AS to generate necessary protein diversity. Conversely, species, such as zebrafish with larger, more duplicated genomes may not require the generation of as many AS transcripts to augment protein diversity (van der Aa et al. 2009). Researchers have previously suggested an inverse relationship between rates of gene duplication and AS in animals (Xing and Lee 2006) and, more recently, in plants (Yuan et al. 2009) based on single gene or gene family investigations.

However, the present genome-wide analysis of AS in several teleost species offers one of the first wide-ranging examinations of this relationship. Further work is needed to quantify more accurately levels of gene duplication in individual teleost species and to begin to examine the evolutionary interactions between duplication and AS across the teleost radiation.

Our study also examined, in part, the utility of cross-species alignments for identification of novel and conserved AS events. Our results indicated that additional putative AS events could be detected applying the transcript sets of related species to a target genome. However, detection power appeared to be limited by transcript length (full-length cDNAs vs. partial transcripts) and evolutionary distance between the aligned species. Additional cross-species examination of AS patterns may be informative for transcript-rich, genome-poor species groups, such as tilapias, carps, and salmonids.

Analysis of shared gene identities across the AS gene sets of the four examined teleost species may be informative in identifying biological processes for which AS-generated diversity is essential. In particular, we identified 182 genes that were alternatively spliced in the genomes of each of the four species (Supplementary Table 4). This gene set appeared to be moderately enriched for genes regulating developmental processes, anatomical structure formation, and immune system processes (Supplementary Fig. 3). Examples of conserved AS gene identities included Fc gamma binding protein, disheveled-associated activator of morphogenesis 1, ligand of numb protein X, ephrin A3, macrophage mannose receptor 1, rhamnose-binding lectin, and complement C3-1. The functional consequences of AS of these genes in teleost fish are unknown in all cases. Additional research is clearly needed to examine the mechanisms supporting maintenance of specific AS genes during teleost species evolution.

Acknowledgments We would like to thank Dr. Bing-bing Wang for sharing the ASviewer software with us. This research was supported in part by funding from the USDA's National Institute of Food and Agriculture.

References

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J, Dehal P, Christoffels A, Rash S, Hoon S, Smit A et al (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. Ann Rev Biochem 72:291–336

Boue S, Letunic I, Bork P (2003) Alternative splicing and evolution. BioEssays 25:1031–1034

Brett D, Hanke J, Lehmann G, Haase S, Delbruck S, Krueger S, Reich J, Bork P (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett 474:83–86

Brett D, Pospisil H, Valcarcel J, Reich J, Bork P (2002) Alternative splicing and genome complexity. Nature Genet 30:29–30

Chen FC, Chen CJ, Ho JY, Chuang TJ (2006) Identification and evolutionary analysis of novel exons and alternative splicing events using cross-species EST-to-genome comparisons in human, mouse and rat. BMC Bioinform 7:136

Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet 17:100–107

Hiller M, Huse K, Platzer M, Backofen R (2005) Creation and disruption of protein features by alternative splicing a novel mechanism to modulate function. Genome Biol 6:R58

Hofstetter G, Berger A et al (2010) Alternative splicing of p53 and p73: the novel p53 splice variant p53delta is an independent prognostic marker in ovarian cancer. Oncogene 29:1997–2004

Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, Barbazuk WB, Li FN, Paw B, Postlethwait JH et al (1999) Radiation hybrid mapping of the zebrafish genome. Proc Natl Acad Sci USA 96:9745–9750

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y et al (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719

Kim N, Alekseyenko AV, Roy M, Lee C (2007) The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. Nucleic Acids Res 35:D93–D98

Maniatis T, Tasic B (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418:236–243

Mironov AA, Fickett JW, Gelfand MS (1999) Frequent alternative splicing of human genes. Genome Res 9:1288–1293

Modrek B, Resch A, Grasso C, Lee C (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res 29:2850–2859

Navratilova P, Fredman D, Lenhard B, Becker TS (2010) Regulatory divergence of the duplicated chromosomal loci sox11a/b by subpartitioning and sequence evolution of enhancers in zebrafish. Mol Genet Genomics 283:171–184

Nigumann P, Redik K, Matlik K, Speek M (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79:628–634

Peatman E, Liu ZJ (2006) CC chemokines in zebrafish: Evidence for extensive intrachromosomal gene duplications. Genomics 88:381–385

Robinson-Rechavi M, Marchand O, Escriva H, Laudet V (2001) An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr Biol 11:R458–R459

Sahoo A, Im SH (2010) Interleukin and interleukin receptor diversity: role of alternative splicing. Int Rev Immunol 29:77–109

Sammeth M, Foissac S, Guigo R (2008) A general definition and nomenclature for alternative splicing events. PLOS Comp Biol 4(8):e1000147

Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol 9:194

van der Aa LM, Levraud JP, Yahmi M, Lauret E, Briolat V, Herbomel P, Benmansour A, Boudinot P (2009) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biol 7:7

Vinogradov AE (1998) Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. Cytometry 31:100–109

Wang BB, O'Toole M, Brendel V, Young ND (2008) Cross-species EST alignments reveal novel and conserved alternative splicing events in legumes. BMC Plant Biol 8:17

Wang S, Peatman E, Abernathy J et al (2010) Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. Genome Biol 11:R8

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875

Xing Y, Lee C (2006) Alternative splicing and RNA selection pressure - evolutionary consequences for eukaryotic genomes. Nature Rev Genet 7:499–509

Yuan YA, Chung JD, Fu XY, Johnson VE, Ranjan P, Booth SL, Harding SA, Tsai CJ (2009) Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. Proc Natl Acad Sci USA 106:22020–22025