



Comparative transcriptome analysis of testes and ovaries for the discovery of novel genes from Amur sturgeon (*Acipenser schrenckii*)

S.B. Jin^{2*}, Y. Zhang^{1*}, X.L. Dong¹, Q.K. Xi², D. Song², H.T. Fu² and D.J. Sun¹

¹Heilongjiang River Fisheries Research Institute,
Chinese Academy of Fishery Sciences, Haebin, China

²Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization,
Ministry of Agriculture, Freshwater Fisheries Research Center,
Chinese Academy of Fishery Sciences, Wuxi, China

*These authors contributed equally to this study.

Corresponding author: D.J. Sun
E-mail: 13845059350@163.com

Genet. Mol. Res. 14 (4): 18913-18927 (2015)

Received August 16, 2015

Accepted October 30, 2015

Published December 28, 2015

DOI <http://dx.doi.org/10.4238/2015.December.28.40>

ABSTRACT. Sturgeons (*Acipenser schrenckii*) are of high evolutionary, economic, and conservation value, and caviar is one of the most valuable animal food products in the world. The Illumina HiSeq2000 sequencing platform was used to construct testicular and ovarian transcriptomes to identify genes involved in reproduction and sex determination in *A. schrenckii*. A total of 122,381 and 114,527 unigenes were obtained in the testicular and ovarian transcriptomes, respectively, with average lengths of 748 and 697 bp. A total of 46,179 genes were matched to the non-redundant nr database. GO (31,266), KEGG (39,712), and COG analyses (20,126) were performed to identify potential genes and their functions. Twenty-six gene families involved in reproduction and sex determination were identified from the *A. schrenckii* testicular and ovarian transcriptomes based on functional annotation of non-redundant transcripts and comparisons with the published literature. Furthermore, 1309 unigenes

showed significant differences between the testes and ovaries, including 782 genes that were up-regulated in the testes and 527 that were up-regulated in the ovaries. Eleven genes were involved in reproduction and sex determination mechanisms. Furthermore, 19,065 simple sequence repeats (SSRs) were identified in the expressed sequence tagged dataset, and 190,863 and 193,258 single nucleotide polymorphisms (SNPs) were obtained from the testicular and ovarian transcriptomic databases, respectively. This study provides new sequence information about *A. schrenckii*, which will provide a basis for the further study of reproduction and sex determination mechanisms in *Acipenser* species. The potential SSR and SNP markers isolated from the transcriptome may shed light on the evolution and molecular ecology of *Acipenser* species.

Key words: *Acipenser schrenckii*; Testes; Ovaries; Transcriptome; Reproduction

INTRODUCTION

Amur sturgeon (*Acipenser schrenckii*), also known as a living fossil, is one of the most primitive *Actinopterygii* species. Amur sturgeon is a species of fish belonging to the Acipenseridae family that can be found in the Amur River, Songhua River, and Heilong River (Li et al., 2012). Sturgeon species have become very popular in Chinese aquaculture for their considerable value in the economy and in ecology. Sturgeons accounted for 15% of the total sturgeon production in Chinese aquaculture (Wei et al., 2011). In recent years, sturgeons have received a considerable amount of interest in many countries (Wang et al., 2009a). However, in the last few decades, increasingly intense fishing practices have caused a dramatic decline in the distribution and population of wild stocks. Most sturgeon species are at risk of extinction; therefore, the International Union for Conservation of Nature has listed sturgeons as one of the most critically endangered group of species. In addition, the high economic value of these animals is a further advantage for the transcriptome characterization of sturgeon species. Sturgeon eggs are sold as caviar and represent one of the most highly prized foods of animal origin (Ludwig, 2008). Female sturgeons, which can produce an overwhelming number of eggs, are desirable in the sturgeon aquaculture program. However, an overriding problem for aquaculture caviar producers is that 50% of the animals are profitless males, which need to be discarded from production as quickly as possible in order to minimize expenditure and maximize space. Thus, in order to maintain their sustainable development, understanding the reproduction and sex-differentiation mechanism of *A. schrenckii* is an urgent topic of research.

The transcriptome includes all the transcripts, mRNA, and non-coding RNA in one cell or in a population of cells during a specific developmental stage or in response to a particular physiological condition, and is determined using high-throughput technology (Mardis, 2008). Transcriptome analysis provides a foundation for gene structure and function research, and reveals the timing of gene expression and associated regulation mechanisms. The development of next-generation sequencing (NGS) technologies allows the acquisition of more sequence data per run at a substantially lower cost than are obtained using long-read technologies (Huang and Marth, 2008). Compared with microarray methods, NGS, developed by Illumina/Solexa, is able to generate over

one billion bases of high quality sequences per run at less than 1% of the cost of capillary-based methods, and is expected to dominate the future analysis of eukaryotic transcriptomes (Cloonan and Grimmond, 2008; Morozova and Marra, 2008; Wang et al., 2009b). Transcriptome analysis is now being widely applied to elucidate genetic factors that confer economically significant traits and/or phenotypes to the management of genetic diversity in cultured fish species. To date, only a limited number of unique sequences have been obtained for *A. schrenckii* and *Acipenser* species. In this study, we investigated the gene expression profile in testes and ovaries of *A. schrenckii*.

We aimed first to generate the first *A. schrenckii* comparative transcriptome analysis using cDNA reverse-transcribed from mRNA isolated from the testes and ovaries, in order to identify candidate genes that are involved in reproduction and that differ in their expression between males and females. Our second goal was to obtain a global view of the molecular mechanism(s) that underlie how gene expression influences reproduction and gonad development in *A. schrenckii*.

MATERIAL AND METHODS

Experimental fishes

Amur sturgeons were collected from the Engineering and Technology Center of Sturgeon Breeding and Cultivation of Chinese Academy of Fishery Science in Beijing, China. Eight healthy *A. schrenckii* (four males and four females) with a wet weight ranging from 12.7 to 28.5 kg (average = 22.36 kg), and a total length ranging from 130.2 to 164.7 cm (average = 158.67 cm) were obtained for transcriptome construction. In order to prevent total RNA degradation, the testes and ovaries were collected and immediately frozen in liquid nitrogen until RNA was extracted for transcriptome sequencing.

RNA isolation for RNA-seq

The prepared testes and ovaries were pooled to provide sufficient RNA for transcriptome sequencing. Total RNA was extracted using the UNIQ-10 Column Trizol Total RNA Isolation Kit (Sangon) following the manufacturer protocol. To confirm the purity of the RNA sample, the OD_{260/280} and OD_{260/230} were measured, and ranged from 1.8 to 2.0 and >2.0, respectively. The total concentration of the RNA sample was adjusted to >5 mg to guarantee transcriptome quality. RNA integrity was confirmed using 2100 Bioanalyzer (Agilent Technologies, Inc.) with a minimum RNA integrity number (RIN) value of 7.0 and an ABI StepOnePlus Real-Time PCR System. The samples used for transcriptome analysis were prepared using a TruseqTM RNA Sample Prep Kit (Illumina) according to the manufacturer recommendation. Briefly, mRNA was isolated from total RNA using oligo (dT) magnetic beads. mRNA was cut into short fragments by the addition of fragmentation buffer in a Thermomixer at a suitable temperature. First-strand cDNA was synthesized using random hexamer-primers, using these short fragments as templates. RNaseH, buffer, dNTPs, and DNA polymerase I were used to synthesize second-strand cDNA. Short fragments were purified with a PCR extraction kit (Takara Bio, Inc.). Sequencing adapters were ligated to form short fragments and then resolved by agarose gel electrophoresis. Fragments were selected and purified, and subsequently PCR amplified to create the final cDNA library template.

Analysis of the transcriptome results

The transcriptome was sequenced using the Illumina HiSeq™ 2000 and each sample was sequenced twice. Four fluorescently labeled nucleotides and a specialized polymerase were used to determine the clusters, base-by-base in parallel. The size of the library was approximately 200 bp and both ends of the library were sequenced. The 200-bp raw paired-end reads were generated on the Illumina sequencing platform. Image deconvolution and quality value calculations were performed using Illumina GA pipeline v1.6. The raw reads were cleaned by removing adaptor sequences, empty reads, and low quality sequences (reads with unknown sequences 'N' or less than 25 bp). The clean reads were assembled into non-redundant transcripts using the Trinity, which has been developed specifically for the *de novo* assembly of transcriptomes using short reads. To obtain non-redundant transcripts, we removed short sequences (shorter than 100 bp in length) and partially overlapping sequences. The resulting sequences were used for BLAST searches and annotation against the Nr protein, the SwissProt, COG, and KEGG databases using an E-value cut-off of 10⁻²⁵. Functional annotation by GO terms (www.geneontology.org) was analyzed by the Blast2go software. The COG and KEGG pathway annotations were performed using Blast all software against the COG and KEGG databases.

Expression analysis

The abundance of each tag was normalized to one transcript per million, in order to allow comparison between testes and ovaries. The low quality sequences, including ambiguous nucleotides and adaptor sequences, were removed, in order to obtain cleaned raw reads. Calculation of unigene expression was conducted using RSEM software, which is an accurate and user-friendly software tool for quantifying transcript abundances from RNASeq data (Li, 2011). It is particularly useful for quantification with *de novo* transcriptome assemblies because it does not rely on the existence of a reference genome (Li, 2011).

RESULTS

Construction and characterization of the normalized cDNA library

The testicular and ovarian transcriptomes yielded approximately 60.8 million and 55.9 million 200-bp paired-end raw reads, respectively. After eliminating adapter sequences and filtering out low-quality reads (<25 bp in each read) before *de novo* assembly by the SeqPreq program, the testicular and ovarian transcriptomes yielded 6,051,265,200 and 5,592,689,200 high-quality transcriptome reads, respectively. The Q20 percentages and GC content of the testicular transcriptomes were 98.25 and 48.32%, respectively, whereas those of the ovarian transcriptome were 98.10 and 49.24%, respectively.

A. schrenckii is a non-model organism, for which there is no reference genome sequence. The raw data were therefore assembled *de novo* using the Trinity program. Totals of 278,257 and 255,179 contigs were assembled from the short, clean reads of the testicular and ovarian transcriptomes, respectively, with median lengths of 292 and 297 bp. From these contigs, 122,381 and 114,527 unigenes were obtained from the testicular and ovarian transcriptomes, respectively, with mean lengths of 748 and 697 bp because of alternative splicing; thus, two or more contigs

may be matched to one unigene (Table 1). Most of the unigenes were 1-200 bp (27.09% for the testicular transcriptome and 27.55% for the ovarian transcriptome) and 201-400 bp (27.48 for the testicular transcriptome and 27.81% for the ovarian transcriptome) in length, followed by 401-600 bp (11.11% for the testes and 11.26% for the ovaries) and 601-800 bp (7.10% for the testicular transcriptome and 7.00% for the ovarian transcriptome) (Figure 1).

Table 1. Summary of Illumina HiSeq 2000 assembly and analysis of *Acipenser schrenckii* transcriptomic sequences.

	Testes	Ovaries
Total unigenes	122,381	114,527
Total contigs	278,257	255,179
Total residues	60,512,652	55,926,892
Average length	748	697
Largest unigenes	8,126	8,317
Smallest unigenes	141	129

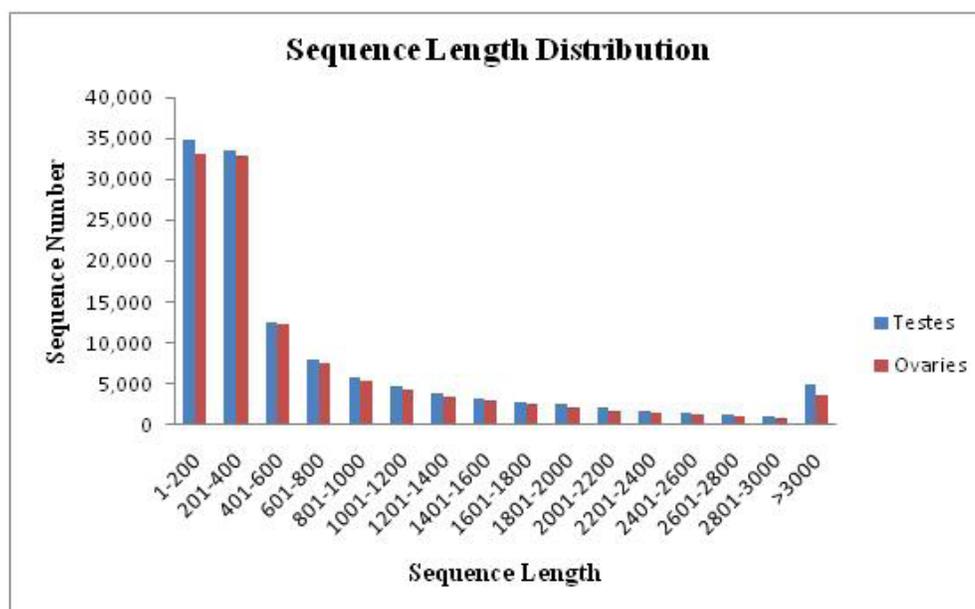


Figure 1. Contig length distribution of *Acipenser schrenckii* transcriptomic expressed sequence tags (ESTs).

Annotation analysis

To identify their putative functions, all unigenes were compared with the non-redundant protein database and nucleotide sequences in NCBI using Blastp and Blastx at an E-value of $<10^{-5}$ in the priority order of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, Gene Ontology (GO), and Cluster of Orthologous Groups (COG) database.

GO terms were assigned to 31,266 *A. schrenckii* contigs based on BLAST matches to proteins with known functions, including 45,126 sequences that were assigned to the molecular

function category, 132,172 to the cellular component category, and 203,572 to the biological process component. The matched contigs comprised 60 functional groups (Figure 2). The number of contigs in each GO term ranged from 1 to 25,099. “Cellular process”, “Cell part”, “Binding”, and “Single-organism process” had more than 20,000 abundant contigs, whereas there were also 12 functional groups in which the numbers of abundant contigs were less than 100.

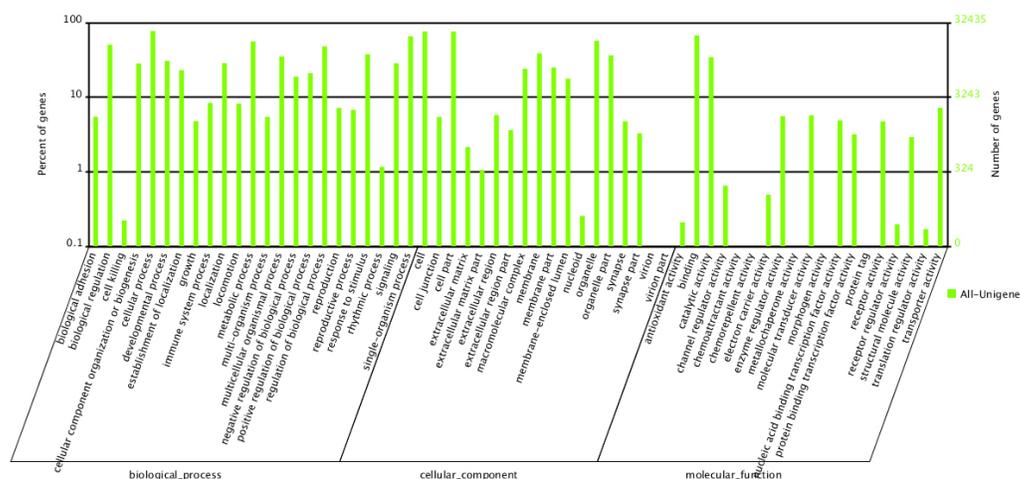


Figure 2. Gene ontology classification of non-redundant transcripts.

In addition, 20,126 unigenes matched known genes in the COG database. Based on their predicted functions, these unigenes were classified into 25 functional categories. Among these 25 functional categories, the cluster for “General function prediction only” represented the largest group with 7,913 unique sequences, followed by “Replication, recombination, and repair” (4123), “Translation, ribosomal structure, and biogenesis” (3513), “Cell cycle control, cell division, chromosome partitioning” (3145), and “Transcription” (3453). “Clusters for RNA processing and modification”, “Intracellular trafficking”, “Secretion vesicular transport”, “Extracellular structure and nuclear structure” represented the smallest of groups, in which the numbers of sequences were <200 (Figure 3).

The KEGG database can map unique sequence to defined metabolic pathways, in order to identify the role or functions of potential candidate transcripts in biological pathways in the ladybird. A total of 39,712 unigenes matched to the metabolic pathways in the KEGG Pathway database, accounting for 258 predicted metabolic pathways, and were grouped into “Amino acid metabolism”, “Genetic information processing”, “Cellular processes”, and “Environmental information processing”. The numbers of unique sequences in various pathways ranged from 1 to 3613. The main metabolic pathways of unique sequences in the *A. schrenckii* transcriptome were “Regulation of actin cytoskeleton”, “Focal adhesion”, “Pathways in cancer”, “Tight junction”, “Endocytosis”, and “MAPK signaling pathway”, in which the numbers of unique sequences exceeded 1200.

Reproduction- and development related genes

Twenty-six reproductive gene families were identified from the testicular and ovarian

transcriptomes constructed from *A. schrenckii*, based on the functional annotation of non-redundant transcripts (Table 2). Most of these functional genes were identified through comparisons with published data for other species (Chen et al., 2006; Hale et al., 2010; Vidotto et al., 2013). However, some were identified based on their GO classification and COG cluster. Double-sex and Mab-3-related transcription factor 1 (DMRT1), forkhead box L2 (FOXL2), Sex determining region of Y chromosome-related high mobility group-box gene 9 (SOX9), and Feminization-1 (FEM1) were screened from the *A. schrenckii* transcriptome database, and were identified as playing essential roles in reproduction and development mechanisms in vertebrates and mammals.

Table 2. Reproduction-related ESTs identified in the *Acipenser schrenckii* transcriptome.

Transcripts	Length (bp)	E-value	Accession No.	Hits
Dmrt1	248-1274	1.00E-06	gij 23428499 gb AY057061.1	7
Foxl2	254-2820	0	sp Q6VFT6 FOX_L2_PIG	5
SOX	200-1806	0	gij 58465464 gb AAW78521.1	112
SRY	236-4111	1.00E-08	gij 170595123 ref XP_001902255.1	30
GnRH	580-1146	3.00E-69	gij 465967294 gb EMP31298.1	9
Estrogen receptor	698-1679	0	gij 291190839 ref NP_001167409.1	17
FHL3	1244	1.00E-162	gij 125840523 ref XP_695478.2	1
FEM1	938-3510	1.00E-79	gij 348501380 ref XP_003438248.1	5
ATRX	229-4615	1.00E-10	gij 126631742 gb AAI34027.1	13
Androgen receptor	1562-2678	1.00E-80	gij 224067351 ref XP_002193622.1	11
Wnt	541-1986	0	gij 147903413 ref NP_001083829.1	58
FGFR	1750-3386	1.00E-47	gij 190888207 gb ACE95859.1	29
Chitinase	273-2059	1.00E-33	gij 327271191 ref XP_003220371.1	11
Cathepsin	221-1715	0	gij 432870116 ref XP_004071815.1	53
Cytochrome	441-4113	0	gij 111115685 gb DQ786406.1	132
Kazal-type serine proteinase inhibitor	393	2.00E-09	gij 58081895 emb CAI46283.1	1
Spermatogenesis-associated protein	531-2573	0	gij 218931210 ref NP_001073660.2	27
Carboxylesterase	1182-3587	1.00E-54	sp Q99K10 NLGN1_MOUSE	42
Peritrophin	902	5.00E-06	gij 307180562 gb EFN68518.1	1
Heat Shock Protein	589-2461	0	gij 432964405 ref XM_004086882.1	86
Gametocyte-specific factor	755	1.00E-53	gij 225714933 gb BT078889.1	1
VASA-like protein	731-779	4.00E-08	gij 270610556 gb ACZ92304.1	2
Vitellogenin	200-616	1.00E-08	gij 305690594 gb ADM64616.1	6
Vitellogenin receptor	1236	1.00E-14	gij 185135214 ref NP_001117847.1	1
Kunitz-type protease inhibitor 1	246-4825	2.00E-07	gij 115527831 gb AAI24740.1	17
Oocyte zinc finger protein	213-1134	1.00E-21	gij 326680600 ref XP_003201569.1	5

Genes differentially expressed between the testes and ovaries were investigated. A total of 1309 unigenes showed significant differences between the testicular and ovarian transcriptomes, including 782 up-regulated unigenes in the testes and 527 in the ovaries. With a view to applicability in caviar production, these 527 genes in the ovary, which may promote ovarian development, were of interest in this study (Figure 4). Among these significantly differentially expressed genes, 11 reproductive and developmental genes were identified (Table 3).

Identification of SNP and SSR markers

A large number of simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers were obtained. A total of 19,542 SSRs were obtained in the transcriptomic dataset, including 53.02% di-nucleotide, 35.62% trinucleotide, and 11.36% tetra/penta/hexanucleotide repeats (Figure 5). Among the di-nucleotide repeat motifs, (GT/TG)_n, (GA/AG)_n, and (CA/AC)_n

were the three dominant di-nucleotide types ($N > 2000$). Among the tri-nucleotide repeat motifs, (GGA/GAG/AGG) $_n$ had a frequency of 18.01% and was the most common type, followed by (GCA/CAG/AGC) $_n$, (GCT/CTG/TGC) $_n$, and (ATA/TAA/AAT) $_n$ ($N > 650$). There was a bias towards tri-nucleotide repeat motifs composed of G and A.

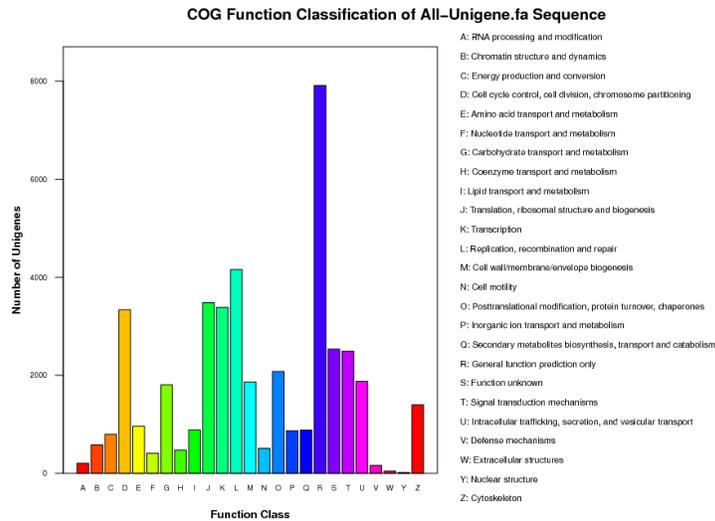


Figure 3. Cluster of orthologous groups (COG) classification of putative proteins.

Table 3. Differentially expressed genes between testes and ovaries transcriptome of *Acipenser schrenckii*.

Unigene name of top BLASPX hit	Length (bp)	E-value	Expression level	
			Testes	Ovaries
Growth hormone	977	1.00E-146	4223	14590
SRY-related box 10	1271	3.00E-46	77	151
Cathepsin F-like	1669	0	299	744
Cytochrome P450	658	9.00E-61	26	3
Cytochrome P4501B	2665	0	9	110
Heat shock protein 10	765	2.00E-42	183	67
Vasa-like protein	779	4.00E-08	59	118
Serine protease HTRA1B precursor	1822	0	36	78
KRAB domain-containing zinc finger protein	3533	2.00E-29	23	59
Spermatogenesis	5042	0	36	74
Carboxylesterase activity	2456	9.00E-116	311	121

SNPs were identified from alignments of multiple sequences used for contig assembly. By excluding those with a mutation frequency of bases less than 1%, a total of 190,863 and 193,258 SNPs, respectively, were obtained from the testicular and ovarian transcriptomic databases (Figure 6). In the testicular transcriptome, 113,226 were putative transitions (Ts) and 77,637 were putative transversions (Tv), giving a mean Ts: Tv ratio of 1.46:1, while in the ovarian transcriptome, 114,304 were putative transitions (Ts) and 78,954 were putative transversions (Tv), giving a mean Ts: Tv ratio of 1.45:1. The AG/GA and CT/TC SNPs were the most common in both transcriptomes, whereas CA/AC, AT/TA, CG/GC, and TG/GT types were the most infrequent SNP types, because of differences in the base structure and the number of hydrogen bonds between different bases.

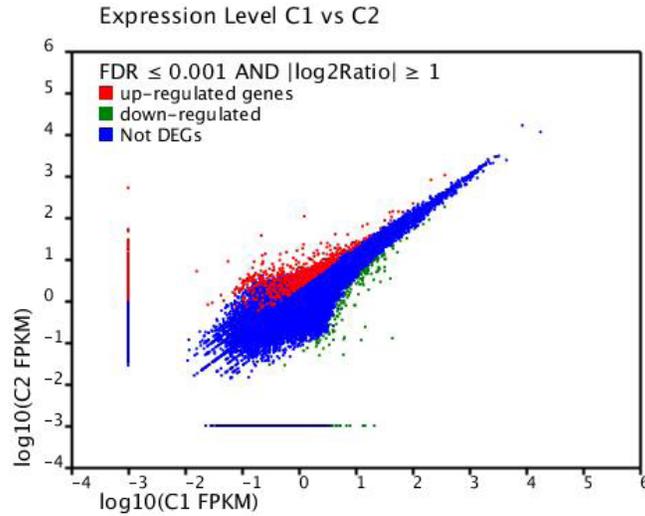


Figure 4. Differentially expressed gene (DEG) analysis of unigenes in the testes and ovaries of *Acipenser schrenckii*.

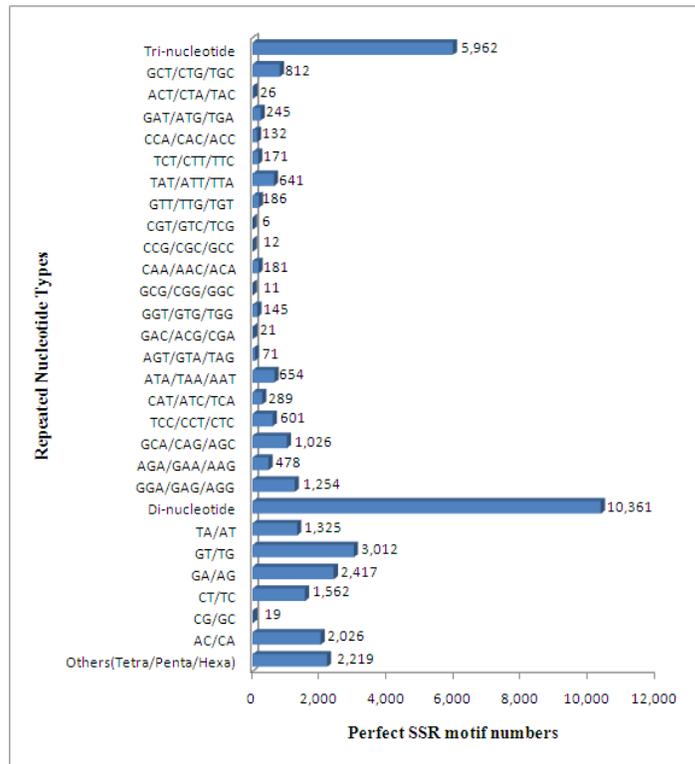


Figure 5. Distribution of simple sequence repeat (SSR) nucleotide classes among different nucleotide types found in the transcriptome of *Acipenser schrenckii*.

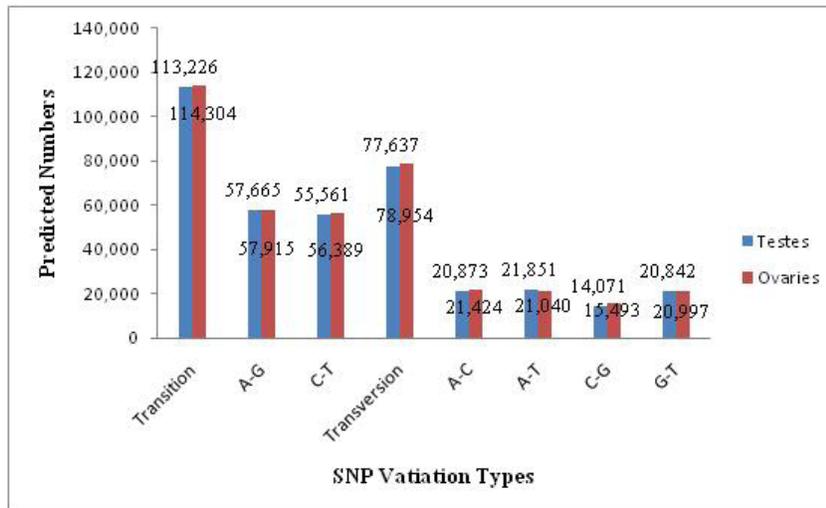


Figure 6. Distribution of putative SNPs in the *Acipenser schrenckii* transcriptome.

DISCUSSION

Transcriptome efficiency

Prior to this study, only 123 cDNAs had been cloned from testicular and ovarian cDNA libraries from *A. schrenckii* (Chen et al., 2006). A gonad transcriptome for all *Acipenser* species was constructed for *Acipenser fulvescens* using Roche 454 GS-FLX. A total of 20,471 unique sequences were obtained with a mean length of 186 bp (Hale et al., 2010). A cDNA library from gonad and brain was constructed for Adriatic sturgeon containing over 55,000 high quality ESTs in order to determine sex-related genes (Vidotto et al., 2013). The lack of a transcriptomic database has delayed studies in *A. schrenckii*, and even in *Acipenser* species. In the current study, we constructed the first testicular and ovarian transcriptomic database for *A. Schrenckii* using the Illumina HiSeq2000 NGS sequencing platform. The current study generated markedly more genes than was generated by previous studies, taking advantage of the capabilities of Illumina HiSeq2000 NGS, which can achieve more high-throughput sequencing and provide more candidate genes. Furthermore, the average length of each unigene is much longer than that of unigenes identified in previous studies after clustering and de novo assembly. This facilitates further studies on these unigenes, including RT-PCR and western blot analyses, to identify patterns of gene expression. The large numbers of unigenes in this study provide an abundant transcriptome database for use in future analyses of genes related to reproduction and sex determination in *A. schrenckii*, and may advance such studies conducted on all *Acipenser* species. Therefore, this transcriptome database accelerates our understanding of the reproductive and sex determination mechanisms in *A. schrenckii*, and in other *Acipenser* species.

Annotation analysis

The total numbers of GO terms and COG categories were much higher than the number of unique sequences, because many contigs can be assigned to more than one GO term or

category. GO and COG can provide a structured and controlled vocabulary for describing gene products. Reproduction and development-related genes of *A. schrenckii* were most likely to be contained in functional groups in “GO assignments”, including “Reproduction”, “Developmental process”, “Reproductive process”, and “Growth”, and functional categories in “COG analysis”, such as “The general functional prediction only” and “Energy production and conversion”. In the current study, genes in these functional groups were significantly more abundant than those identified in previous studies in *Acipenser* species (Chen et al., 2006; Hale et al., 2010; Vidotto et al., 2013), providing basic information to aid the further analysis of reproduction and development mechanisms in *A. schrenckii* and in other *Acipenser* species.

Furthermore, KEGG analysis dramatically advances research on the relationship between different genes in the transcriptome. A large number of genes involved in reproduction exist in these pathways, including DMRT1, FOXL2, and SOX9. Considering the growth process of *A. schrenckii*, identifying these pathways permits the analysis of reproduction and sex determination mechanisms in *A. schrenckii*, although not all of the major genes reported in the putative KEGG pathways were found in the current study.

Genes related to reproduction and sex determination

Research into reproduction and sex determination is needed to maintain the sustainable development for *A. schrenckii*, based on the high economic value of this species. Some homologous genes, which are reported to be involved in reproduction and sex determination in vertebrates, were found from the testicular and ovarian transcriptome.

DMRT1 is required for mammalian testis differentiation, and is reported to be a sexual regulator in worm and fly (Raymond et al., 2000). Sex in birds and chicken is determined by both heterogamous (ZW) and homogamous (ZZ) genotypes. The Z chromosome is the male-related sex chromosome, whereas the W chromosome is the female-related sex chromosome. In all birds, previous studies have confirmed that DMRT1 is located on the Z chromosome but is absent from the W chromosome. In chicken embryos, although DMRT1 is expressed in the early bipotential gonad, it is more highly expressed in males than in females (Raymond et al., 1999). Knockdown of DMRT1 in early chicken embryos can result in the feminization of male gonads (Smith et al., 2009). Recent research showed that DMRT1, which is located on the Z chromosome, plays an essential role in male development in *Cynoglossus semilaevis* (Chen et al., 2014). A DMRT1 homologous transcript was identified in the transcriptome database in the present study. However, there was no significant difference in its expression between the testes and ovaries. Therefore, the function of this gene in *A. schrenckii* requires further study.

Foxl2 is a member of the forkhead/HNF-3-related family of transcription factors. Its expression is detectable in the mammalian ovaries at the moment of gonadal determination (Crespo et al., 2013). Females lacking the Foxl2 gene appear male. Foxl2 knockout in mature female mice causes them to develop testes, although oocytes are still formed (Uda et al., 2004). Mutations in this gene have been shown to cause blepharophimosis and premature ovarian failure (Ottolenghi et al., 2005). A recent study showed that Foxl2 suppressed testicular differentiation mainly through repression of the SOX9 regulatory element (Uhlenhaut et al., 2009). In vertebrates, Sox9 plays an essential role in the sex determination and developmental process. Sox9 has diverse functions in mammals, including the formation of cartilaginous tissues and testis differentiation. Sox9 promotes the development of male sexual organs through the production of anti-Müllerian hormone (AMH),

which can inhibit the development of a female reproductive system (De Santa Barbara et al., 1998). Two different Sox9 isoforms, Sox9a and Sox9b, were found in *Danio rerio*, based on the analysis of differentially expressed genes. Sox9a was only expressed in the testes, whereas Sox9b was only expressed in the ovarian developmental process (Liang et al., 2014). Two different Sox9 isoforms were also found in *Oryzias latipes*, *Salmo irideus*, and *Pelteobagrus fulvidraco*. Both Foxl2 and Sox9 transcripts were found in the transcriptome in the present study. Based on the analysis of differentially expressed genes, Foxl2 was found to be more highly expressed in the ovaries, whereas Sox9 expression was higher in the testes. This implies that these two genes may have similar functions to those found in other vertebrate and mammalian species. The relationship between these two genes needs further research.

FEM1 was first reported to regulate signal-transduction in *Caenorhabditis elegans*, in which it has diverse functions in the sex determination pathway (Gaudet et al., 1996). The expression of a single FEM1 transcript and protein led to no significant difference in either sex, suggesting that its activity was regulated post-transcriptionally and post-translationally (Gaudet et al., 1996). Its homologs, including FEM1A, FEM1B, and FEM1C, have been found to be involved in sex determination in both human and the house mouse (*Mus musculus*). In this study, FEM1A, FEM1B, and FEM1C transcripts were identified, suggesting that FEM1 may play an important role in sex determination in *A. schrenckii*.

Vitellogenin is the major precursor of the egg-yolk proteins, which is primarily functioned to form yolk protein, providing nutrients for the developing embryo and larvae (Hook et al., 2007). However, many other biological functions of vitellogenin have also been reported, including temporal division of labor and foraging specialization, regulation of hormonal dynamics and change in gustatory responsiveness (Nelson et al., 2007). Vitellogenin is synthesized in different tissues in different species, and eventually accumulates in the ovary, where it promotes ovarian development (Jeon et al., 2010). Vitellogenin was recently reported to play a role in the host immune system in fish (Garcia et al., 2010). In the present study, vitellogenin and the vitellogenin receptor were not found to be expressed in testes, whereas there was moderate expression in the ovary, suggesting that vitellogenin may be involved in ovarian development. In addition, different isoforms of vitellogenin were found in this study, implying that vitellogenin may have diverse biological functions in *A. schrenckii* and that further research is needed.

Furthermore, many other reproduction-related genes were discovered in the transcriptome. Cathepsin L is involved in yolk processing during vitellogenesis in rainbow trout (Kwon et al., 2001). Cathepsin C was reported to play essential roles in the final stages of oocyte maturation in the kuruma prawn *Marsupenaeus japonicus* (Qiu et al., 2008). Cathepsin D was found to have high tissue-specific expression in the testis in the rat male reproductive system (Burdan et al., 2006).

Identification of differentially expressed genes involved in reproduction

Growth hormone (GH) is a peptide hormone that stimulates growth, cell reproduction, and regeneration in humans and other animals. GH promotes growth by increasing the concentration of glucose and free fatty acids (Ranabir and Reetu, 2011). In addition, GH also promotes growth by stimulating the production of insulin-like growth factor 1 (IGF-1) (Scarth, 2006). IGF-1 is generated in target tissues, making it both an endocrine and autocrine hormone (Jansen et al., 1983). In this study, GH was found to have significantly higher expression in ovaries than in testes, implying that GH has vital functions in promoting growth in females.

A vasa-like protein transcript was found to be significantly more highly expressed in ovaries than in testes. VASA belong to the DEAD-box family of ATP-dependent RNA helicases, which have diverse functions in germ cell development (Lasko and Ashburner, 1988). The product of this gene is required maternally for germplasm assembly and completion of oogenesis. VASA is considered a reliable molecular marker that can be used to trace the origin, migration, and differentiation of primordial germ cells because it is specifically and stably expressed in germ cells in most animals (Chang et al., 2002). Higher expression in the ovaries suggests that the vasa-like protein may also have an important role in the promotion of germplasm development and in the completion of oogenesis in *A. schrenckii*.

Cytochrome P450 (CYP) belongs to a superfamily of proteins that contain a heme cofactor. CYP enzymes have been identified in all living organisms (Lamb et al., 2009) except for *E. coli* (Sigel et al., 2007). Most CYP enzymes require a protein partner to deliver one or more electrons to reduce iron and eventually oxygen. The expression levels of different types of CYPs varied in the testes and ovaries of *A. schrenckii*. For example, Cytochrome P450 is more highly expressed in testes, whereas cytochrome P4501B is more highly expressed in ovaries, implying that members of the CYP family have diverse functions in *A. schrenckii*.

Identification of SSR and SNP markers

SSRs, or microsatellites, are polymorphic loci present in genomic DNA that consist of repeated core sequences of 2-6 base pairs in length (Queller et al., 1993). SNPs represent the most common type of variation in the genome. SNPs provide the best genome coverage for analyzing the performance and production of traits. Genomes with high-density SNP coverage are powerful tools for whole-genome association studies because they permit the detection of linkage disequilibrium (Wang et al., 1998). Thus, SSR and SNP markers have been widely used in molecular genetic research, such as evolution, molecular ecology, and quantitative trait loci (QTL) analysis.

Considering the advantages of SSRs and SNPs, the development of such markers to analyze the evolution, molecular ecology, and QTLs from *A. schrenckii* is desirable. The potential markers identified in this study will provide an invaluable resource for such studies. SSR and SNP markers will be identified from the transcriptomic database in future studies.

In conclusion, a comparative testicular and ovarian transcriptome of *A. schrenckii* was constructed in this study. This provides new sequence information for *A. schrenckii*, which will provide a basis for further genetic studies of *Acipenser* species. Many genes related to reproduction and development were identified in this study, which improves our understanding of reproductive and developmental mechanisms in *A. schrenckii* and *Acipenser* species. Potential SSR and SNP markers isolated from the transcriptome may shed light on the evolution and molecular ecology of *Acipenser* species. These data will advance research in the field of *Acipenser* molecular genetics.

Data deposition

The IlluminaHiSeq2000 reads of *A. schrenckii* were submitted to http://www.hrfri.ac.cn/news_nei.asp?newsid=1508&SortID=12&SortPath=0,12.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank International Science Editing for their help revising the English. Research supported by the National Natural Science Foundation of China (Grant #31201971) and the National Science & Technology Supporting Program of the 12th Five-year Plan of China (Grant #2012BAD25B10).

REFERENCES

- Burda F, Szumilo J, Dudka J, Dąbrowski A, et al. (2006). The activity and immunoexpression of cathepsin D in rat male reproductive organs. *Folia Morphol.* 65: 111-115.
- Chang CC, Dearden P and Akam M (2002). Germ line development in the grasshopper *Schistocerca gregaria*: vasa as a marker. *Dev. Biol.* 252: 100-118.
- Chen J, Wang B, Wu W, Liang B, et al. (2006). Cloning and screening of differentially expressed genes from testis and ovaries of Amur sturgeon. *J. Fish. Sci. China* 13: 8-12.
- Chen SL, Zhang GJ, Shao CW, Huang QF, et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46: 253-260.
- Cloonan N and Grimmond SM (2008). Transcriptome content and dynamics at single nucleotide resolution. *Genome Biol.* 9: 234.
- Crespo B, Lan-Chow-Wing O, Rocha A, Zanuy S, et al. (2013). foxl2 and foxl3 are two ancient paralogs that remain fully functional in teleosts. *Gen. Comp. Endocrinol.* 19: 481-93.
- De Santa Barbara P, Bonneaud N, Boizet B, Desclozeaux M, et al. (1998). Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. *Mol. Cell Biol.* 18: 6653-6665.
- Garcia J, Munro ES, Monte MM, Fourier MC, et al. (2010). Atlantic salmon (*Salmo salar* L.) serum vitellogenin neutralises infectivity of infectious pancreatic necrosis virus (IPNV). *Fish Shellfish Immunol.* 29: 293-297.
- Gaudet J, VanderElst I and Spence AM (1996). Post-transcriptional regulation of sex determination in *Caenorhabditis elegans*: widespread expression of the sex determining gene fem-1 in both sexes. *Mol. Biol. Cell* 7: 1107-1121.
- Hale MC, Jackson JR and DeWoody JA (2010). Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (*Acipenser fulvescens*). *Genetica* 138: 745-756.
- Hook SE, Skillman AD, Small JA and Schultz IR (2007). Temporal changes in gene expression in rainbow trout exposed to ethinyl estradiol. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 145: 73-85.
- Huang W and Marth G (2008). A genome assembly viewer for next-generation sequencing technologies. *Genome Res.* 18: 1538-1543.
- Jansen M, van Schaik F, Ricker A, Bullock B, et al. (1983). Sequence of cDNA encoding human insulin-like growth factor I precursor. *Nature* 306: 609-611.
- Jeon JM, Lee SO, Kim KS, Baek HJ, et al. (2010). Characterization of two vitellogenin cDNAs from a Pandalus shrimp (*Pandalopsis japonica*): expression in hepatopancreas is down-regulated by endosulfan exposure. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 157: 102-112.
- Kwon JY, Prat F, Randall C and Tyler CR (2001). Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout (*Oncorhynchus mykiss*). *Biol. Reprod.* 65: 1701-1709.
- Lamb DC, Lei L, Warrilow AG, Lepesheva GI, et al. (2009). The first virally encoded cytochrome P450. *J. Virol.* 83: 8266-8269.
- Lasko PF and Ashburner M (1988). The product of the *Drosophila* gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* 335: 611-617.
- Li D (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.
- Li D, Liu Z and Xie C (2012). Effect of stocking density on growth and serum concentrations of thyroid hormones and cortisol in Amur sturgeon, *Acipenser schrenckii*. *Fish Physiol. Biochem.* 38: 511-520.
- Liang QY, Wang XY, Sun DJ and Yan JZ (2014). Dynamic expression patterns of two Sox9 genes in developing embryo and gonads of zebra fish. *Shandong Agric. Sci.* 46: 20-24.
- Ludwig A (2008). Identification of *Acipenseriformes* species in trade. *J. Appl. Ichthyol.* 24: 2-19.
- Mardis E (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24: 133-141.
- Morozova O and Marra MA (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255-264.

- Nelson CM, Ihle KE, Fondrk MK, Page RE, et al. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol.* 5: e62.
- Ottolenghi C, Omari S, Garcia-Ortiz JE, Uda M, et al. (2005). Foxl2 is required for commitment to ovaries differentiation. *Hum. Mol. Genet.* 14: 2053-2062.
- Qiu LH, Jiang SG, Huang JH, Wang W, et al. (2008). Molecular cloning and mRNA expression of cathepsin C gene in black tiger shrimp (*Penaeus monodon*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 150: 320-325.
- Queller DC, Strassman JE and Hughes CR (1993). Microsatellites and kinship. *Trends Ecol. Evol.* 8: 285-288.
- Ranabir S and Reetu K (2011). Stress and hormones. *Indian J. Endocrinol. Metab.* 15: 18-22.
- Raymond CS, Kettlewell JR, Hirsch B, Bardwell VJ, et al. (1999). Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* 215: 208-220.
- Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, et al. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* 14: 2587-2595.
- Sigel R, Sigel A and Sigel H (2007). The ubiquitous uoles of cytochrome P450 proteins: metal ions in life sciences. Wiley, New York.
- Scarth J (2006). Modulation of the growth hormone-insulin-like growth factor (GH-IGF) axis by pharmaceutical, nutraceutical and environmental xenobiotics: an emerging role for xenobiotic-metabolizing enzymes and the transcription factors regulating their expression. A review. *Xenobiotica* 36: 119-218.
- Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, et al. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* 461: 267-271.
- Uda M, Ottolenghi C, Crisponi L, Garcia JE, et al. (2004). Foxl2 disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* 13: 1171-1181.
- Uhlenhaut NH, Jakob S, Anlag K, Eisenberger T, et al. (2009). Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* 139: 1130-1142.
- Vidotto M, Grapputo A, Boscari E, Barbisan F, et al. (2013). Transcriptome sequencing and de novo annotation of the critically endangered Adriatic sturgeon. *BMC Genomics* 14: 407.
- Wang DG, Fan JB, Siao CJ, Berno A, et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077-1082.
- Wang D, Lu TY and Liu HB (2009a). Characterization and genetic diversity of the sturgeon *Acipenser schrenckii*. Ig heavy chain. *Immunobiology* 214: 359-366.
- Wang Z, Gerstein M and Snyder M (2009b). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57-63.
- Wei QW, Zou Y, Li P and Li L (2011). Sturgeon aquaculture in China: progress, strategies and prospects assessed on the basis of nation-wide surveys (2007-2009). *J. Appl. Ichthyol.* 27: 162-168.