

# Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (*Acipenser fulvescens*)

Matthew C. Hale · James R. Jackson ·  
J. Andrew DeWoody

Received: 16 December 2009 / Accepted: 30 March 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** Modern pyrosequencing has the potential to uncover many interesting aspects of genome evolution, even in lineages where genomic resources are scarce. In particular, 454 pyrosequencing of nonmodel species has been used to characterize expressed sequence tags, xenobiotics, gene ontologies, and relative levels of gene expression. Herein, we use pyrosequencing to study the evolution of genes expressed in the gonads of a polyploid fish, the lake sturgeon (*Acipenser fulvescens*). Using 454 pyrosequencing of transcribed genes, we produced more than 125 MB of sequence data from 473,577 high-quality sequencing reads. Sequences that passed stringent quality control thresholds were assembled into 12,791 male contigs and 32,629 female contigs. Average depth of coverage was  $4.2 \times$  for the male assembly and  $5.5 \times$  for the female assembly. Analytical rarefaction indicates that our assemblies include most of the genes expressed in lake sturgeon gonads. Over 86,700 sequencing reads were assigned gene ontologies, many to general housekeeping genes like protein, RNA, and ion binding genes. We searched specifically for sex determining genes and documented significant sex differences in the expression of two genes involved in animal sex determination, *DMRT1* and *TRA-1*. *DMRT1* is the master sex determining gene in birds and in medaka

(*Oryzias latipes*) whereas *TRA-1* helps direct sexual differentiation in nematodes. We also searched the lake sturgeon assembly for evidence of xenobiotic organisms that may exist as endosymbionts. Our results suggest that exogenous parasites (trematodes) and pathogens (protozoans) apparently have infected lake sturgeon gonads, and the trematodes have horizontally transferred some genes to the lake sturgeon genome.

**Keywords** Transcriptome · Pyrosequencing 454 · Polyploidy · *Trichomonas* · *Schistosoma*

## Abbreviations

SNP Single nucleotide polymorphism

## Introduction

In many organisms, genes dictate sex. Among mammals and birds, for example, the mode of sex determination is extremely conserved. The primary sex-determining gene in eutherian mammals is *Sry* (Sinclair et al. 1990) whereas *DMRT1* controls bird sex (Smith et al. 2009). Only one master sex determining gene has been discovered in a lower vertebrate. In the medaka fish (*Oryzias latipes*) a homologue of the *DMRT1* gene (called *DMY*) is located on the Y chromosome, and is the genetic switch that determines medaka sex (Matsuda et al. 2002; Nanda et al. 2002). The medaka seems to be the exception, and otherwise sex determination is much more plastic in reptiles, amphibians, and fishes. Temperature and other environmental switches determine sex in some such taxa (see reviews in Devlin and Nagaham 2002 and in DeWoody et al. 2010), but many different systems of genotypic sex determination have evolved

**Electronic supplementary material** The online version of this article (doi:10.1007/s10709-010-9455-y) contains supplementary material, which is available to authorized users.

M. C. Hale (✉) · J. A. DeWoody  
Department of Forestry and Natural Resources, Purdue  
University, West Lafayette, IN 47906, USA  
e-mail: mchale@purdue.edu

J. R. Jackson  
Cornell Biological Field Station, Cornell University, Bridgeport,  
NY 13030, USA

independently during the course of vertebrate evolution (Ferguson-Smith 2007). For instance, many fishes have evolved different means of sex determination and the plasticity of these mechanisms is of considerable interest to evolutionary biologists (Devlin and Nagahama 2002; Avise and Mank 2010; DeWoody et al. 2010).

We are interested in fish sex-determining mechanisms. Thus, a thorough characterization of genes expressed in male and female gonads could prove enlightening, and 454 pyrosequencing of gonad transcriptomes may provide a powerful platform for identifying genes involved in sexual differentiation and development. Several “next-generation” sequencing platforms now are commercially available, including Roche’s 454 pyrosequencer (Hudson 2008). The newest iteration of 454 chemistry, termed Titanium, provides hundreds of thousands of sequences in a single run with a mean read length of >300 bps, roughly an order of magnitude longer than other next-generation sequencing technologies (Margulies et al. 2005; Wicker et al. 2006; Weber et al. 2007). Roche 454 pyrosequencing has been used to quantify differences in gene expression (Torres et al. 2008), locate single nucleotide polymorphisms (SNPs; Barbazuk et al. 2007; Novaes et al. 2008; Hale et al. 2009), scan for signatures of selection (Novaes et al. 2008), and to characterize diets (Deagle et al. 2009). Pyrosequencing error rates are low (<1%) and most errors, such as those in homopolymer (mononucleotide) repeat regions, can be resolved with deep coverage and adequate quality control (Margulies et al. 2005; Moore et al. 2006; Huse et al. 2007). Pyrosequencing has been applied to very few non-model organisms including the Glanville fritillary butterfly (genus *Melitaea*; Vera et al. 2008), the rose gum tree (*Eucalyptus*; Novaes et al. 2008), and a plant pathogen (*Pythium*; Cheung et al. 2008). Collectively, these studies suggest the *de novo* assembly of 454 sequences has the potential to identify and characterize evolutionarily important genes in non-model organisms.

Studies utilizing 454 sequencing technology have also uncovered evidence of xenobiotic organisms associated with the primary host. For example, Miller et al. (2008) applied 454 sequencing technology to genomic DNA isolated from hairs of a woolly mammoth (*Mammuthus primigenius*) preserved in permafrost. Their sequencing results suggested that between 3.8 and 15.8% of reads were from non-target microbial organisms. BLAST searches determined that 0.2% of presumptive mammoth sequences were viral and 3.5–15.5% of reads were bacterial in origin (Miller et al. 2008). Similarly, Vera et al. (2008) also found evidence of xenobiotics in the transcriptome of the Glanville fritillary. There, 452 sequence reads produced a top BLAST hit to microsporidia, which are known to infect insects and can affect insect population dynamics (Kohler and Hoiland 2001). The presence of Microsporidia was not

previously appreciated in Lepidopterans, and this example serves as an illustration of how 454 sequencing can lead to novel xenobiotic discoveries.

We are interested in sturgeon, an ancient lineage of ray-finned fishes that are of conservation concern (Noakes et al. 1999; Peterson et al. 2007). The restoration of lake sturgeon (*Acipenser fulvescens*) populations is complicated by their biology, as they exhibit delayed sexual maturity (between 10 and 30 years of age), infrequent spawning (every few years), and sexual monomorphism (Peterson et al. 2007). DNA sexing assays have proven invaluable in the conservation of wildlife and could be valuable in fish conservation as well, where they could be used to determine sex ratios at stocking as well as sex-specific mortality rates (DeWoody et al. 2010). However, the search for sturgeon sex determining genes has been unsuccessful using traditional molecular approaches such as AFLPs, RAPDs, and subtractive hybridization (Wuertz et al. 2006; Keyvanshokoo et al. 2007; McCormick et al. 2008), probably because the sturgeon genome is polyploid and most loci are inherited tetrasomically or octasomically (Blackledge and Bidwell 1993; Ludwig et al. 2001; DeWoody et al. 2010). Here, we use next-generation 454 sequencing of gonad transcriptomes to annotate lake sturgeon genes involved in sex determination as well as genes from xenobiotic organisms.

## Materials and methods

### RNA isolation and cDNA library construction

Two lake sturgeon, one adolescent female with immature eggs and one sexually mature that was expressing milt, were sampled from Lake Oneida, New York in May 2008 with assistance from research support staff at the Cornell Biological Field Station. These fish were estimated to be 13–14 years old based on the stocking history of the lake, their size relative to growth rates in the lake, and gonad development (Bruch et al. 2001; Jackson et al. 2002; Colombo et al. 2007). From these fish, we collected gonad biopsies with the assistance of veterinarians from Cornell University (P. Bowser, N. Abou-Madi, and their coworkers). Biopsies were immediately frozen in liquid nitrogen and RNA was extracted using TRIZOL reagent (Invitrogen) following the manufacturer’s protocol, with the resulting RNA pellet resuspended in 50  $\mu$ l of RNase free water. Quantity and quality of total RNA was analyzed using a spectrophotometer (Nanodrop) and by gel electrophoresis.

Originally, two cDNA libraries (one from each sex) were constructed for 454 sequencing using GS-FLX chemistry. However, the advent of Titanium chemistry led to the construction of a third cDNA library because of our search for sex-determining genes in female lake sturgeon, the

presumptive heterogametic sex (Blackledge and Bidwell 1993; Van Eenennaam et al. 1999). Thus, library 1 (L1) was from the male whereas libraries 2 (L2) and 3 (L3; Titanium) were from the female. L2 and L3 were constructed from 1 µg of RNA whereas L1 was constructed with 1.5 µg of RNA. All libraries were constructed using a modified version of the CloneTech SMART kit (Zhu et al. 2001) as detailed in Hale et al. (2009).

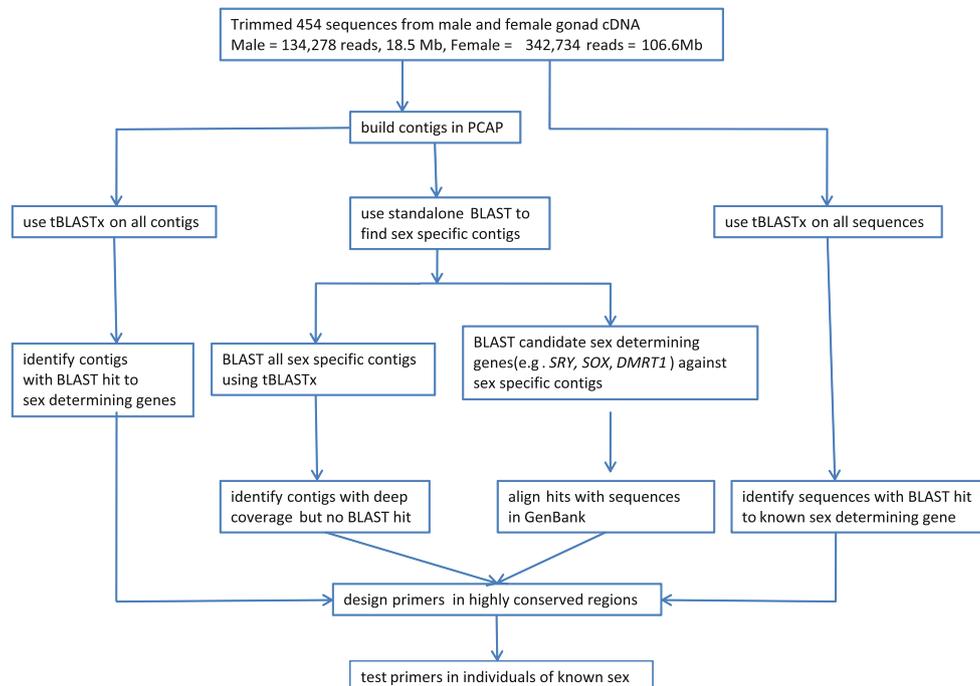
#### 454 sequencing and assembly

Approximately 4 (L1 and L2) to 6 (L3) µg of amplified cDNA were used for 454 library construction and sequencing following established protocols (Margulies et al. 2005). Sequences were screened and those with primer concatamers, weak signal, and/or poly A/T tails were culled. Then, all sequences from L2 and L3 were pooled prior to assembly of the female transcriptome. Assembly was conducted with the GsAssembler and PCAP (Huang et al. 2003) due to the poor performance of the GsAssembler software when lacking a reference sequence (see Cheung et al. 2006; Vera et al. 2008; Hale et al. 2009). Male and female assemblies were annotated using BLASTx with homology settings of  $e \leq 1.00 \times 10^{-03}$  and bit scores  $\geq 40$ . All sequences with significant BLASTx hits were noted, as was the species which produced the top BLAST hit. Blast2Go software was used to search for Gene Ontology (GO) terms and so infer gene function (Ashburner et al. 2000; Shaw et al. 1999). Significant ontology was confirmed using the same settings as the

BLASTx search and ontology was categorized with respect to Molecular Function, Biological Process, and Cellular Component. Gene expression differences between the male and female assemblies were evaluated by calculating the exact binomial probability of obtaining the observed proportion of counts of a particular gene in a particular sex by chance alone. In other words, our null hypothesis was there was no difference in the expression of genes between the sexes. GO terms with a  $P$ -value  $\leq 0.05$  were significantly more abundant in one assembly than in the other; this approach is similar to those used to evaluate cDNA normalization (Hale et al. 2009).

#### Rarefaction

We generated rarefaction curves to compare the rate of gene discovery in both the male and female assemblies (as in Hale et al. 2009). Thus, we compared the efficiency of both assemblies and compared the absolute number of different genes expressed in the male and female gonad transcriptome at the point of sampling. The number of sequences that produced a significant BLAST hit and the number of sequences from the male and female assemblies were determined. Rarefaction curves were constructed using resampling procedures; all identified genes within a library were recorded along with their frequency. This list was then randomized; the number sorted in ascending order, and the first appearance of the new gene was recorded along with its frequency. This procedure was repeated 1,000 times with the program EcoSim700 (Gotelli and Ensminger 1997).



**Fig. 1** Methods employed in the search for candidate sex determining genes

## Candidate gene discovery and PCR validation

In order to identify potential sex-determining markers for lake sturgeon, a variety of approaches were used. Standalone BLAST was used to find expressed sequence tags (ESTs) that were sex specific, i.e. unique to one sex. Amino acid sequences of known sex-determining genes (*SRY*, *DMY*) or genes involved in the vertebrate sex-determination cascade (*Sox2*, *Sox4*, *Sox17*, *Sox21*, *Sox 9*, *RSPO1*, *DMRT1*, *WT1*, *WNT4*, *FOXL2*, *TRA-1*, *FEMI*) were downloaded from GenBank. These amino acid sequences were then compared to the (translated) sex-specific sturgeon ESTs using standalone BLAST (e-value  $\leq 1.00 \times 10^{-3}$ , bit score  $\geq 40$ ). ESTs with significant similarity to candidate sex-determining genes were selected and BLASTed to Genbank using BLASTx in case the ESTs had greater sequence similarity to a different gene. We designed PCR primers to amplify these genes as candidate sexing markers.

A significant proportion of sex-specific contigs did not produce a significant BLAST hit to any known candidate sex-determining gene. However, because sturgeon represent an ancient lineage poorly represented in the sequence databases and because of the plasticity of sex-determination in fishes, we considered the possibility that lake sturgeon have a novel and hitherto unknown sex-determining gene. Thus, we also designed primers from ESTs that were unique to one sex but produced no GenBank hits. We focused our efforts on ESTs with the best coverage, both in terms of length and depth. Furthermore, we BLASTed all ESTs using tBLASTx against the GenBank database for annotation purposes, but also to search for any hits to genes involved in the sex determining pathway that were missed by the methods described above. Figure 1 depicts a flowchart of our methods for identifying potential sex-determining genes.

All candidate sexing primers were validated using four male and four female lake sturgeon samples graciously provided by R. Elliot (USFWS). PCRs consisted of 1 × NEB buffer, 2.0 mM MgCl<sub>2</sub>, 1 μM of each primer, 250 μM dNTPs, 0.5 units of NEB *Taq*, 30–50 ng of template DNA, and water to 10 μl. Primers were tested using the following PCR conditions: an initial denature at 94°C for 3 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 45 s. PCRs were terminated with a final 5 min extension at 72°C. PCR products were analyzed in a 1.2% agarose gel stained with ethidium bromide and viewed under UV light.

## Identification of xenobiotics

The cDNA we sequenced was derived from tissue biopsies of lake sturgeon gonads. Thus, we expected that the vast majority of genes sequenced should be lake sturgeon in origin. However, it is possible that genes from

microorganisms that are parasitic, symbiotic, or commensal might also be sequenced. To identify such sequences, we used BLASTx to query the entire GenBank database with all of our 454 sequences. The mean e-value for metazoan hits was compared to those of non-metazoan hits as one means of determining sequence identity. Furthermore, apparent xenobiotic sequences were assembled into contigs and then were individually evaluated. Ultimately, we considered organisms with multiple BLAST hits to different genes as the most likely xenobiotics. In some cases, we used gene-specific PCR primers to verify the presence of xenobiotics (see Results).

## Results

### Sequence assembly

A total of 473,577 sequence reads remained after quality control and primer trimming; over 72,000 reads were culled (Table 1). Mean read length was over twice as long for the Titanium chemistry (343 bp) versus the GS-FLX chemistry (163 bp). PCAP consistently produced longer and deeper contigs than GsAssembler, so we present only the PCAP assembly (Table 2). The female assembly (L2 and L3 combined) produced 32,629 contigs with a mean contig length of 597 bp (mode = 506, median = 529) and

**Table 1** Characteristics of three cDNA libraries (L1–L3) and the sequences derived from them

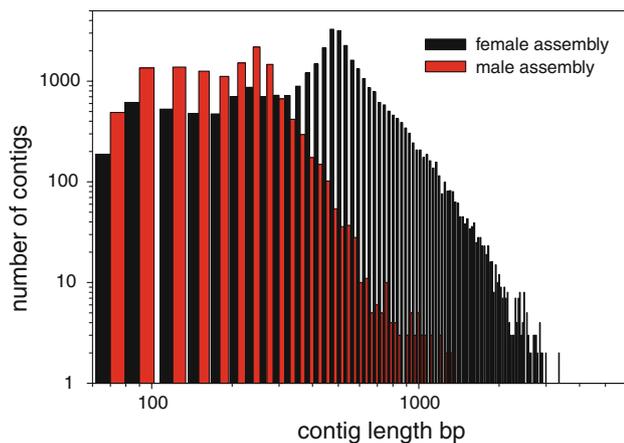
	L1	L2	L3	Total
Tissue source	Male	Female	Female	n/a
Chemistry	GS-FLX	GS-FLX	TITANIUM	n/a
N. reads	192,260	80,468	273,371	546,099
N. reads after culling	134,278	69,366	269,913	473,577
N. nucleotides	18,521,429	12,940,424	93,700,732	125,162,585
Mean read length	138	187	342	222

BLAST criteria for a significant hit were an e-value  $\leq 1 \times 10^{-3}$  and a bit score  $\geq 40$

**Table 2** Results of PCAP assembly using the default setting

	Male assembly	Female assembly
Number of contigs	12,791	32,629
Mean depth of contig (>200 bp)	4.2	5.5
Mean length of contig (>200 bp)	277	597
Number of reads in contigs	56,057	206,996
Number of singletons	99,713	154,724

cDNA libraries L2 and L3 are both included in the female assembly



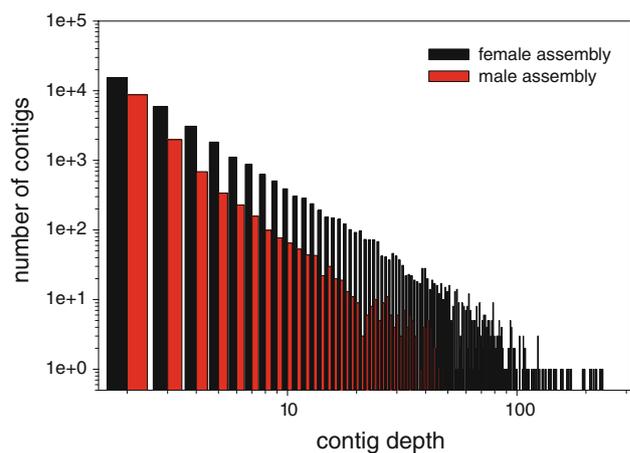
**Fig. 2** Length of contigs of the male and female assemblies. Note the logarithmic axes. Reprinted with permission from DeWoody et al. (2010)

a mean depth of 5.5 (mode = 2, median = 3) reads. In comparison, the male assembly (all from GS-FLX reads of L1) produced 12,791 contigs with a mean contig length of 277 bp (mode = 244, median = 252) and a mean depth of 4.2 sequences (mode = 2, median = 2). Over half (57%) the reads from female gonads were assembled into contigs, whereas about a quarter (22%) of the reads from the male were assembled into contigs. Figure 2 shows contig length in each assembly whereas Fig. 3 shows contig depth. Regression analysis indicated there was a significant positive relationship between contig length (bp) and contig depth (number of sequences within a contig;  $r^2 = 0.25$ , slope =  $3.372 (\pm 0.09 \text{ SD})$ ,  $P = < 0.001$ ). Rarefaction analyses (Fig. 4) indicate that our sampling effort was sufficient to identify most genes expressed in lake sturgeon gonads.

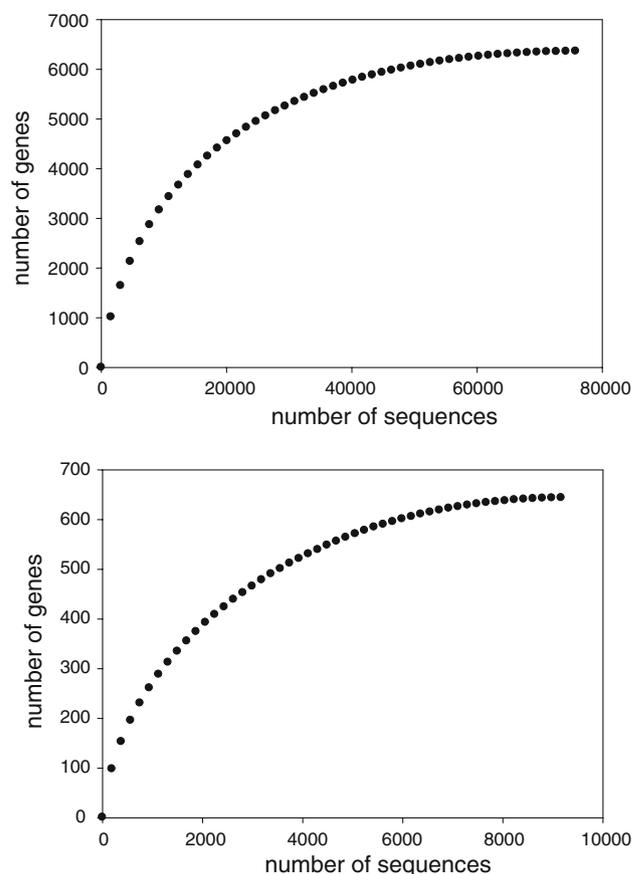
#### Gene annotation

GO assignments placed ESTs from the female assembly into 1081 different Molecular Function categories whereas male ESTs were placed into 267 such categories (Supplementary Table 1). The three most common categories in the female assignment were protein binding (8.02%), ATP binding (6.32%), and DNA binding (2.85%). In the male assembly, the three most common categories were cytochrome-c oxidase activity (8.37%), DNA binding (6.57%), and transposase activity (6.3%). Binomial probability tests revealed a statistically significant difference in gene expression across 28 Molecular Function categories. Expression of ten categories was greater in the female library than the male library, whereas expression of the other 18 categories was greater in the male library.

GO analysis identified 1362 Biological Process categories from the female assembly and 157 such categories



**Fig. 3** Depth of contigs of the male and female assemblies. Reprinted with permission from DeWoody et al. (2010)



**Fig. 4** Rarefaction curves illustrating gene discovery (number of genes) as a function of effort (number of sequencing reads). **a** rarefaction analysis comparing empirical gene discovery in the female assembly, **b** rarefaction analysis comparing empirical gene discovery in the male assembly

from the male assembly (Supplementary Table 2). The three most common categories in the female assembly were translation (4.39%), transport (3.39%), and DNA-dependent regulation of transcription (2.49%). The three most

common categories in the male assembly were translation (19.22%), electron transport (9.31%), and transport (7.68%). Binomial probability tests revealed 35 Biological Process categories that showed a statistically significant difference in expression between the sexes. Expression of 14 categories was higher in the female assembly than the male assembly; expression of the remaining 21 categories was higher in the male assembly.

GO analysis identified 21 different categories of Cellular Component among the lake sturgeon sequences (Supplementary Table 3). The top three categories in the female assembly were nucleus (14.53%), cytoplasm (13.16%), and structural constituent of ribosomes (8.78%). The top three categories in the male assembly were structural constituent of ribosomes (28.07%), ribosome (12.82), and integral to membrane (11.75%). All 21 Cellular Component categories showed a significant difference in expression between the sexes, with eleven categories showing higher expression in females than in males and 10 higher expression in males than females.

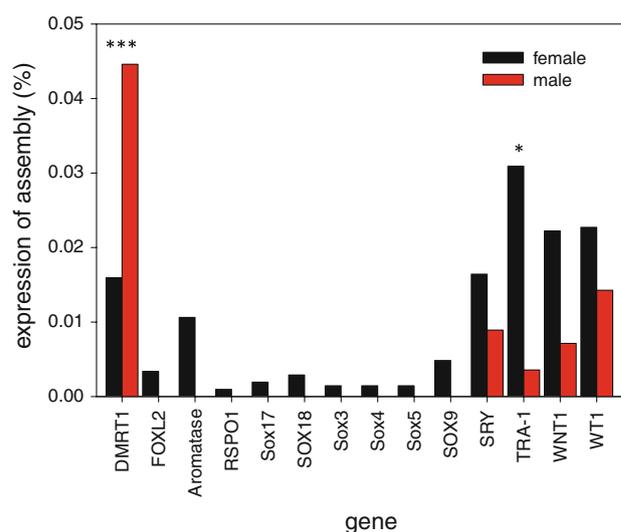
#### Search for sex determining genes in lake sturgeon

We evaluated a grand total of 73 candidate sex determining markers in lake sturgeon. Sequences from 12 candidate sex determining genes (*Sox2*, *Sox4*, *Sox17*, *Sox21*, *Sox9*, *DMRT1*, *RSPO1*, *WT1*, *WNT4*, *FOX12*, *TRA-1*, *FEM1*) produced significant BLAST hits from which we designed 44 PCR primer pairs. All but one of these (43) amplified in both males and females ( $n = 4$  of each sex); *FEM1* failed to amplify. We also identified 31 sex-specific contigs that were  $\geq 10$  sequences deep; 9 of these 31 were specific to males and 22 contigs were specific to females (i.e., not present in the assembly of the opposite sex or in Genbank). Nine PCR primer pairs were designed from the nine male-specific contigs and 24 primer pairs were developed from the 22 female-specific contigs. Again, many of these primer pairs (30 of 33 primer pairs) amplified lake sturgeon DNA, but none produced a sex-specific product.

The expression of two candidate sex-determining genes differed between the adolescent female and the young adult male (Fig. 5). Statistically significant differences in gene expression were noted between the sexes in *DMRT1* with males expressing significantly more *DMRT1* than females ( $P = < 0.001$ ), as seen in chickens (Smith et al. 2009). A difference in expression was also seen in *TRA-1*, with an expression bias towards females ( $P = 0.05$ ).

#### Xenobiotics

Analyses of the top BLASTx hits suggests 331 sequences out of 342,737 female reads (0.1%), and 183 sequences out of 134,278 male reads (0.14%), produced a positive



**Fig. 5** Differences in gene expression of candidate sex determining genes. The relative expression of these genes in the female assembly is shown in *black*, the male assembly in *red*. Significant differences in expression between the sexes was determined by a binomial distribution test, \* =  $P < 0.05$ , \*\*\*  $P < 0.001$

BLASTx hit that matched a non-metazoan taxa. These sequences were distributed among 17 female contigs and 22 male contigs, respectively. The mean BLASTx e-value for non-metazoan hits was  $5.95 \times 10^{-5}$  ( $\pm 1.29 \times 10^{-4}$ ) compared to the average BLASTx e-value for metazoan hits  $8.171 \times 10^{-6}$  ( $\pm 6.44 \times 10^{-5}$ ). We found 39 contigs with a top BLAST hit to a candidate xenobiotic of some kind (Table 3).

Fifteen different contigs (totaling 145 individual reads; mean BLAST hit =  $8.9 \times 10^{-6}$ ) from the male assembly each produced a top BLAST hit to a gene from a trematode of the genus *Schistosoma*. These contigs produced hits to 9 different unannotated proteins (Table 3). To our knowledge, *Schistosoma* trematodes have not been previously reported in sturgeon, but there is evidence for horizontal transfer of genes between *Schistosoma* trematodes and salmonid fishes (Melamed et al. 2004; Matveev and Okada 2009). To further evaluate the possibility of horizontal gene transfer, we designed PCR primers for *Schistosoma* (5'-TAGGGACAGTGGGAATCTCG-3' and 5'-GTGTTGACGCGATGTGATTT-3') and used them to try and amplify trematode DNA from the lake sturgeon genome. As a PCR template, we used genomic DNA isolated from fin clips, and this produced amplicons of the expected size based on the *Schistosoma* sequences we recovered from cDNA (Fig. 6).

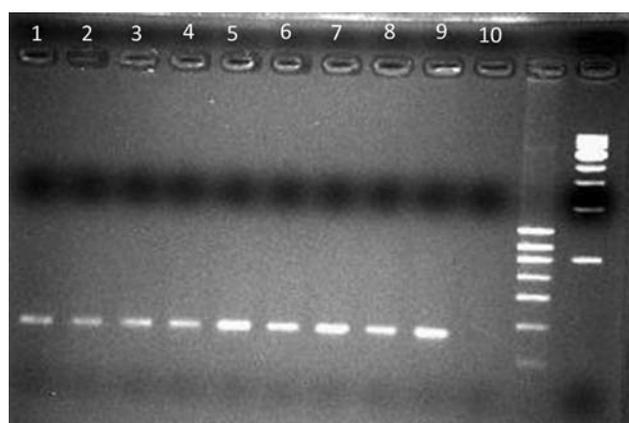
Four different contigs (totaling 21 individual reads) from the female assembly all produced a top BLAST hit to the protozoan *Trichomonas vaginalis*. Furthermore, these four contigs produced BLAST hits to four different genes from *Trichomonas vaginalis* (Tlp1-like cysteine peptidase,

**Table 3** Evidence of xenobiotics in the lake sturgeon transcriptome

Sex	Contig ID	Depth	e-value	Blast hit	Species	Effect on host
F	679.1	31	0.00E + 00	Hypothetical protein	<i>Leishmania infantum</i>	Causes leishmiasis in humans
F	5.1	173	1.10E-52	Senescence-associated protein	<i>Tetrahymena thermophila SB210</i>	Ciliate can be pathogenic
F	276.1	52	6.43E-28	mflj00348 protein	<i>Brugia malayi</i>	Nematode causes elephantitis
F	1519.1	18	6.86E-28	mflj00348 protein	<i>Brugia malayi</i>	Nematode causes elephantitis
F	13601.1	3	1.44E-18	chk1 checkpoint homolog	<i>Brugia malayi</i>	Parasitic nematode
F	22475.1	2	5.31E-18	Endonuclease reverse transcriptase	<i>Lymphocystis disease virus</i>	Common viral disease of fish
F	6541.1	6	2.38E-15	sjchgc01393 protein	<i>Entamoeba histolytica HM-1:IMSS</i>	Pathogenic in humans
F	18576.1	2	3.05E-15	Cytochrome c-type protein	<i>Leifsonia xyli</i> subsp. <i>xyli</i>	Causes root disease in sugar cane
F	5074.1	7	1.71E-08	lp1-like cysteine peptidase	<i>Trichomonas vaginalis G3</i>	Vaginal infection in humans
F	11201.1	4	2.26E-07	sr-like protein ra9	<i>Trichomonas vaginalis G3</i>	Vaginal infection in humans
F	7406.1	5	1.18E-06	c4sr protein	<i>Chaetomium globosum</i>	Pathogenic fungus in humans
F	3532.1	9	3.24E-05	ftsk spoiie family protein	<i>Bacillus cereus NVH0597-99</i>	Food associated disease in humans
F	8192.1	5	3.46E-05	sjchgc01974 protein	<i>Candida albicans SC5314</i>	Causes thrush
F	21699.1	2	8.05E-05	Membrane family cbs domain	<i>Burkholderia mallei PRL-20</i>	Pathogenic in humans and vertebrates
F	6430.1	6	1.15E-04	DNA ligase	<i>Trichomonas vaginalis G3</i>	Vaginal infection in humans
F	8467.1	4	4.12E-04	Chagas antigen	<i>Trichomonas vaginalis G3</i>	Vaginal infection in humans
F	29439.1	2	5.50E-04	Minicircle sequence binding protein	<i>Trypanosoma cruzi</i>	Causes chagas disease In humans
M	30.1	57	9.11E-47	SJCHGC06398 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	1158.1	5	7.12E-28	SJCHGC03019 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	74.1	35	1.35E-26	Unknown	<i>Schistosoma japonicum</i>	Trematode parasite
M	929.1	6	1.08E-20	Unknown	<i>Schistosoma japonicum</i>	Trematode parasite
M	1679.1	4	3.26E-20	SJCHGC03026 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	1252.1	5	3.43E-13	SJCHGC01615 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	2161.1	3	3.26E-12	Putative ribosomal protein L7Ae-like	<i>Staphylococcus haemolyticus</i>	Common on skin of vertebrates
M	4631.1	2	2.09E-11	SJCHGC09227 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	1111.1	5	1.36E-10	SJCHGC04882 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	2585.1	3	6.31E-08	TPA: reverse transcriptase	<i>Schistosoma mansoni</i>	Trematode parasite
M	2973.1	3	1.85E-07	SJCHGC03009 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	527.1	9	2.43E-07	SJCHGC01615 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	9054.1	2	4.55E-06	SJCHGC06004 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	494.1	9	6.05E-06	Hypothetical protein Bm1_22195	<i>Brugia malayi</i>	Nematode pathogenic in humans
M	1723.1	4	9.13E-06	SJCHGC03026 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	2714.1	3	1.02E-05	Histidine-rich glycoprotein precursor	<i>Brugia malayi</i>	Nematode pathogenic in humans
M	2668.1	3	2.26E-05	SJCHGC03026 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	6933.1	2	8.46E-05	SJCHGC09227 protein	<i>Schistosoma japonicum</i>	Trematode parasite
M	1210.1	5	1.16E-04	Hypothetical protein Btr_1952	<i>Bartonella tribocorum</i>	Pathogenic in humans
M	355.1	12	1.91E-04	Hypothetical protein Bm1_22195	<i>Brugia malayi</i>	Nematode pathogenic in humans
M	5444.1	2	2.20E-04	Hypothetical protein	<i>Theileria parva</i> strain Muguga	Pathogenic in cattle
M	1670.1	4	4.31E-04	Hypothetical protein	<i>Theileria parva</i> strain Muguga	Pathogenic in cattle

DNA ligase, chagas antigen, and sr-like protein ra9). This concordance across genes strongly suggests that these protozoan hits are not spurious. To evaluate the possibility

of horizontal gene transfer, we designed three sets of PCR primers for *Trichomonas* and used as template cDNA isolated from the gonads as well as genomic DNA isolated



**Fig. 6** The amplification of a trematode gene from lake sturgeon genomic DNA. From the left, lanes 1–4 are from male lake sturgeon genomic DNA isolated from fin clips whereas lanes 5–8 are from female lake sturgeon genomic DNA, again isolated from fin clips. Lane 9 is a positive control of *Schistosoma mansoni* genomic DNA (courtesy of E. Thiele and D. Minchella). Lane 10 is a no-template negative control reaction

from fin clips. The amplifications from lake sturgeon genomic DNA (fin clips) failed, whereas those from cDNA succeeded (data not shown), indicating that the infection was likely contemporary. Trichomonad infections are known to occur in many terrestrial vertebrates including dogs (*T. canistomae*), domestic chickens (*T. gallinarum*), rodents (*T. microti*), and Rhesus macaques (*T. macaovaginase*). If confirmed, this discovery of protozoans in lake sturgeon would be the first trichomonad infection known in lower vertebrates, or in an aquatic animal.

## Discussion

### Transcriptome assembly

We successfully generated over 473,000 sequencing reads that corresponded to 125,000,000 nucleotides of sequence data (125 MB) from genes expressed in lake sturgeon gonads. These sequences were assembled into 45,420 contigs (32,629 in females and 12,791 in males). The Titanium chemistry used for L3 produced more and longer reads than did the GS-FLX chemistry with L1 and L2. There were only 2-fold more reads with the Titanium run, but those longer reads produced 5-fold more data. The increased number of reads and read lengths are both important for assembly in non-model organisms that lack a reference genome sequence, and are important for SNP identification. These data include thousands of lake sturgeon SNPs (Hale et al. 2009).

Our lake sturgeon assemblies compare favorably to those of other non-model species. Mean contig length for

the male assembly was 277 bp, slightly more than the fritillary assembly (197 bp; Vera et al. 2008) but less than the eucalypt assembly (353 bp; Novaes et al. 2008). Mean contig length from the female transcriptome is much higher (597 bp), which suggests the Titanium chemistry substantially improves contig assembly. Sequences from longer contigs are more likely to produce a positive BLAST hit, which in turn produces more annotated sequences. There was little difference in the mean contig depth between the lake sturgeon assemblies (4.2 vs. 5.5 in males and females, respectively). This compares favorably to the fritillary assembly (2.9 reads; Vera et al. 2008) but is less than the eucalypt assembly (9.9 reads; Novaes et al. 2008). The difference in contig depth between lake sturgeon and eucalypt may be due to ploidy. The ploidy of lake sturgeon is uncertain; there is evidence for disomic, tetrasomic, and octasomic inheritance of lake sturgeon microsatellite loci but no consensus on the overall level of ploidy (Ludwig et al. 2001). Polyploidy no doubt affects transcriptome (and genome) assembly because paralogs may on occasion be inadvertently incorporated into the same contig. This complicates direct comparisons between diploid and polyploid assemblies.

The annotation of the lake sturgeon gonad transcriptome is difficult due to the lack of sequence data in closely related species. There are two lineages of ray-finned fishes, Chondrostei (primarily cartilaginous fishes with some ossification, including sturgeon) and Neopterygii (true bony fishes). There are fewer than 3000 Chondrostei sequences available in GenBank whereas there are >1,500,000 sequences in Neopterygii fishes. This means there are very few sequences of sturgeon relatives in GenBank. Thus, we used a Gene Ontology approach to characterize male and female transcriptomes, and to evaluate the expression of different GO categories. Our approach did not utilize normalization of mRNA because it can obfuscate gene expression (Hale et al. 2009). For example, Supplementary Tables 1–3 indicate there are sex-specific biases in the expression of many genes. The overall trend in the female assembly is the expression of genes representing many different GO categories, none of which alone constitutes a large fraction of the total reads. By contrast, genes expressed in the male assembly represent fewer GO categories and those often contain a large fraction of the total reads. For example, more than 20% of the total reads from the male assembly are related to translation and ribosome production (i.e., protein synthesis, see Supplementary Tables 1 and 3). These same genes are also expressed in ovary tissue, but to a much lesser proportion.

Overall, 35 different Biological Process GO subcategories are differentially expressed between the sexes. Of these, males express significantly more genes involved in translation and in transport than females. Analyses of the Molecular Function GO subcategory indicates that male

expression was higher in genes associating with DNA binding, cytochrome-c oxidase activity, and transposase activity. The female assembly indicates there is increased expression of genes associated with ATP-binding and protein binding. The Cellular Component GO subcategory indicates greater expression of membrane and ribosome genes in testes, whereas genes found in the cytoplasm and nucleus are more highly expressed in female gonads. Our analyses of the testes transcriptome repeatedly points towards the increased expression of genes in regions of the cell (cytoplasm) associated with active translation. This implies to us a greater rate of protein synthesis in testes compared to ova, perhaps because the male was sexually mature (readily expressing milt) whereas the adolescent ovaries were not fully developed. This is confirmed by histology, where slides from the male depict the active production of seemingly-functional gametes but slides from female tissue do not. The immature nature of the female gonads used here might mean that more genes are expressed but most at low levels, and thus the components of translation are not particularly in demand.

#### Sex-determining genes

Conventional genetic (mapping) and molecular approaches (subtractive hybridization, cloning, candidate genes, etc.) have failed to identify master sex-determining genes in a variety of fishes, including sturgeon (Wuertz et al. 2006; Keyvanshokoo et al. 2007; McCormick et al. 2008). Our strategy was to diminish the overall complexity of the lake sturgeon genome by sequencing cDNA instead of genomic DNA. We then used two primary approaches to identify sex-specific genes. First, we evaluated candidate genes whereby gene sequences associated with sex-determining genes in other taxa were used to query the lake sturgeon transcriptome. This strategy is reasonable because downstream genes in the sex-determining cascade of one species may in fact become a master sex-determining gene in another species. However, none of the candidate genes that we evaluated in lake sturgeon were sex-specific. This is not too surprising, as the master sex-determining gene in one fish species may not be the cascade trigger even in a closely related congeneric species (as demonstrated by *DMY* in the genus *Oryzias* which is not the master sex determining gene in closely related species to *O. latipes*, Kondo et al. 2004). We did find significant differences in the sex-specific expression of many important genes (Supplementary Tables 1–3), including two genes involved in the sex-determining pathway of other species (i.e., *DMRT1* and *TRA-1*; Fig. 5).

Our second approach was not based on candidate genes *per se*, but on evaluating contigs that were present in one sex but absent in the other. Such contigs could

represent a novel sex-determining gene, especially if the contig was relatively deep (>10 sequences). We isolated 3,332 contigs unique to the male assembly and 4,008 unique to the female assembly. BLAST searches of these contigs revealed that some were similar to known genes in other organisms, but 1,004 sex-specific contigs had no counterpart in GenBank. In theory, these could represent novel sex-determining genes heretofore uncharacterized in cartilaginous fishes, but PCR tests repeatedly documented their presence in both sexes and thus they are unlikely to determine sex by dominance (as with *Sry* in mammals).

Some sturgeon have heteromorphic sex chromosomes (Blackledge and Bidwell 1993; Van Eenennaam et al. 1999) and anecdotal evidence suggests hatchery sex ratios are near unity, which both suggest genetic sex determination. However, sturgeon geneticists have failed to find sex-determining genes despite using a plethora of molecular approaches (Hett and Ludwig 2005; Wuertz et al. 2006; Keyvanshokoo et al. 2007; McCormick et al. 2008). Similarly, the research described herein provides no evidence for a master gene whose presence or absence determines lake sturgeon sex. However, we have identified two genes that are differentially expressed between the sexes. *DMRT1* expression is higher in our male sample than in our female (as in eels and grouper; Huang et al. 2005; Alam et al. 2008), whereas *TRA-1* is expressed more highly in our female sample than in our male sample. Thus, the dosage of either (or both) of these genes may dictate sex in developing lake sturgeon. *DMRT1* is now thought to be the master sex-determining gene in chickens, where it acts in a dosage dependent manner (Smith et al. 2009). There, ZW females receive a single dose of *DMRT1* whereas ZZ males receive two doses. Our lake sturgeon data are consistent with this idea of sex by dosage, but more research is needed to determine if one or both of these genes is the primary determinant of sturgeon sex.

Of course, an alternative model might suggest that a hypothetical master sex-determining gene is not expressed in the gonads of adolescent females or young adult males, but is expressed much earlier in development. In theory, one could repeat our work in gonads at different points in development (i.e., embryos, fingerlings, etc.) but in practice this would be quite difficult because of the problems associated with collecting the relevant tissues in juveniles, delayed sexual maturity (>10 years), and the costs associated with transcriptome analyses.

#### Xenobiotics

The identification of microsymbionts via 454 transcriptome sequencing was first described by Vera et al. (2008). Their

paper on the glanville fritillary presented evidence of a microsporidia intracellular parasite that was known to adversely affect insect population dynamics, but had not been reported in any lepidopteran species. Our searches for xenobiotics similarly suggest the presence of previously unknown exogenous organisms in lake sturgeon gonads.

Our dataset consists of 145 sequencing reads that are most similar to genes from trematodes of the genus *Schistosoma*; 14 out of the 15 contigs that produced a top BLAST hit to *S. japonicum*. These 145 reads do not all BLAST to the same *Schistosoma* gene, which suggests that we isolated multiple genes from a trematode parasite of lake sturgeon. *Schistosoma* infections have not been reported in sturgeon but lake sturgeon are bottom feeders that regularly ingest snails, which are often the first intermediate host of *Schistosoma* parasites (Peterson et al. 2007; Matveev and Okada 2009).

Most *Schistosoma* infections have been reported in mammals (He et al. 2001), but there is indirect evidence that these trematodes have infected salmonid fishes (Melamed et al. 2004; Matveev and Okada 2009). Several schistosome EST libraries contain sequences with high homology (~95%) to LINE elements derived from the rainbow trout genome (Melamed et al. 2004). It seems likely that these mobile elements were inserted into the schistosome genome as a result of parasitic infection of rainbow trout (*Oncorhynchus mykiss*) followed by horizontal gene transfer. We have no direct evidence for contemporary trematode infections of lake sturgeon, but schistosomes infect most vertebrate classes and *S. japonicum* can utilize many different hosts (He et al. 2001). Our data do show that trematode genes have been horizontally transferred to the lake sturgeon genome, as PCR tests (Fig. 6) amplify trematode DNA from lake sturgeon genomic DNA isolated from fin snips.

Four contigs from our lake sturgeon libraries produced a top BLAST hit to the protozoan *Trichomonas vaginalis*. We think the evidence for a protozoan infection in lake sturgeon gonads is strong, and in some respects mirrors the evidence for a schistosome infection. In both cases, the evidence is based on multiple sequences that form different (independent) contigs. In the case of *Trichomonas*, two contigs encode proteins of very general function (DNA ligase and cysteine peptidase). If these sequences originated in lake sturgeon or in a human contaminant rather than a microorganism, then we would expect a stronger vertebrate hit (e.g., zebrafish or medaka) for the same function. Furthermore, the *Trichomonas* BLAST hits were only documented in the female gonads, which mirrors findings from other vertebrates with trichomonad infections. Unlike the trematode case, however, our PCR tests suggest the *Trichomonas* infection is contemporary and not due to horizontal gene transfer, as we generated

*Trichomonas* amplicons only from cDNA and not from genomic DNA.

In humans, *T. vaginalis* is the causative agent of a sexually-transmitted vaginal infection that, according to World Health Organization estimates, infects ~180 million people each year. This infection is normally not severe for the host, but has negative consequences for the health of the infant born of an infected mother (e.g., low birth weight, increased mortality, and preterm delivery; Schwebke and Burgess 2004). Other species of *Trichomonas* infect different terrestrial vertebrates, including rodents, carnivores, galliform birds, and ruminants. In those species, infection sites appear to be either the urogenital tract or the oral cavity. If our evidence for a contemporary trichomonad infection in lake sturgeon is confirmed, it would be the first search reported in anything but a mammal or bird. It would also be the first time a trichomonad infection has been found in an aquatic organism. Given the ill-effects of *Trichomonas* infection in the human population, one wonders if they may similarly affect lake sturgeon. If so, this could prove important to their conservation.

## Conclusions

We have demonstrated that it is possible to use next-generation pyrosequencing to characterize the transcriptome of a non-model polyploid species; we have characterized most of the genes expressed in the gonads of adolescent and young adult lake sturgeon. Here, we briefly summarize some of our major conclusions. First, our assemblies reveal sex-specific differences in the expression of many genes, and GO analyses suggest there is considerably more protein synthesis in ripe male gonads than in immature female gonads. Second, we identified two genes (*DMRT1* and *TRA-1*) that are differentially expressed between sexes and may play key roles in sex determination. Third, we provide evidence that xenobiotic protozoa and trematodes have infected the gonads of female lake sturgeon, and that genes from the trematode were horizontally transferred to the lake sturgeon genome. The fitness implications of these infections (if any) remain to be determined. Overall, these findings illustrate how next-generation sequencing provides a powerful platform for the study of ecologically and evolutionarily important genes.

**Acknowledgments** We thank P. San Miguel and P. Parker for their help with sequencing of the cDNA. We are grateful for helpful comments from the DeWoody lab group and from A. Chin-Baarstad, G. Dharmarajan, S. Edwards, B. Hecht, and K. Nichols. In addition, we are indebted to M. Ducore for his help with PCR validation. This work was funded by grants to JAD from the Great Lakes Fishery Trust and the Indiana Department of Natural Resources through a State Wildlife Grant (T07R05).

## References

- Alam MA, Kobayashi Y, Horiguchi R et al (2008) Molecular cloning and quantitative expression of sexually dimorphic markers *Dmrt1* and *Foxl2* during female-to-male sex change in *Epinephelus merra*. *Gen Comp Endocrinol* 157:75–85
- Ashburner MC, Ball JA, Blake A et al (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29
- Avise JC, Mank JE (2010) Evolutionary perspectives on hermaphroditism in fishes. *Sex Dev*, in press
- Barbazuk WB, Emrich SJ, Chen HD et al (2007) SNP discovery via 454 transcriptome sequencing. *Plant Journal* 51:910–918
- Blackledge KH, Bidwell CA (1993) Three ploidy levels indicated by genome quantification in *Acipenseriformes* of North America. *J Hered* 84:427–430
- Bruch RM, Dick TA, Choudhury A (2001) A field guide for the identification of stages of gonad development in lake sturgeon (*Acipenser fulvescens* Rafinesque) with notes on lake sturgeon reproductive biology and management implications. *Sturgeon for Tomorrow*, Fond du Lac, WI
- Cheung F, Haas BJ, Goldberg SMD et al (2006) Sequencing *Medicago truncatula* expressed sequenced tags using 454 life sciences technology. *BMC Genomics* 7:272
- Cheung F, Win J, Lang JM et al (2008) Analysis of the *Pythium ultimum* transcriptome using Sanger and Pyrosequencing approaches. *BMC Genomics* 9:542
- Colombo RE, Garvey JE, Wills PS (2007) A guide to the embryonic development of the shovelnose sturgeon (*Scaphirhynchus platyrhynchus*), reared at a constant temperature. *J Appl Ichthyol* 23:402–410
- Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by Pyrosequencing prey DNA in faeces. *Mol Ecol* 18:2022–2038
- Devlin RH, Nagahama Y (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208:191–364
- DeWoody JA, Hale MC, Avise JC (2010) Vertebrate sex determining genes and their potential utility in conservation, with particular emphasis on fishes. In: DeWoody JA, Bickham JW, Michler C, Nichols K, Rhodes OE, Woeste K (eds). *Molecular approaches in natural resource conservation*. Cambridge University Press (in press)
- Ferguson-Smith M (2007) The evolution of sex chromosomes and sex determination in vertebrates and the key role of *DMRT1*. *Sex Dev* 1:2–11
- Gotelli NJ, Ensminger GL (1997) EcoSim: null models software for ecologists. *Bull Ecol Soc Am* 78:96
- Hale MC, McCormick CR, Jackson CR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics* 10:203. doi:10.1186/1471-2164-10-203
- He YX, Salafsky B, Ramaswamy K (2001) Host-parasite relationships of *Schistosoma japonicum* in mammalian hosts. *Trends Parasitol* 17:320–324
- Hett AK, Ludwig A (2005) SRY-related (*Sox*) genes in the genome of European Atlantic sturgeon (*Acipenser sturio*). *Genome* 48:181–186
- Huang XQ, Wang JM, Aluru S et al (2003) PCAP: A whole-genome assembly program. *Genome Res* 13:2164–2170
- Huang X, Hong CS, O'Donnell M et al (2005) The doublesex-related gene *XDmrt4*, is required for neurogenesis in the olfactory system. *Proc Natl Acad Sci USA* 102:11349–11354
- Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour* 8:3–17
- Huse SM, Huber JA, Morrison HG et al (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8:R143
- Jackson JR, VanDeValk AJ, Brooking TE et al (2002) Growth and feeding dynamics of lake sturgeon, *Acipenser fulvescens*, in Oneida Lake, New York: results from the first five years of a restoration program. *J Appl Ichthyol* 18:439–443
- Keyvanshokoh S, Pourkazemi M, Kalbassi MR (2007) The RAPD technique failed to identify sex-specific sequences in beluga (*Huso huso*). *J Appl Ichthyol* 23:1–2
- Kohler SL, Hoiland WK (2001) Population regulation in an aquatic insect: the role of disease. *Ecology* 82:2294–2305
- Kondo M, Nanda I, Hornung U et al (2004) Evolutionary origin of the medaka Y chromosome. *Curr Biol* 14:1664–1669
- Ludwig A, Belfiore NM, Pitra C et al (2001) Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics* 158:1203–1215
- Margulies MM, Egholm W, Altman E et al (2005) Genome sequencing in microfabricated high-density pico litre reactors. *Nature* 437:376–380
- Matsuda M, Nagahama Y, Schinomiya A et al (2002) DMT is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–563
- Matveev V, Okada N (2009) Retroposons of salmonid fishes (Actinopterygii: Salmonoidei) and their evolution. *Gene* 434:16–28
- McCormick CR, Bos DH, DeWoody JA (2008) Multiple molecular approaches yield no evidence for sex-determining genes in lake sturgeon (*Acipenser fulvescens*). *J Appl Ichthyol* 24:643–645
- Melamed P, Chong KL, Johansen MV (2004) Evidence for lateral gene transfer from salmonids to two Schistosome species. *Nat Genet* 36:786–787
- Miller W, Drautz DI, Ratan A et al (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:380–387
- Moore MJ, Dhingra A, Soltis PS et al (2006) Rapid and accurate Pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* 6:17
- Nanda I, Kondo M, Hornung U et al (2002) A duplicated copy of *DMRT1* in the sex determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc Natl Acad Sci USA* 99:11778–11783
- Noakes DLG, Beamish FWH, Rossiter A (1999) Conservation implications of behaviour and growth of the lake sturgeon, *Acipenser fulvescens*, in northern Ontario. *Environ Biol Fishes* 55:135–144
- Novaes ED, Drost R, Farmerie WG et al (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9:312
- Peterson DL, Vecsei P, Jennings CA (2007) Ecology and biology of the lake sturgeon: a synthesis of current knowledge of a threatened North American *Acipenseridae*. *Rev Fish Biol Fish* 17:59–76
- Schwebke JR, Burgess D (2004) Trichomoniasis. *Clin Microbiol Rev* 17:794–803
- Shaw DR, Ashburner M, Blake JA et al (1999) Gene Ontology: a controlled vocabulary to describe the function, biological process and cellular location of gene products in genome databases. *Am J Hum Genet* 65:1419
- Sinclair AH, Berta P, Palmer MS et al (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244
- Smith CA, Roeszler KN, Ohnesirg T et al (2009) The avian Z-linked gene *DMRT1* is required for male sex determination in chicken. *Nature* 461:267–271

- Torres TT, Metta M, Ottenwalder B, Schlotterer C (2008) Gene expression profiling by massively parallel sequencing. *Genome Res* 18:172–177
- Van Eenennaam AL, Van Eenennaam JP, Medrano JF et al (1999) Evidence of female heterogametic genetic sex determination in white sturgeon. *J Hered* 90:231–233
- Vera JC, Wheat CW, Fescemyer HW et al (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 17:1636–1647
- Weber APM, Weber KL, Carr K et al (2007) Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiol* 144:32–42
- Wicker T, Schlagenhauf E, Graner A et al (2006) 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 7:275
- Wuertz S, Gaillard S, Barbisan F et al (2006) Extensive screening of sturgeon genomes by random screening techniques revealed no sex-specific marker. *Aquaculture* 258:685–688
- Zhu YY, Machleder EM, Chenchik A et al (2001) Reverse transcriptase template switching: A SMART (TM) approach for full-length cDNA library construction. *Biotechniques* 30:892–897