Mar. Biotechnol. 1, 437–447, 1999

MARINE BIOTECHNOLOGY

Development of Polymorphic EST Markers Suitable for Genetic Linkage Mapping of Catfish

Zhanjiang Liu,* Attila Karsi, and Rex A. Dunham

The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, 203 Swingle Hall, Auburn University, Auburn, AL 36849, U.S.A.

Abstract: Expressed sequence tag (EST) markers are important for gene mapping and for marker-assisted selection (MAS). To develop EST markers for use in catfish gene mapping, 100 randomly picked complementary DNAs from the channel catfish (*Ictalurus punctatus*) pituitary library were sequenced. The EST sequences were used to design primers to amplify channel catfish and blue catfish (*I. furcatus*) genomic DNAs. Polymerase chain reaction products of the ESTs were analyzed to determine length polymorphism between the channel catfish and blue catfish. Eleven polymorphic EST markers were identified. Five of the 11 EST markers were from known genes and the other six were from unidentified ESTs. Seven ESTs were found to be associated with microsatellite sequences. Analysis of channel catfish gene sequences indicated highly biased codon usage, with 16 codons being preferably used. These codons were more preferably used in highly expressed ribosomal protein genes and in highly expressed pituitary hormone genes. G/C-rich codons are less used in channel catfish than those in other vertebrates suggesting AT-richness of the channel catfish genome.

Key words: Expressed sequence tags, catfish, marker, polymorphism, gene mapping

INTRODUCTION

Genetic markers are classified into two groups: type I markers are associated with a gene of known function, and type II markers are associated with anonymous gene segments (O'Brien, 1991). Polymorphic type I markers are more useful as anchorage points for comparative gene mapping than type II markers. Studies have indicated high levels of conservation in genomic organization among vertebrates (Harless et al., 1995; Reeves et al., 1998). In a syntenic group, gene organization is conserved; therefore, genes can be mapped by establishing syntenic groups between a "map-rich" species and a "map-poor" species. Markers already mapped in the map-rich species can be mapped in the map-poor species by analogy in the same syntenic group. Humans, mice, and several mammalian livestocks are map-rich, whereas all aquaculture species are currently map-poor. While conservation of genomic organization between mammals and fish is largely unknown, conservation between the model organism and other fish should be high. Development of type I markers thus should facilitate location of mapped markers from the map to the catfish map. The catfish map should in turn facilitate mapping in other aquaculture species.

Expressed sequence tags (ESTs) are short, single-pass complementary DNA sequences generated from randomly selected library clones. Characterization of ESTs is a convenient and rapid way to identify new genes in various

Received June 29, 1998; accepted March 29, 1999.

*Corresponding author; telephone (334)-844-4054; fax (334)-844-9208; e-mail zliu@acesag.auburn.edu.

organisms. An EST approach is being used extensively to analyze expressed genes from humans (Adams et al., 1992), *Caenorhabditis elegans* (Waterston et al., 1992), *Arabidopsis thaliana* (Hofte et al., 1993), *Plasmodium falciparum* (Chakrabarty et al., 1994), *Schistosoma mansoni* (Franco et al., 1995), *Entamoeba histolytica* (Azam et al., 1996), rice (*Oryza sativa*) (Aliyeva et al., 1996), medaka (*Oryzias latipes*) (Hirono and Aoki, 1997), and Japanese flounder (*Paralichthys olivaceus*) (Inoue et al., 1997). There are numerous ongoing EST projects that provide expressed sequences by single-pass sequencing both upstream and downstream of cDNAs using arbitrarily selected cDNA libraries. This means that discovery of novel genes takes as little time as it takes to perform a computer search of a sequence database (Tilghman, 1996), rather than the months or even years that were required to work at the bench. Because of the relative ease of EST characterization, the EST database called dbEST is the fastest growing division of GenBank (Wolfsberg and Landsman, 1997). The 415,000 human ESTs, for instance, are believed to represent 40% to 80% of the total number of human genes (Wolfsberg and Landsman, 1997).

The EST approach is useful not only for identification of new genes, but also for genome mapping. The generation of human ESTs, for example, was a crucial step in the progress of the human genome project (Boguski and Schuler, 1995). The availability of these human ESTs enabled the recent mapping of 16,000 genes in the human genome (Schuler et al., 1996). Although many other types of molecular markers are available for use in genome mapping, EST markers have some advantages; they represent genes and some can serve as type I markers. Development of EST markers is straightforward and technically simple although tedious. Single-pass sequencing analysis of either or both upstream and downstream cDNA would produce sequences for designing primers for polymerase chain reaction (PCR). PCR analysis of the genomic DNA from the parents used to produce the reference or resource families would reveal whether or not the markers are informative (polymorphic) in the reference family. Polymorphic EST markers are needed for gene mapping analysis.

The number of EST sequences deposited to public databases is huge. Of the 810,000 ESTs, about 70% were reported from various human tissues. ESTs of teleostei have accounted for less than 1% of the total ESTs reported (Gong et al., 1994; Hirono and Aoki, 1997; Inoue et al., 1997; Karsi et al., 1998). The EST analysis is fast and technically easy. Development of ESTs provides not only potential markers for genomic research, but also information on codon usage in a specific species of teleostei, allowing proper gene construction in genetic engineering programs.

As part of catfish genome mapping, we have initiated development of EST markers using a channel catfish (*Ictalurus punctatus*) and blue catfish (*I. furcatus*) hybrid system. The hybrids of channel catfish and blue catfish are fertile. Their F_2 and backcross hybrid reference/resource families have been produced (Liu and Dunham, 1998). We previously reported the transcriptional activities of the channel catfish pituitary during hormone-induced ovulation as revealed by EST analysis (Karsi et al., 1998). Here we report the identification of 11 polymorphic EST markers developed from 100 randomly picked cDNA clones from the pituitary cDNA library, the association of ESTs with microsatellite sequences, and codon usage in channel catfish as analyzed from the ESTs and sequences extracted from the GenBank databases.

Materials and Methods

Construction of the Channel Catfish Pituitary cDNA Library

The λ Uni-ZAP II cDNA library constructed from pituitary RNA of channel catfish used for this work has already been described (Liu et al., 1997). The λ phage libraries were converted into plasmid libraries by a mass *in vivo* excision procedure (Stratagene, La Jolla, Calif.; Karsi et al., 1998). The resulting plasmid libraries contained cDNA clones harbored in the plasmid pBluescript SK^-. Conversion into a plasmid library simplified manipulation and handling.

DNA Sequencing

Plasmid DNA was prepared by an alkaline lysis procedure (Sambrook et al., 1989). The upstream and downstream ends were sequenced using the cycle sequencing kit (Applied Biosystems) with the reverse and universal sequencing primers, respectively. For most clones, readable sequences of 250 to 350 bases were generated from single-pass sequencing. Because of the specialized function of the pituitary, we expected that some genes were expressed at high levels and therefore would be detected at high frequencies. To reduce repeated sequencing, a DNA dot blot analysis was conducted immediately after a specific cDNA was encountered twice.

Sequence Analysis

Nucleotide sequences were edited by using DNASIS or DNA Star software packages. Vector sequences were removed manually before searching for homologies using BLASTN or BLASTX servers, through the Internet (NCBI, www.ncbi.nlm.nih.gov). Matches were considered to be significant only when the smallest sum probability (p) was less than .0001 and scores were greater than 160 for BLASTN and greater than 80 for BLASTX.

PCR Primer Design

PCR primers were designed using the OLIGO software package version 5.1 (National Biosciences, Inc., Plymouth, Minn.). In each case, primers were designed at the highest possible stringency. Factors considered included PCR product length, duplex formation (homo primer dimers and hetero primer dimers), hairpin formation, false priming sites, and the melting temperature. For analysis of potentially small differences of the PCR products, PCR primers were designed to generate products of about 150 to 200 bp, but in some cases primers were designed to generate larger PCR products considering other factors such as primer dimer formation, high A/T-richness, etc. In this range, a length difference of several base pairs should be readily detected on a high percentage polyacrylamide gel. In addition to the factors analyzed by the OLIGO software, several other factors were considered for selection of primers after generation of a primer pair list from the computer. Primers containing stretches of simple sequence (e.g., a single base tract) were avoided. Primers with a 3′ C or G were preferred for stronger annealing for initiation of PCR reaction.

PCR Amplification and Analysis

Approximately 200 ng of genomic DNA was amplified in PCR reactions of 50 μl containing 50 mM KCl, 10 mM Tris (pH 9.0 at 25°C), 0.1% Triton X-100, 0.25 mM each of deoxynucleotide triphosphate (dNTPs), 1.5 mM $MgCl_2$, 20 μM each of the upper and lower PCR primers, and 2.5 units of *Taq* DNA polymerase. The temperature profiles used to amplify the cDNAs were 94°C for 30 seconds, 57°C for 1 minute, and 72°C for 2 minutes, for 35 to 40 cycles. Following amplification, samples of 3 μl were mixed with 1 μl loading dye and electrophoresed on 10% acrylamide (19:1 acrylamide:bis-acrylamide) gels.

Codon Usage Analysis

Sequences from the GenBank databases were extracted by using the BLAST text searching with key words: catfish, *Ictalurus punctatus.* One hundred eighty-six sequences were extracted from GenBank Re. 97.0. Only sequences from channel catfish (*I. punctatus*) were used. Coding sequences of messenger RNAs were directly copied into DNASIS program. Coding sequences from genes were spliced according to the GenBank specifications. The sequences from the EST analysis were first checked against the GenBank databases using the cross-checking program BLASTX for proper open reading frames (ORFs). Only regions with a proper ORF were used for analysis. Unidentified clones were not used in the analysis. After being copied into the DNASIS software, the sequence was first translated to look for a proper ORF. Then each sequence was analyzed for codon usage using the "codon usage" command of DNASIS. The results were saved as files, which were then copied into Microsoft Excel for codon usage analysis. Codon usage was expressed as codons used per thousand, which itself is a reflection of amino acid composition, but the relative frequencies among synonymous codons reflect codon usage preferences.

Results and Discussion

The gene identities of the ESTs and their representation in the transcriptional products in the channel catfish pituitary were previously summarized (Karsi et al., 1998). Their sequence has been deposited to the dbEST. The analysis of the 100 clones generated ESTs representing 59 distinct gene products, of which 30 are known genes and 29 are unknown gene products.

Identification of Polymorphic EST Markers

Genomic DNAs of channel catfish and blue catfish were used to amplify fragments of ESTs to reveal length polymorphism using PCR primers designed from the EST sequences (Table 1). Among 47 loci tested, length polymorphism was detected from 11 loci (Figure 1 and Table 2). Among the 11 polymorphic EST markers, 5 were identified from known genes and the other 6 from unidentified genes. The five known gene loci were gonadotropin I-β subunit, ribosomal proteins L30, L35, and L41, and proopiomelanocortin. These five EST markers are type I markers.

Primers were designed from both upstream and downstream of the cDNA to compare polymorphism levels between the two regions. Of the 24 primers designed from the 5′ ESTs, 5 (21%) showed polymorphism. Similarly, of the 12 primers designed from the 3′ ESTs, 6 (50%) showed polymorphism. It appeared that the polymorphic rates were

Table 1. Primers Used to Amplify EST Fragments

EST	Gene identity	Upper primer	Lower primer
E001	Cyclin 2b	agtcacgggatgaacgaac	aggccgaattatgagaaacag
E004	Ribosomal protein S9	atctgccatgccggttgc	cttcgaagagacgcctggtcc
E005	Unknown	atattgcagaatacccaagac	tttagccttagtggtttgtg
E006	Unknown	aacaaggcttccacagagacc	ccatggacaggaactgctg
E008	Unknown	attattcgaagagttccaccc	gcagcgtagaagaaaccctc
E009	Prolactin	gcgtgatatcgatttattgac	taaaccaggaggaaaacatc
E010	Unknown	ggtgaaagtactgctggctac	aacatcatcaacaaagacgg
E011	Unknown	tttcacacgtggtcagattcg	ggttccttcctagccatacgg
E015	GTP-binding protein	gctactgatcaccctccaccc	gcttcccatgttcctgttctg
E016	RAP1A	tctttgtggagaagtatgaccc	atactgtagccttgaatctgtcg
E019	Unknown	atattcatgccacgagatcagc	tgcactgcatctacacctacgtt
E020	Unknown	tgtagatctgttcctcgaagcac	cagaggaattactgcgttcacac
E022	Unknown	ggaagtggtagaaggcggtcac	gatagagcagggcagcatagtcg
E024	Cytochrome c oxidase	aagttgcactgtaccgtatcagc	tccaggacacattcattgagg
E027	Ribosomal protein S11	cttaggcttcaaaacccccagag	tcatggcaatacagtggaaccc
E028	Unknown	cacacgcgcatggacacatacg	cgaattcggcacgaggtgaacg
E031	Unknown	ataaagtgcagaaggttcctccc	ggcgatcgtatcaaacagagc
E032	Acylphosphatase	cgaaaatatacacagtctgagg	cagtgtgactagcgaacaac
E033	Unknown	ggaggatgatgaaggccagag	cgggaggacctatcttgagacc
E034	MHC complex protein	gctgggacaagatggtcaagg	cacagagccagtagcggttagg
E035	Ribosomal protein S24	ccatggtctacgattcactggac	aggatgatcatctccactgaagg
E036	Unknown	taagggctgcagctgtcaatcag	caaacccgaccaatcagacgac
E038	Unknown	ggagaaggggaaaaatattacc	gcatgtacctcacttcaacag
E040	Unknown	tgtgcttatatttttaatggc	tattgattagagtctgatggtc
E043	Ribosomal protein L18	cctggctcgccgttcagatgc	cgagctcctgcagtcaccctc
E044	Gonadotropin I-β	aatctggatcactttcagcatc	ccttcatgatattcagggtcac
E046	Ribosomal protein L30	gactgcagctggtgatgaagag	gccagtgtgcacaccctgtag
E048	Unknown	acgtcgagtgaaagaacttgag	agcttagatctctgcagccac
E050	Cystatin	cacttctggcggtctttctggc	caccacccttctgcaggctgtc
E051	Unknown	aatcacattttagagcactccc	gtacccaaaggacattctgac
E052	Unknown	acgtcgagtgaaagaacttgag	tgatgcattccaggatattgtc
E053	Unknown	tccctgtcgcagtaattatcc	aagtgcagaaggttccatccc
E054	Unknown	tgaaaggctttgcagcgtggag	ggtgcctctgaaccggtccttc
E057	Ribosomal protein L7a	ctgcatcgttaaaggaaaagcc	ggacccatgatgttacctccc
E061	Ribosomal protein L35	cagctggatgacctgaaagtg	ccagaggcttgtacttcttccc
E064	cAMP-responsive element modulator	gcctcagggaaatagctttgtg	gcagcttccctgttcttcatc
E067	Somatolactin	aatttgcagaaaagatgtatgg	cgctcacaagatgagacattc
E071	Unknown	cagcagttttgtaaagatgagg	ctgatctgtgtagtgtctcgtg
E078	Unknown	gatcactaccagaggttttcac	ctgaagaggtctttaagtggag
E083	Ribosomal protein L41	cgaggctgaagcgtaaaaggag	gccatgcaaggtcaagtccaac
E084	Unknown	aacttgatctgaaggtgtgtg	tcttatcgctgcttttacag
E085	Unknown	tgctagtcacgtgcagttctga	ggaggtgagagaagccatgtg
E086	Proopiomelanocortin	gcgtgatatcgatttattgac	gtgcactgatctgtgtagtgtc
E088	Unknown	tttgagtgcatgaaagactaga	tgagctgtgataactgtagcaa
E093	MHR23B	ctagcgaaaatagcctggactc	tccagcgacaggaaaactctc
E095	Ribosomal protein L41	gaaacgtctccaacatgagagc	gcctagcaaggtcaagtccaac
E098	Unknown	tatcaagcttatcgagaccgtc	cagaggattaacatggaagcc

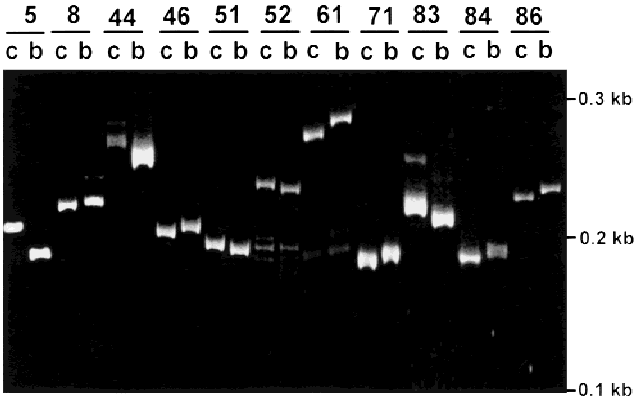

Figure 1. Polymorphic EST markers suitable for gene mapping of catfish. EST loci were amplified by PCR and analyzed on 10% polyacrylamide gels. For each locus, channel catfish, *Ictalurus punctatus* (c lanes) and blue catfish, *I. furcatus* (b lanes) genomic DNA was amplified by using primers designed from the channel catfish EST sequences. Names of the 11 loci are labeled at the top with numbers Ip0005 (5), Ip008 (8), Ip044 (44), Ip046 (46), Ip051 (51), Ip052 (52), Ip061 (61), Ip071 (71), Ip083 (83), Ip084 (84), Ip086 (86). Molecular weight size markers are indicated on the right margin of the gel.

higher from the 3′ ESTs than the 5′ ESTs. Eleven pairs of primers were designed to amplify the whole gene, with the upper primer designed from the 5′ EST, and the lower primer designed from the 3′ EST. Most of the 11 primer pairs failed to amplify, probably owing to the large sizes of the products. Among the loci amplified, no polymorphism could be detected. Longer PCR product covering the whole gene or most of the gene offers the opportunity to include introns thus increasing the chances of detecting polymorphism. However, large PCR product makes it difficult to directly identify length polymorphism. Opportunities open to analyze these products after cleavage with restriction enzymes. Nonetheless, considering the potential application of these markers to gene mapping, in which large numbers of samples must be analyzed, complicated processes of analysis are not desirable.

The length differences were subtle in all cases except for Ip005 (Figure 1), but were readily detected by using 10% acrylamide gel electrophoresis. Thus, these markers can be analyzed by PCR-facilitated genotyping without extensive analysis.

The level of polymorphism of the catfish ESTs indicates that it is possible to obtain several hundred EST markers for mapping in catfish using the channel catfish × blue catfish interspecific hybrid system (Liu et al., 1998a, 1998b, 1998c). However, the efforts involved in characterizing several thousand ESTs would be significant. Our experience with the channel catfish pituitary cDNA libraries indicates that fewer genes are expressed in highly specialized tissues such as pituitary than in less-specialized tissues. For instance, hormone clones were highly represented in the pituitary cDNA libraries (Karsi et al., 1998). Hormones accounted for 32% and translational machinery proteins accounted 11% of the 100 EST clones. For a thorough characterization of a large number of ESTs without repeated sequencing, cDNA libraries constructed from mRNA of several tissues should be used. In addition, if initial selection against already sequenced clones can be conducted to eliminate repeated clones from sequencing, the efforts for EST analysis should be reduced.

Potential Application of the ESTs and Polymorphic EST Markers

The identified EST polymorphic markers and ESTs have three potential applications. First, the polymorphic EST markers can be used in a genetic linkage mapping project involving the interspecific reference/resource families. Segregation analysis of these polymorphic EST markers should allow linkage to be established with other markers and assignment of these markers to chromosomes on which they reside. Second, the sequence information generated from this study should facilitate physical mapping of linkage using radiation hybrid panels (Cox et al., 1990) as they are being developed from channel catfish. The sequences between channel catfish and rodent cells such as Chinese hamster cells should be divergent enough to allow specific amplification of channel catfish sequences, but not the hamster sequences. Third, the ESTs should be useful for identification and mapping of single nucleotide polymorphism (SNuP) markers (Wang et al., 1998) in channel catfish.

Association of Microsatellite Sequences with ESTs

Among the 100 clones analyzed, 7 distinct clones contained simple sequence repeats (SSR, or microsatellites) (Table 3). The genes that are associated with microsatellites include the GTP-binding protein (Ip015), the nucleotide excision repair gene (Ip093), the gonadotropin α subunit (Ip095), the growth hormone (Ip0097), and three unknown genes (Ip005, Ip082, and Ip085). This observation indicates that transcribed microsatellites are abundant and should be exploited in gene mapping analysis. The use of transcribed

Table 2. Identified Polymorphic EST Markers and PCR Conditions

EST	Gene identity	Primers designed from	Annealing temperature (°C)	Expected PCR products (bp)
Ip005	Unknown	Downstream EST	57	178
Ip008	Unknown	Downstream EST	57	194
Ip044	Gonadotropin I-b	Downstream EST	57	243
Ip046	Ribosomal protein L30	Upstream EST	57	233
Ip051	Unknown	Upstream EST	57	255
Ip052	Unknown	Upstream EST	57	257
Ip061	Ribosomal protein L35	Upstream EST	57	190
Ip071	Unknown	Downstream EST	57	187
Ip083	Ribosomal protein L41	Upstream EST	57	212
Ip084	Unknown	Downstream EST	57	177
Ip086	Proopiomelanocortin	Downstream EST	57	227

Table 3. EST Clones Associated with Microsatellite Sequences

EST	Repeat units	Repeat location	Gene identity	Reference
Ip005	$(A_{3-9}C)_5$	5′-NTR	Unknown	This study
Ip015	$(CA)_9$	5′-NTR	GTP-binding protein	M35522
	$T_{12}N_{17}(AAT)_{10}(AAC/T)_8$	3′-NTR		
Ip082	$(GT)_8Nn(GT)_{11}$	5′-NTR	Unknown	This study
Ip085	$(CA)_{16}$	5′-NTR	Unknown	This study
Ip093	$(CA)_{12}$	5′-NTR	MHR23B	X92411
Ip095	$(TCAA)_6$	3′-NTR	Gonadotropin α subunit	AF112190
Ip097	$(GA)_9$	3′-NTR	Growth hormone	S69215

microsatellites may have a twofold advantage: serving as better anchorage points for comparative gene mapping, and perhaps allowing easier PCR optimization with unique gene sequences flanking microsatellites. In general, microsatellites are type II markers. By identifying microsatellite markers in transcribed sequences, they can be associated with a gene of known function and thus can be converted into type I markers.

Identification of transcribed microsatellites can be an effective way to develop polymorphic EST markers. The mechanisms for polymorphism are different for EST and microsatellite markers. Length difference in EST is caused by deletion or insertion between the PCR primers whereas length difference in microsatellites is believed to be caused by difference in repeat numbers due to slippage of DNA polymerase (Tautz et al., 1986; Levinson and Gutman, 1987; Tautz and Schlotterer, 1994). Although insertion/deletion rates in gene sequences are low, the microsatellite repeats mutate at frequencies as high as 10^{-2} per generation (Dallas, 1992; Weber and Wong, 1993; Crawford and Cuthbertson, 1996). High polymorphic levels were observed with microsatellites in channel catfish (Waldbieser and Bosworth, 1997; Liu et al., 1998c). Thus, as transcribed microsatellites are identified, it is likely that polymorphic EST markers can be identified by amplifying the region of transcribed microsatellites. It may be difficult, however, to isolate large numbers of unique transcribed microsatellite clones because their representation in cDNA libraries is proportional to the expression levels of the genes with which they are associated. Abundantly expressed messages containing microsatellites may be overwhelmingly represented if a microsatellite-enriched library is constructed based on cDNA

libraries. Alternatively, they can be identified by screening microsatellite-enriched libraries with first-strand cDNA probes.

Codon Usage in Channel Catfish

Codon usage in channel catfish was analyzed with the sequenced ESTs as well as the gene sequences of channel catfish extracted from the GenBank databases. The unidentified ESTs were not analyzed because their ORFs were unidentified. The codon usage in channel catfish was similar to that of higher vertebrates such as humans (Table 4). This is consistent with what was found from medaka (Hirono and Aoki, 1997). Several differences were noticed between codon usage in channel catfish and humans. The usage of CCC, GCC, CGG, and GGG was much lower in channel catfish than in humans (Table 4). The trend was that the discriminated codons in channel catfish were all G/C-rich. That may result from the lower G/C content overall or regionally (also known as genome compartmentation) (Stenico et al., 1994) in channel catfish genome. This is consistent with our previous findings that the channel catfish genome was AT-rich as revealed by AFLP analysis (Liu et al., 1998b).

In total, 11,502 codons were included for analysis, of which 9,268 were extracted from the GenBank databases and 2,234 were from this EST analysis. To our best knowledge, this is the first thorough analysis of codon usage in catfish, an organism with which much genetic engineering is carried out (Dunham, 1996). In *E. coli* and yeast, the codons identified with their increased usage coincide with those predicted to be translationally optimal on the basis of the anticodon sequences and relative abundance of transfer RNAs (Gouy and Gautier, 1982; Ikemura, 1985; Sharp and Cowe, 1991). In the absence of such detailed knowledge of channel catfish tRNAs, only an inference by analogy can be made. If the inference is correct, the codon usage information of channel catfish will provide the basis for gene construction for genetic engineering in catfish. As some beneficial genes are isolated for sources other than fish, modification of sequences based on the codon usage in channel catfish would be preferable for efficient translation of the transgenes.

Correlation of Preferred Codon Usage, Expression Level, and Tissue-Specific Expression

The EST analysis using the pituitary cDNA library offers opportunities to analyze correlation of codon usage with expression levels and tissue-specific gene expression. Clearly, the channel catfish codon usage is not random, with 16 codons being preferably used (Table 4). This biased codon usage is even more biased (preferred) in ribosomal protein genes (Table 4), which are believed to be highly expressed (Senico et al., 1994). When sequences from the EST analysis representing other than pituitary hormones or ribosomal proteins were analyzed, the overall trend was that there was less codon bias in these genes than in ribosomal proteins, indicating correlation of codon usage bias in channel catfish as a function of gene expression levels (Sharp et al., 1993; Stenico et al., 1994; Pouwels and Leunissen, 1994; Kim and Lee, 1997).

Codon usage in the tissue-specific hormones such as prolactin, growth hormone, proopiomelanocortin, somatolactin, and gonadotropins (α subunit and two β subunits) was highly biased as compared with either that of all channel catfish genes (with 11,502 codons included) or that from the previous databases (with 9,268 codons included) (Table 4). Fifteen of the 16 codons with increased usage in channel catfish were more biased in the pituitary hormone genes (Table 4). Only CAG encoding Gln was not preferably used in hormone genes, but was preferably used in other genes (Table 4). Obviously, overall amino acid compositions affect usage of the 61 codons (Karlin and Mrazek, 1996). Pituitary hormones are highly cysteine-rich (5.6% vs. 2.6% in other genes). Many of the cysteines are from gonadotropins (Liu et al., 1997, and unpublished results). It is possible that the biased codon usage in the tissue-specific pituitary hormones is also related to high levels of expression of these genes in pituitaries.

Acknowledgments

This project was supported by the Department of Fisheries and Allied Aquacultures, the College of Agriculture, and the Office of the Vice President for Research, Auburn University grant 233078 to Z.J.L., and the Auburn University Competitive Research-Grant-in-Aid grant 213966 to Z.J.L., USDA NRICGP No. 98-35205-6738, and partially by a grant from BARD to Z.J.L., R.A.D., and Boaz Moav (US-2954-97) for isolating the catfish growth hormone cDNA. We thank Ping Li for her technical assistance. This paper is journal number 8-985975 of the Alabama Agricultural Experiment Station.

Table 4. Codon Usage Comparison

	Channel catfish (GenBank)	Channel catfish pituitary hormones	Channel catfish ribosomal proteins	Channel catfish other genes	Channel catfish (Total)	Human
Phe-TTT	15.6	10.3	11.0	16.8	15.0	16.3
TTC	24.5	33.6	18.9	18.5	24.4	21.3
Leu-TTA	4.9	6.5	4.7	3.3	4.8	6.3
TTG	12.0	10.3	15.7	13.5	12.2	11.5
CTT	9.7	12.9	22.0	8.4	10.5	11.7
CTC	19.2	14.2	6.3	16.8	18.1	19.5
CTA	7.5	6.4	3.1	3.3	7.0	6.4
CTG	41.6	49.1	51.9	35.4	41.9	40.5
Ile-ATT	15.0	12.9	17.3	6.7	14.3	15.8
ATC	28.5	23.2	31.5	32.0	28.3	23.7
ATA	7.1	0.0	1.6	3.3	6.0	6.7
Met-ATG	21.5	33.6	26.7	27.0	22.8	22.6
Val-GTT	18.3	11.6	7.9	10.1	17.0	10.7
GTC	16.0	15.5	22.0	16.8	16.1	15.6
GTA	7.5	3.9	3.1	11.8	7.1	6.6
GTG	29.2	31.0	29.9	25.3	28.9	29.9
Ser-TCT	20.5	16.8	14.2	11.8	19.7	14.1
TCC	13.4	18.1	7.9	18.5	13.5	17.7
TCA	15.5	10.3	14.2	8.4	14.5	10.9
TCG	5.1	5.1	1.6	6.7	4.9	4.4
AGT	12.6	9.0	7.9	5.0	11.7	11.2
AGC	17.5	32.3	20.4	21.9	19.1	19.1
Pro-CCT	14.5	20.7	29.9	10.1	15.5	16.8
CCC	9.1	6.5	14.2	8.4	9.1	20.0
CCA	13.0	15.5	7.9	11.8	13.0	16.2
CCG	6.3	16.8	3.1	5.0	6.6	6.9
Thr-ACT	17.9	11.6	9.4	13.5	16.4	12.8
ACC	21.9	29.7	22.0	27.0	22.5	21.1
ACA	23.0	12.9	12.6	13.5	21.0	14.8
ACG	10.1	7.8	3.1	13.5	9.6	6.7
Ala-GCT	25.9	19.4	31.5	15.2	24.7	18.6
GCC	17.7	12.9	25.2	27.0	18.4	29.1
GCA	17.0	9.0	11.0	18.5	16.5	15.3
GCG	7.9	6.5	11.0	11.8	8.1	7.5
Tyr-TAT	12.8	11.6	7.9	15.2	12.5	12.3
TAC	20.3	20.7	26.8	18.5	20.3	17.0
His-CAT	7.4	6.5	7.9	5.0	7.1	9.7
CAC	12.8	18.1	15.7	11.8	13.5	14.6
Gln-CAA	12.1	14.2	4.7	13.6	12.0	11.4
CAG	34.5	18.1	22.0	50.6	33.6	33.8
Asn-AAT	16.0	16.8	12.6	23.6	16.3	16.6
AAC	25.1	25.8	20.4	20.2	24.4	21.0
Lys-AAA	26.4	23.2	28.3	23.6	26.2	23.2
AAG	33.2	34.9	59.8	54.0	36.7	33.9
Asp-GAT	22.8	19.4	17.3	13.5	21.7	22.0
GAC	23.1	25.8	15.8	25.3	22.9	26.9

Table 4. Continued

	Channel catfish (GenBank)	Channel catfish pituitary hormones	Channel catfish ribosomal proteins	Channel catfish other genes	Channel catfish (Total)	Human
Glu-GAA	20.8	15.5	9.4	15.2	19.4	27.9
GAG	38.1	47.8	29.9	45.6	38.9	40.7
Cys-TGT	11.2	23.3	12.6	10.1	12.1	9.5
TGC	12.4	32.3	14.2	5.0	13.6	12.8
Trp-TGG	15.2	6.5	7.9	27.0	14.8	13.3
Arg-CGT	6.7	6.5	12.6	3.3	6.9	4.6
CGC	9.5	14.2	7.9	6.7	9.7	11.0
CGA	3.9	5.2	9.3	6.7	4.2	5.9
CGG	4.1	5.1	3.1	5.0	4.2	11.3
AGA	10.7	5.2	29.9	6.7	11.7	10.8
AGG	9.9	12.9	31.5	10.1	11.3	10.9
Gly-GGT	16.4	10.3	23.6	10.1	16.2	11.1
GGC	15.5	18.1	28.3	35.4	17.6	23.8
GGA	20.6	14.2	17.3	28.7	20.5	16.9
GGG	9.2	6.5	3.1	8.4	8.5	16.6
Stop-TAA	2.4	ND	ND	ND	ND	0.7
TAG	0.9	ND	ND	ND	ND	0.5
TGA	0.6	ND	ND	ND	ND	1.2
Total no. of codons	9268	774	636	592	11502	4693617
No. of sequences	36	7	11	12	66	9808

Data for humans were from codon usage tabulated from GenBank (Re.97.0). AF016042, AF028014–AF028016, AF038156–AF038162, AF044402, AF053546–AF053549, AJ000267, AJ003122, J00944, J00945, M25903, M27230, M64400, M83111, S69215, S75715, U09721, U21163, U22460, U25703, U25704, U33273, U34410, U39193, U39194, U58505, U97019, X52617, X79482, Z68499. Channel catfish pituitary hormones in the analysis included Ip009, Ip044, Ip067, Ip086, Ip094, Ip095, Ip097. Channel catfish ribosomal proteins in the analysis included Ip027, Ip030, Ip035, Ip039, Ip043, Ip046, Ip049, Ip057, Ip061, Ip068, Ip095. Other genes in the analysis include Ip001, Ip015, Ip016, Ip024, Ip029, Ip032, Ip034, Ip058, Ip064, Ip076, Ip093, Ip099.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651–1656.

Aliyeva, E., Metz, A.M., and Browning, K.S. (1996). Sequences of two expressed sequence tags (EST) from rice encoding different cap-binding proteins. *Gene* 180:221–223.

Azam, A., Paul, J., Sehgal, D., Prasad, J., Bhattacharya, S., and Bhattacharya, A. (1996). Identification of novel genes from *Entamoeba histolytica* by expressed sequence tag analysis. *Gene* 181: 113–116.

Boguski, M.S., and Schuler, G.D. (1995). Establishing a human transcript map. *Nature Genet* 10:369–371.

Chakrabarty, D., Reddy, G.R., Dame, J.B., Almira, E.C., Laipis, P.J., Ferl, R.J., Yang, T.P., Rowe, T.C., and Schuster, S.M. (1994). Analysis of expressed sequence tags from *plasmodium falciparum*. *Mol Biochem Parasitol* 66:97–104.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S., and Myers, R.M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250:245–250.

Crawford, A.M., and Cuthbertson, R.P. (1996). Mutations in sheep microsatellites. *Genome Res* 6:876–879.

Dallas, J.F. (1992). Estimation of microsatellite mutation rates in recombinant strains of mouse. *Mamm Genome* 3:452–456.

Dunham, R.A. (1996). Contribution of genetically improved aquatic organisms to global food security. Presented at the International Conference on Sustainable Contribution of Fisheries to Food Security, Government of Japan and FAO, Rome, Italy.

Franco, G.R., Adams, M.D., Bento, S.M., Simpson, A.J.G., Venter, J.C., and Pena, S.D.J. (1995). Identification of new *Schistosoma*

mansoni genes by the EST strategy using a directional cDNA library. *Gene* 152:141–147.

Gong, Z., Hu, Z., Gong, Z.Q., Kitching, R., and Hew, C.L. (1994). Bulk isolation and identification of fish genes by cDNA clone tagging. *Mol Mar Biol Biotechnol* 3:243–251.

Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* 10:7055–7074.

Harless, J., Obermoeller, R.D., Walter, R.B., Svensson, R., Nairn, R.S., Vielkind, J.R., Kallman, K.D., and Morizot, D.C. (1995). Characterization of a tyrosine kinase–peptidase A synteny in linkage group XIII of *Xiphophorus* fishes (*Teleostei: Poeciliidae*): implications for vertebrate chromosome evolution. *J Hered* 86:235–240.

Hirono, I., and Aoki, T. (1997). Expressed sequence tags of medaka (*Oryzias latipes*) liver mRNA. *Mol Mar Biol Biotechnol* 6:345–350.

Hofte, H., Desprez, T., Amselem, J., Chiapello, H., Rouze, P., Caboche, M., Moisan, A., Jourjon, M.F., Charpenteau, J.L., Berthomieu, P., et al. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana. Plant J* 4:1051–1061.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2:13–34.

Inoue, S., Nam, B., Hirono, I., and Aoki, T. (1997). A survey of expressed sequence tags in Japanese flounder (*Paralichthys olivaceus*) liver and spleen. *Mol Mar Biol Biotechnol* 6:376–380.

Karlin, S., and Mrazek, J. (1996). What drives codon choices in human genes? *J Mol Biol* 262:459–472.

Karsi, A., Li, P., Dunham, R.A., and Liu, Z.J. (1998). Transcriptional activities in the pituitaries of channel catfish before and after induced ovulation by injection of carp pituitary extract as revealed by expressed sequence tag analysis. J Mol Endocrinol. 21:121–129.

Kim, C.H., Oh, Y., and Lee, T.H. (1997). Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. *Gene* 199:293–301.

Levinson, G., and Gutman, G.A. (1987). High frequency of short frameshifts in poly CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* 15:5323–5338.

Liu, Z.J., and Dunham, R.A. (1998). Considerations for genetic linkage and QTL mapping of ictalurid catfish. *Alabama Agr Exp Stat Cir Bull* 321:1–19.

Liu, Z.J., Li, P., Argue, B., and Dunham, R.A. (1997). Gonadotropin α subunit glycoprotein from channel catfish (*Ictalurus punctatus*) and its expression during hormone-induced ovulation. *Mol Mar Biol Biotechnol* 6:221–231.

Liu, Z.J., Li, P., Argue, B.J., and Dunham, R.A. (1998a). Inheritance of RAPD markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*), and their F_1, F_2 and backcross hybrids. *Anim Genet* 29:58–62.

Liu, Z.J., Nichols, A., Li, P., and Dunham, R.A. (1998b). Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*), and their F_1, F_2 and backcross hybrids. *Mol Gen Genet* 258:260–268.

Liu, Z.J., Tan G., Kucuktas, H., Karsi, A., Li, P., Zheng, X., Kim, S., and Dunham, R.A. (1998c). Evolutionary conservation of microsatellite loci between channel catfish (*Ictalurus punctatus*) and blue catfish (*I. furcatus*). *J Hered* 90:307–312.

O'Brien, S.J. (1991). Molecular genome mapping: lessons and prospects. *Curr Opin Genet Dev* 1:105–111.

Pouwels, P.H., and Leunissen, J.A.M. (1994). Divergence in codon usage of *Lactobacillus* species. *Nucleic Acids Res* 22:929–936.

Reeves, R.H., Rue, E., Yu, J., and Kao, F.T. (1998). *Stch* maps to mouse chromosome 16, extending the conserved synteny with human chromosome 21. *Genomics* 49:156–157.

Sambrook, J., Frisch, E.F., and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P.J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., et al. (1996). A gene map of the human genome. *Science* 274:540–546.

Sharp, P.M., and Cowe, E. (1991). Synonymous codon usage in *Saccharomyces cerevisiae. Yeast* 7:657–678.

Sharp, P.M., Stenico, M., Peden, J.F., and Lloyd, A.T. (1993). Codon usage: mutational bias, translational selection, or both? *Biochem Soc Trans* 21:835–841.

Stenico, M., Lloyd, A.T., and Sharp, P.M. (1994). Codon usage in *Caenorhabditis elegans:* delineation of translational selection and mutational biases. *Nucleic Acids Res* 22:2437–2446.

Tautz, D., and Schlotterer, C. (1994). Simple sequences. *Curr Opin Genet Dev* 4:832–837.

Tautz, D., Trick, M., and Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322:652–656.

Tilghman, S.M. (1996). Lessons learned, promises kept: a biologist's eye view of the Genome Project. *Genome Res* 6:773–780.

Waldbieser, G.C., and Bosworth, B.G. (1997). Cloning and characterization of microsatellite loci in channel catfish, *Ictalurus punctatus. Anim Genet* 28:295–298.

Wang, D.G., Fan, J., Siao, C., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., and Lander, E.S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–1082.

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J., and Sulston, J. (1992). A survey of expressed genes in *Caenorhabditis elegans. Nature Genet* 1:114–123.

Weber, J.L., and Wong, C. (1993). Mutation of human short tandem repeats. *Hum Mol Genet* 2:1123–1128.

Wolfsberg, T.G., and Landsman, D. (1997). A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res* 25:1626–1632.