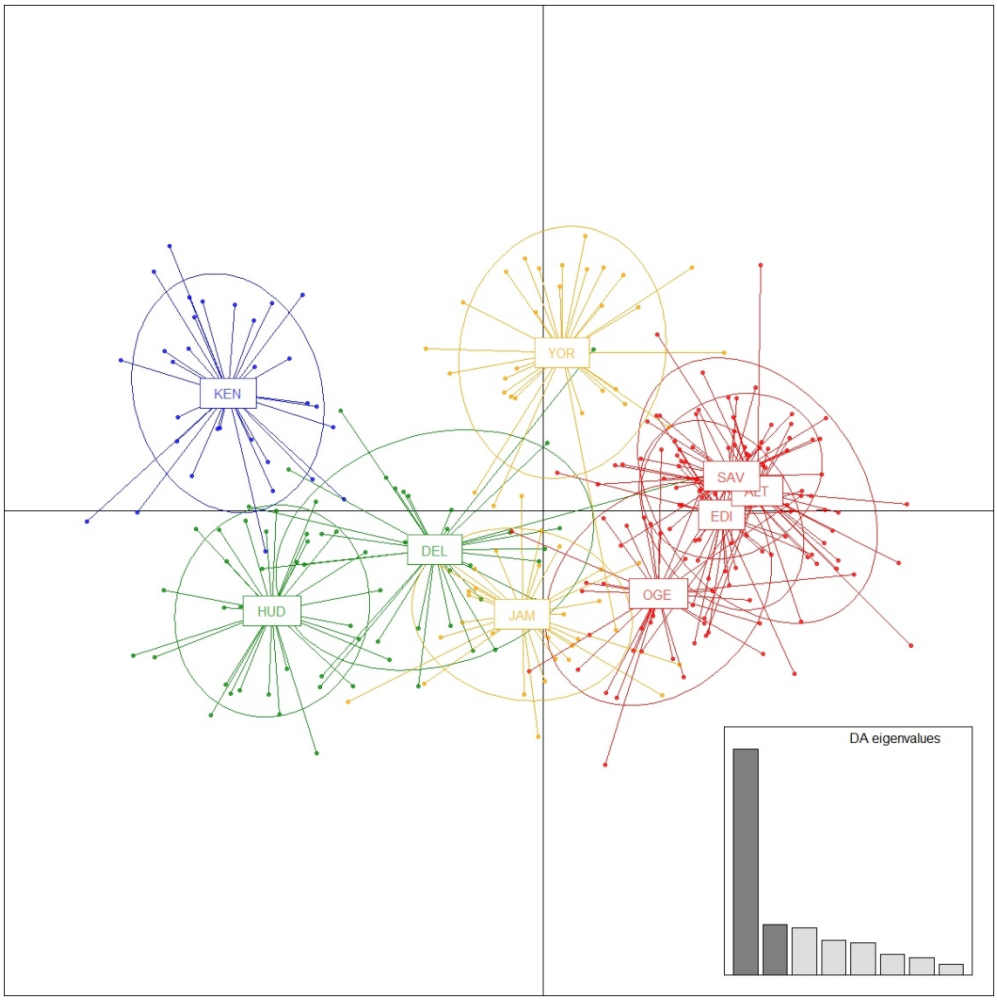
OCS Study
BOEM 2020-062

Using Advanced Population Genomics to Better Understand the Relationship Between Offshore and Spawning Habitat Use for Atlantic Sturgeon



US Department of the Interior
Bureau of Ocean Energy Management
Environmental Studies Program

BOEM
Bureau of Ocean Energy Management

OCS Study
BOEM 2020-062

Using Advanced Population Genomics to Better Understand the Relationship Between Offshore and Spawning Habitat Use for Atlantic Sturgeon

January 2020

Authors:
David C. Kazyak, Aaron Aunins, Robin Johnson, Barb Lubinski, Michael Eackles, and Tim L. King

Prepared under IAA No. M15PG00044
By
U.S. Geological Survey Leetown Science Center
11649 Leetown Road
Kearneysville, WV 25430

US Department of the Interior
Bureau of Ocean Energy Management
Environmental Studies Program

DISCLAIMER

This study was funded, in part, by the US Department of the Interior, Bureau of Ocean Energy Management (BOEM), Environmental Studies Program, Washington, DC, through Interagency Agreement Number M15PG00044 with the US Geological Survey (USGS) Leetown Science Center. This report has been technically reviewed by BOEM, and it has been approved for publication. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of BOEM. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REPORT AVAILABILITY

To download a PDF file of this report, go to the US Department of the Interior, Bureau of Ocean Energy Management Data and Information Systems webpage (http://www.boem.gov/Environmental-Studies-EnvData/), click on the link for the Environmental Studies Program Information System (ESPIS), and search on 2020-062. The report is also available at the National Technical Reports Library at https://ntrl.ntis.gov/NTRL/.

CITATION

Kazyak DC, Aunins A, Johnson R, Lubinski B, Eackles M, King TL. 2019. Using advanced population genomics to better understand the relationship between offshore and spawning habitat use for Atlantic sturgeon. Sterling (VA): US Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2020-062. 70 p.

ABOUT THE COVER

The cover image is an ordination plot showing the relationships among nine populations of Atlantic sturgeon as resolved from 118 nuclear single nucleotide polymorphism (SNP) markers using discriminant analysis of principal components.

ACKNOWLEDGMENTS

Our foray into sturgeon genomics owes its existence to Dr. Tim King, who had the vision to propose this initiative and led work on the project during its early stages. Unfortunately, he passed away unexpectedly before the research was completed. We also thank the numerous sturgeon researchers who submitted tissue samples to the Atlantic Coast Sturgeon Tissue Research Repository, including Jason Kahn, Amanda Higgs, Mark Fritts, Matt Fisher, Gayle Zydlewski, and Bill Post, without whom this work would have been intractable. In addition, we specifically thank Evan Ingram at Stony Brook University for providing tissues samples collected in a separate BOEM-funded project (cooperative agreement M16AC00003) so that assignment testing could be conducted on Atlantic sturgeon from the continental shelf. Many other researchers were responsible for collecting samples that were used to develop our genetic and genomic baselines, and we thank them for their contributions. This study was funded by the U.S. Department of the Interior, Bureau of Ocean Energy Management through Interagency Agreement M15PG00044 with the U.S. Geological Survey.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

BOEM	Bureau of Ocean Energy Management
bp	base pair(s)
DAPC	discriminant analysis of principal components
DPS	Distinct Population Segment
ESA	Endangered Species Act
GT-seq	Genotyping-in-Thousands by sequencing
HWE	Hardy-Weinberg equilibrium
LSC	Leetown Science Center
LSC-KCGL	Leetown Science Center – King Conservation Genetics Laboratory
mtDNA	Mitochondrial DNA
PCR	polymerase chain reaction
RAD-seq	restriction site associated DNA sequencing
RRJ	river-resident juvenile
SNP	single nucleotide polymorphism
TL	total length
USGS	U.S. Geological Survey
VCF	vCard file

Executive Summary

Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus*) are a large-bodied anadromous fish that historically supported important fisheries along the east coast of the United States. Following years of overharvest and habitat degradation, populations experienced severe declines. In 2012, the National Marine Fisheries Service listed Atlantic sturgeon under the Endangered Species Act (ESA; 61 FR 4722). Their listing named five Distinct Population Segments (DPSs), predicated on genetic groups composed of geographically proximate populations.

Federal management of Atlantic sturgeon presents challenges, as sturgeon from each of the five DPSs mix extensively in coastal and marine habitats, and take and recovery progress must be evaluated separately for each unit. Genetic assignment testing based on mitochondrial and microsatellite markers allows individuals to be assigned back to their natal river and DPS. However, this approach is not perfect, and some individuals may be incorrectly assigned. Recent advances in genomics offer the potential of a higher resolution approach to genetic assignment testing, and thus may reduce uncertainty associated with assignment testing. In addition, genomics allows a greater number of markers to be examined from across a broader portion of the sturgeon genome, thus may provide an enhanced perspective of population structure for the species and potentially allow other previously intractable questions to be addressed (Bernatchez et al. 2017, Supple and Shapiro 2018).

We used next-generation sequencing to develop a draft genome for Atlantic sturgeon and identify single nucleotide polymorphisms (SNPs) that could be used to resolve the natal river and DPS of individual Atlantic sturgeon. We identified 1,210 candidate SNPs within the nuclear genome as well as 49 SNPs within the mitochondrial genome. After filtering and review, we selected 161 nuclear SNPs and 39 mitochondrial SNPs for further testing and evaluation. We used genotyping-in-thousands by sequencing (GT-seq) to simultaneously sequence nuclear SNP loci, mitochondrial SNP loci, and the existing panel of 12 microsatellite loci. This effort required a pilot sequencing run on a single sturgeon sample to test marker amplification and refine primer strengths, followed by a series of sequencing runs to generate baseline data for 288 individuals representing nine populations of Atlantic sturgeon in four DPSs.

Using baseline data from the nine populations, we ran a series of genomic analyses to characterize diversity within and among populations, providing a benchmark for this species using the new SNP markers. Allelic richness was similar for all populations, although there was a general trend of more northern population containing greater levels of allelic richness. Interestingly, we observed linkage disequilibrium among many pairs of loci within many populations. This might be the result of physical linkage but could also suggest these populations are recovering from genetic bottlenecks and/or are effectively small, leading to specific haplotypes to be favored by chance. Pairwise differentiation among populations varied among the populations (F_{ST} range: 0.010–0.098) and was significantly correlated (r = 0.771; $P < 0.001$) to pairwise F_{ST} observed using microsatellite markers). Population clustering and ordination techniques using the new genomic data both support an overall population structure that is similar to the current DPS management units (which were developed primarily based on microsatellite genetic data). Overall, this suggests that existing microsatellite markers and the panel of SNP markers developed in this study provide similar information about the populations structure and ecology of Atlantic sturgeon. Given the observed differences in allele frequencies among populations, our genomic baseline supports previous assertations that Atlantic sturgeon show natal homing, despite mixing extensively in marine waters during non-breeding periods. Lower levels of differentiation between populations in the South Atlantic DPS suggest that populations in this region may have greater levels of gene flow relative to their more northerly conspecifics, which has also previously been suggested based

on microsatellite data. The observed differentiation among populations provides the necessary foundation for determining the natal river and DPS of Atlantic sturgeon using assignment testing.

We tested the utility of our new genomic baseline for resolving the population and DPS of Atlantic sturgeon. Our nuclear SNP markers showed utility for identifying the origin of unknown Atlantic sturgeon samples, as 86.5% were assigned to the correct DPS, and 66.3% were assigned to the correct natal river. However, since this study was funded, the U.S. Geological Survey (USGS) Leetown Science Center – King Conservation Genetics Laboratory has made significant improvements to their microsatellite genetic baseline, which now performs more effectively than our new genomic approach (the genetic baseline includes 12 populations and 5 DPSs, and correctly assigns 95.8% of individuals to DPS and 84.9% of individuals to their natal population using 12 microsatellite loci). We conducted an ad hoc exploration of how additional microsatellite or nuclear SNP loci may further improve the accuracy of assignment testing. We found that additional microsatellite markers are likely to result in greater improvements in assignment efficiency than additional nuclear SNPs. However, a much larger number of SNP loci (which if identified could be sequenced using other methods that are now available; e.g., the RAD-Capture approach published by Ali et al. 2016) could produce assignment efficiencies that are greater than what is currently feasible using microsatellites. In the absence of further research and development of additional SNP markers for Atlantic sturgeon (possibly using an approach other than GT-seq), the existing microsatellite loci are the most effective means available to determine the natal river and DPS of Atlantic sturgeon encountered in offshore waters.

Because our new genomic markers were less effective than the existing panel of 12 microsatellite markers, we chose to use the existing microsatellite markers to assign Atlantic sturgeon captured in another study funded by the Bureau of Ocean Energy Management (BOEM) (cooperative agreement M16AC00003; “Monitoring Endangered Atlantic Sturgeon and Commercial Finfish Habitat Use Offshore New York”) following consultation with our project officer at BOEM. Using this approach, we genotyped and assigned 186 Atlantic sturgeon captured in coastal waters off the Rockaway Peninsula, New York. The vast majority of these sturgeon were assigned to the New York Bight DPS (94.62%), and most appear to belong to the Hudson River population (87.10%) with smaller contributions from the Delaware River population (7.53%). Smaller contributions ($\leq$ 2.15%) were observed from six other populations, including those from the James, York, Kennebec, Ogeechee, and Edisto rivers. Although most of the fish we assigned were assigned to the nearest spawning rivers (Hudson and Delaware), the contributions from distant rivers is consistent with the propensity of this species to move long distances and form mixed stock aggregations along the continental shelf. This finding indicates that spawning populations (and their corresponding DPS) from distant locations may potentially be impacted by offshore activities. In fact, activities in this region of the New York Bight could negatively impact Atlantic sturgeon population from at least four different DPSs. Genetic or genomic assignment testing remains an essential tool to characterize potential impacts to Atlantic sturgeon populations and could be applied more broadly to better characterize potential impacts of activities in other locations.

1 Background and Overview of Study Approach

Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus*) are a large, ancient fish native to the Atlantic coast of North America. As an anadromous species, Atlantic sturgeon spawn in freshwater over rocky substrate in rivers from the southeastern U.S. to maritime Canada (Hilton et al. 2016). Although juvenile sturgeon are confined to their natal rivers and estuaries, sub-adult and adult Atlantic sturgeon spend the majority of their lives in nearshore coastal areas and the continental shelf, where they are known to exhibit long distance migrations and occupy habitats far from their natal rivers. The species is long-lived and late to mature, and does not spawn annually. The life history of Atlantic sturgeon makes them vulnerable to numerous anthropogenic impacts.

Historically, Atlantic sturgeon supported important fisheries for meat and caviar. Archaeological evidence suggests that sturgeon were important to early European settlers as a food source (Bowen and Andrews 2000) and an early commercial export (Balazik et al. 2010). Harvest of Atlantic sturgeon peaked in the late 1800s, when 3,200 metric tons were reported to have been harvested in a single year (Secor 2002). By 1901, Atlantic sturgeon populations had crashed coastwide (US Commission of Fish and Fisheries 1884–1905). Landings continued at greatly reduced levels through most of the 20th century and populations remained severely depleted. By the mid-1990s, many populations were thought to be at historic lows.

As the result of dramatic population declines and continued harvest pressure, many states enacted and implemented harvest restrictions or moratoria in their jurisdictions. In 1998, the Atlantic States Marine Fisheries Commission established a coastwide moratorium on the sturgeon fishery in U.S. waters. In 2012, the National Marine Fisheries Service enacted further legal protections when it listed Atlantic sturgeon under the Endangered Species Act (ESA). The final rules established five Distinct Population Segments (DPSs; USFWS and NMFS 1996), based on distinct genetic and physical characteristics that had been described in a previous status review (ASSRT 2007). The New York Bight, Chesapeake Bay, Carolina, and South Atlantic DPSs were each listed as endangered, whereas the Gulf of Maine DPS was listed as threatened.

Under the ESA, DPSs are essentially managed as different species. This presents a significant challenge for Atlantic sturgeon (75 FR 61904), as sub-adult and adult Atlantic sturgeon frequently form mixed stock aggregations in marine and estuarine waters (Waldman et al. 1996, Dunton et al. 2012, Wirgin et al. 2015). Consequently, non-breeding sub-adult and adult individuals can only be assigned to their DPS using molecular genetic techniques.

To date, most molecular work on Atlantic sturgeon has employed a modest number of microsatellite markers (described in May et al. 1997, King et al. 2001, Henderson-Arzapalo and King 2002) to assess population structure and run individual-based assignment tests. This approach provided the underlying information that supported the establishment of DPSs and has proven fairly useful in addressing applied conservation questions. Notably, the USGS Leetown Science Center (LSC) – King Conservation Genetics Laboratory (LSC-KCGL) has developed a microsatellite baseline comprised of over 1,600 individuals from 12 rivers. This genetic baseline can be used to assign Atlantic sturgeon to their natal area (DPS or river) and provides much needed information to help assess the impacts of specific permitted activities on each management unit. This genetic baseline (as well as a similar baseline produced by New York University) has provided the basis for many studies characterizing the composition of mixed stock aggregations and describing patterns of population structure and genetic diversity (Grunwald et al. 2008, Waldman et al. 2013, O'Leary et al 2014, Wirgin et al. 2015).

Despite considerable research and management attention, Atlantic sturgeon continue to face a broad suite of threats to their recovery, and these most likely vary by DPS. Sturgeon are taken as incidental catch in

commercial fisheries and may be killed by ship strikes (NMFS 2010). Dredging activities and marine construction can also result in sturgeon mortality (NMFS 2010). Recent interest in the development of offshore energy resources on the continental shelf raises the importance of understanding potential impacts of development activities on this federally listed species. Until recently, Atlantic sturgeon were thought to primarily use shallow inshore habitats while in marine environments (Stein et al. 2004). However, recent telemetry work funding by the Bureau of Ocean Energy Management (BOEM) demonstrates that Atlantic sturgeon use offshore wind energy areas extensively during certain periods (Ingram et al. 2019), and the stock composition of these fish is largely unknown. An improved understanding of the makeup of Atlantic sturgeon using offshore areas will help managers understand the impacts of development activities and enhance recovery and monitoring efforts (NMFS 2010).

Rapid advances in conservation genomics offer new opportunities to provide conservation relevant information to support the management of fishery resources. In particular, new technologies allow sequencing vast amounts of DNA sequence at reasonable costs, offering unprecedented insight into the molecular underpinnings of living organisms. For conservation practitioners, these new approaches offer the prospect of much larger numbers of loci from throughout the genome, including not only neutral markers but also those that may be under selective pressures. The current study was designed to leverage recent advances in genomics to develop new molecular tools to support the management of Atlantic sturgeon. A genomic approach could provide a higher resolution tool to characterize mixed stock aggregations, reveal undetected population structure, and provide insight into patterns of natural selection.

Our approach to develop genomic tools for Atlantic sturgeon was based on four key components:

1. Marker discovery
2. Refinement of a cost-effective higher throughput genomics approach
3. Generation of a genomic baseline
4. Application to select set of BOEM-funded samples to demonstrate proof of concept

For the marker discovery component, we used two approaches. First, we performed shotgun genomic sequencing to generate a draft reference genome to assist with identifying the locations of the single nucleotide polymorphism (SNP) loci to be ascertained and to provide flanking sequence to the SNPs for the design of primers for polymerase chain reaction (PCR) amplification. Next, we sequenced a small number of individuals (collected by cooperating scientists from other agencies) from select populations across the native range of the species using double digest RAD-seq for mapping against the new draft reference genome to identify a suite of candidate SNP loci that were variable among populations and functionally diploid. To develop these SNPs into a cost-effective approach for genotyping large numbers of individuals for population scale work, we operationalized a genotyping-in-thousands by sequencing (GT-seq; Campbell et al. 2014) technique based on these markers. With this new approach, we genotyped hundreds of Atlantic sturgeon at the new loci we identified to create a representative baseline for populations across the eastern U.S. We conducted basic population genetic analyses using this new class of markers and evaluated the effectiveness of our genomic approach for assigning individuals to their population of origin. Although we initially intended to apply our new genomic approach to provide assignments for samples collected in another BOEM-funded study, our preliminary results suggested that we were unable to attain comparable levels of assignment success using the new GT-seq panel, and many more SNP loci would likely be necessary to achieve the assignment power of the current microsatellite-based approach. As a result, we consulted with BOEM and instead applied the existing microsatellite markers to Atlantic sturgeon collected by Stony Brook University in marine waters near Long Island, New York.

2 Marker Discovery

As part of our research plan to identify candidate SNP markers for studying Atlantic sturgeon, we first generated a draft genome using individuals from multiple populations across their native range. The purpose of this draft assembly was to provide a genomic reference for mapping short reads from RAD-seq libraries against to identify candidate SNP markers. In addition, by having long contig sequences to which the RAD-seq reads could be mapped, we would potentially have flanking sequence that could be used for primer design to amplify the SNP. Following completion of the draft genome, we performed double digest RAD-seq on a sample of sturgeon from throughout their native range to ascertain their SNP. The following sections describe the construction of the reference genome, double digest RAD-seq, and bioinformatic processing of these data to obtain candidate SNP markers for further investigation.

2.1 Shotgun Sequencing for *De* Novo Reference Assembly

Several shotgun genomic libraries of Atlantic sturgeon from multiple populations were 150 base pairs (bp) single end sequenced on a NextSeq 500 at LSC utilizing high output cartridges (**Table 1**). Briefly, 1 nanogram of DNA was used to make dual indexed libraries with the Illumina Nextera XT library preparation kit following the manufacturer's instructions. Reads from each demultiplexed library were downloaded from Illumina's Basespace and imported into CLC Genomics Workbench (Qiagen). CLC parameters for length and quality trimming included a quality limit of 0.025, ambiguous nucleotide limit of 1, and minimum read length of 30 bp. After trimming, all reads from among the libraries were combined (n = 1,014,479,716) and used for *de novo* assembly (automatic bubble size = yes, minimum contig length = 200, word size = 25, perform scaffolding = yes). This assembly yielded 1,317,463 contigs with an average length of 753 bp (min = 200 bp, max = 13,704 bp), and N_{50} = 1,104 (deposited at GenBank under the accession JABEPO000000000).

When developing a draft reference genome assembly, the initial assembly may yield multiple highly similar contigs as a consequence of sequencing repetitive genomic regions, especially within a tetraploid genome like the Atlantic sturgeon. To reduce the over-representation of these repetitive regions in the final draft assembly, we clustered the contigs at 95% similarity using cd-hit-est (Li and Godzik 2006). This reduced the number of contigs to 1,215,580.

2.2 Double Digest RAD-seq

DNA of 16 adult Atlantic sturgeon from among six rivers sampled during their respective spawning migrations was identified for double digest RAD-seq (**Table 2**). These samples were chosen because they span a large geographic range of Atlantic sturgeon natal rivers and should therefore represent genetic variation from throughout this range.

Double digest RAD-seq library preparation and sequencing was performed by the sequencing core facility Admera Health, South Plainfield, NJ. Briefly, two 4 base-cutting restriction enzymes NlaIII and MluCI were used to digest 200 ng DNA of each sample in separate reactions at 37 ºC for 2 hours, followed by heat inactivation of the enzyme at 65 ºC for 20 minutes. Five base pair unique indexed sequencing adapters were ligated to the NlaIII cut site to enable unambiguous demultiplexing of each sample after sequencing. The size selection of adapter-ligated fragments was 410 base pairs (this includes primers and barcode lengths), which would result in sequenced fragments of approximately 284 bp after demultiplexing. Samples were 150 bp paired-end sequenced on an Illumina HiSeq.

Table 1. Atlantic sturgeon samples used for SNP discovery.

Individual	Population	Run	Number of quality trimmed single end reads
AoxyDe-5097	Delaware	Aoxy-GBS-3-05-07-16	14940892
AoxyDe-5099	Delaware	Aoxy-GBS-3-05-07-16	393514
AoxyDe-5102	Delaware	Aoxy-GBS-3-05-07-16	9957104
AoxyDe-5101	Delaware	Aoxy-GBS-3-05-07-16	9315814
AoxyDe-5100	Delaware	Aoxy-GBS-3-05-07-16	10088973
AoxyDe-5103	Delaware	Aoxy-GBS-3-05-07-16	10844866
AoxyDe-5091	Delaware	Aoxy-GBS-3-05-07-16	8435277
AoxyDe-5098	Delaware	Aoxy-GBS-3-05-07-16	12143633
AoxySav-11263	Savannah	Aoxy-GBS-3-05-07-16	22491929
AoxySav-11286	Savannah	Aoxy-GBS-3-05-07-16	21426709
AoxySav-11275	Savannah	Aoxy-GBS-3-05-07-16	13285966
AoxySav-11278	Savannah	Aoxy-GBS-3-05-07-16	8404217
AoxySav-11258	Savannah	Aoxy-GBS-3-05-07-16	28589038
AoxySav-11279	Savannah	Aoxy-GBS-3-05-07-16	14082252
AoxySav-11251	Savannah	Aoxy-GBS-3-05-07-16	13146957
AoxySav-11280	Savannah	Aoxy-GBS-3-05-07-16	15987005
AOS-95	Edisto	AoxyGBS-04-11-16	7861517
AOS-94	Edisto	AoxyGBS-04-11-16	8760466
AOS-98	Edisto	AoxyGBS-04-11-16	15741224
AOS-92	Edisto	AoxyGBS-04-11-16	22144652
AOS-96	Edisto	AoxyGBS-04-11-16	4387607
AOS-97	Edisto	AoxyGBS-04-11-16	17904667
AOS-93	Edisto	AoxyGBS-04-11-16	17746879
AOS-99	Edisto	AoxyGBS-04-11-16	11661017
AOK-62	Kennebec	AoxyGBS-04-11-16	36143734
AOK-59	Kennebec	AoxyGBS-04-11-16	5028440
AOK-61	Kennebec	AoxyGBS-04-11-16	4682542
AOK-63	Kennebec	AoxyGBS-04-11-16	6093374
AOK-58	Kennebec	AoxyGBS-04-11-16	1673641
AOK-60	Kennebec	AoxyGBS-04-11-16	4363991
AOK-57	Kennebec	AoxyGBS-04-11-16	19627774
AOK-56	Kennebec	AoxyGBS-04-11-16	15332004
AOOG-102	Ogeechee	AoxyGBS-04-11-16	8347448
AOOG-111	Ogeechee	AoxyGBS-04-11-16	11769619
AOOG-107	Ogeechee	AoxyGBS-04-11-16	9400498
AOOG-104	Ogeechee	AoxyGBS-04-11-16	12288583
AOOG-112	Ogeechee	AoxyGBS-04-11-16	18721556
AOOG-116	Ogeechee	AoxyGBS-04-11-16	26657280
AOOG-113	Ogeechee	AoxyGBS-04-11-16	6701612
AOOG-115	Ogeechee	AoxyGBS-04-11-16	6996640

Individual	Population	Run	Number of quality trimmed single end reads
AOSL-01	Saint Lawrence	AoxyGBS-04-11-16	8796548
AOSL-08	Saint Lawrence	AoxyGBS-04-11-16	13705049
AOSL-06	Saint Lawrence	AoxyGBS-04-11-16	9700928
AOSL-07	Saint Lawrence	AoxyGBS-04-11-16	10915608
AOSL-05	Saint Lawrence	AoxyGBS-04-11-16	16163180
AOSL-02	Saint Lawrence	AoxyGBS-04-11-16	8874970
AOSL-03	Saint Lawrence	AoxyGBS-04-11-16	9095166
AOSL-04	Saint Lawrence	AoxyGBS-04-11-16	9409230
AOY-15-35	York	Aox-GBS-12-02-15	6191806
AOY-15-65	York	Aox-GBS-12-02-15	4327142
AOY-15-23	York	Aox-GBS-12-02-15	4993244
AOY-15-48	York	Aox-GBS-12-02-15	4283733
AOY-15-20	York	Aox-GBS-12-02-15	3902816
AOY-15-42	York	Aox-GBS-12-02-15	7039248
AOY-15-29	York	Aox-GBS-12-02-15	6546380
Aoxy-6556	Altamaha	Aox-GBS-12-02-15	27706376
Aoxy-6548	Altamaha	Aox-GBS-12-02-15	35882449
Aoxy-6550	Altamaha	Aox-GBS-12-02-15	49593457
Aoxy-6555	Altamaha	Aox-GBS-12-02-15	82440135
Aoxy-6553	Altamaha	Aox-GBS-12-02-15	54294381
Aoxy-6554	Altamaha	Aox-GBS-12-02-15	45526251
James_Spring-5	James	Aox-GBS-12-02-15	3359337
James_Fall-5	James	Aox-GBS-12-02-15	5178678
AOH-96-13	Hudson	Aox-GBS-12-02-15	20096662
AOH-96-11	Hudson	Aox-GBS-12-02-15	5982964
AOH-96-04	Hudson	Aox-GBS-12-02-15	7503405
AOH-96-15	Hudson	Aox-GBS-12-02-15	10368553
AOH-97-06	Hudson	Aox-GBS-12-02-15	8897969
AOH-97-09	Hudson	Aox-GBS-12-02-15	4351801
AOH-97-01	Hudson	Aox-GBS-12-02-15	8487941
AOH-97-05	Hudson	Aox-GBS-12-02-15	7293398

Table 2. Samples of 16 Atlantic sturgeon DNA subjected to double digest RAD-sequencing for the identification of nuclear SNPs.

Natal River	Collection Year	Leetown Science Center ID	Admera ID	Concentration (ng/µl)
St. Lawrence	2013	AOSL-13-012	AOX_01	56.0
St. Lawrence	2013	AOSL-13-027	AOX_02	78.2
Kennebec	2010	AOK-051	AOX_03	21.1
Kennebec	2011	AOK-064	AOX_04	26.1
Hudson	2015	AoxyH15-9167	AOX_05	62.3
Hudson	2015	AoxyH15-9168	AOX_06	44.0
Hudson	2015	AoxyH15-9171	AOX_07	56.1
Hudson	2015	AoxyH15-9173	AOX_08	108.0
James	2013	AoxyJ13-9633	AOX_09	88.2
James	2012	AoxyJ12-9639	AOX_10	74.0
James	2012	AoxyJ12-9640	AOX_11	83.5
James	2012	AoxyJ12-9643	AOX_12	95.4
Edisto	1998	AOE-108	AOX_13	20.6
Edisto	1998	AOE-109	AOX_14	22.8
Altamaha	2011	AoxyALT11-6789	AOX_15	68.8
Altamaha	2011	AoxyALT11-6824	AOX_16	157.0

Notes: Samples were sequenced by the provider Admera Health, South Plainfield, NJ. Please refer to the text for details of library construction.

Two sequencing runs of the same libraries were ultimately performed to reach the minimum read number threshold set by Admera Health. The first run produced 133,891,708 read pairs with 49,988,713 forward reads demultiplexed among the 16 samples. The average number of reads per sample was 3.1 million, and the standard deviation was 1.5 million. For the second run, the total number of read pairs was 94,097,618 with 92,177,684 reads demultiplexed among the 16 samples. The average number of reads in the second run was 5.8 million, and the standard deviation was 0.6 million. When the reads for each individual were combined across runs, the average number of reads per sample increased to 8.9 million, and the standard deviation was 1.4 million.

2.3 SNP Identification

The result of the double digest RAD-seq libraries was large numbers of genomic fragments beginning with the NlaIII cut site on the forward read of each pair. To identify SNPs within these tags, we first sought to reduce the tags to only a set of those that could be mapped to a contig in our draft sturgeon genome assembly. The main reason for this was to place each tag within the larger contig sequence from which they were derived, allowing more opportunity for primer design in a high throughput SNP genotyping assay. For example, if an SNP within a tag is close to the 5` or 3` end of the 150 bp RAD-seq fragment, there is limited opportunity for primer development, whereas if the contig sequence the tag maps to is available, potentially a few hundred bases of contig sequence may be used for primer design. To accomplish this, all tag reads from the 16 sturgeon across both sequencing runs were clustered into provisional loci using a pairwise-identity threshold of 85% in the program Vsearch (Rognes et al. 2016). Representatives of each cluster were then aligned to the draft sturgeon genome contigs using BLASTN, resulting in $157x10^3$ tags. Only these tags that mapped to the draft genome were retained for subsequent SNP identification.

Variable sites within tags were identified by mapping all reads to the cluster representatives identified by Vsearch using bowtie2 (Langmead and Salzberg 2012) with the fast and local parameter switches and a minimum mapping quality of 30 (phred scaled). The samtools mpileup function (Li et al. 2009) was then used to tabulate the base counts occurring at each position and their statistical distributions, with base

alignment qualities recalibrated using the '-E' flag in mpileup. Variant sites were then discriminated from sequence error using the bcftools call function and multiallelic calling model (Li et al. 2009). However, given the functional diploidization of the Atlantic sturgeon genome (Ludwig et al. 2001), ultimately only bi-allelic variants were retained. Additional quality filters included the exclusion of sites within three bases of an indel (due to potentially reduced alignment accuracy near indels), exclusion of sites within five bases of either end of a tag (also due to potentially reduced alignment accuracy), and exclusion of tags with more than three variant sites (due to potential mapping bias against the most diverged haplotypes). Finally, only loci with a minor allele frequency of ≥ 5% within the read pool were retained, as low frequency alleles are more difficult to discriminate from error, and the autosomal sample size limited the theoretical minimum allele frequency to 1/32 (~3%). Application of these additional quality filters resulted in a set of 1,210 candidate SNP loci.

A histogram of minor allele frequencies among these 1,210 loci indicated an over-representation of SNPs with intermediate allele frequencies, suggesting undetected paralogy. Paralogous alignments, in which similar but non-allelic sequences are mistakenly treated as orthologous, is a significant source of error for SNP detection in polyploid organisms, even if the organism is functionally diploid (Clevenger et al. 2015). We therefore performed additional analyses of locus coverage, as well as comparisons with known population structure from existing microsatellite-based population genetic analyses, to further restrict the number of candidate SNPs. The goal of this additional filtering was to identify loci that had similar coverage ranges in test populations and gave estimates of genetic distances between populations comparable to existing microsatellite data.

We first used a subset of the sequence data previously generated by shotgun sequencing for our *de novo* reference assembly to estimate allele frequencies at the 1,210 candidate loci after grouping the sequenced samples into eight populations (Kennebec, Hudson, Ogeechee, Satilla, Saint Lawrence, Savannah, Delaware, and York). These shotgun genomic reads were mapped against the contigs of the reference assembly containing the 1,210 candidate SNPs with bowtie2 using the sensitive and local parameter switches and a minimum phred-scaled mapping quality of 20. SAMtools mpileup was performed as before to generate allele counts for each variant. Based on these results, we removed 90 SNPs with excessive or low coverage within each population. Next, we added sequence data from the RAD-seq runs of four populations for which there were comparable shotgun genomic data from the same populations (Hudson, Kennebec, Saint Lawrence, Savannah). Balanced coverage in all four populations was found for 625 SNPs (i.e., within the range 15–30X, which was chosen subjectively based inflection points as well as the number of samples from each population). We then calculated Nei's pairwise genetic distance (Nei 1972, 1978) for these 625 SNPs and compared them to the same metric for previously estimated microsatellite allele frequencies for those populations. Because the most variable 161 SNPs (i.e., with allele frequency coefficient of variation ≥ 0.5) showed somewhat greater concordance with microsatellite patterns than did the entire set of 625, these were chosen for development of a high throughput SNP genotyping assay.

3 Development of a Cost-effective High Throughput Genomics Approach and Creation of a Population Genomic Baseline

The next step in our study required the genotyping of large numbers of Atlantic sturgeon at the SNP loci we identified and selected for further analysis. Although double digest RAD-seq can be used for repeatable genotyping of specific loci if enzyme digestion and size selection of fragments are precisely controlled in each sequencing run, these conditions are not always feasible in practice. Consequently, some loci may have varying levels of recovery across runs leading to missing data and analytical challenges. Therefore, RAD-seq is usually employed for initial identification of a large number of loci, which are then ported over to a different sequencing methodology for genotyping a large number of individuals for population genomic analyses, especially when sequencing needs to be conducted repeatedly over longer periods of time and conducted at multiple laboratories.

In order to develop a tractable method for repeatably genotyping large numbers of Atlantic sturgeon at the SNP loci we identified, we reviewed the scientific literature and evaluated a suite of different published methods. We considered chip-based assays, capture-based approaches, methods that focus on restriction site associated loci, and techniques that use multiplexed amplicon sequencing. Key factors in our review included the number of loci that could be assayed, repeatability of sequencing specific loci, cost per sample, and flexibility in running batches containing different numbers of samples. A viable approach needed to be reproducible, scalable, and cost-effective. SNP "chips" using microarray technology (e.g., the 'ssalar01' Affymetrix array for Atlantic salmon SNP genotyping; Houston et al. 2014) and hybridization probe based sequencing technologies (e.g., RAD-Capture; Ali et al. 2016) are efficient means of genotyping thousands of SNP loci in hundreds of individuals but require a substantial up-front investment in array or probe design. Because we had identified 161 SNP loci for additional investigation and needed to first generate more data to assess their utility before application to thousands of individuals, we chose to develop a genotyping method called genotyping-in-thousands by sequencing (GT-seq; Campbell et al. 2014) for Atlantic sturgeon that would be readily repeatable and cost less than developing a custom microarray or designing hybridization.

GT-seq is an SNP genotyping method developed by Campbell et al. (2014) that leverages highly multiplexed PCR and massively parallel sequencing to genotype a modest number (~50 to ~500) SNP markers at a reasonable cost per sample. Briefly, primers for all loci that flank either side of each SNP locus are designed and combined into a primer cocktail. These primers contain adapters that enable incorporation of unique barcode indices and sequencing adapters in subsequent PCR steps to allow multiple samples to be pooled and sequenced on a massively parallel sequencer (Illumina platform, Ion Torrent, etc.). We developed our GT-seq approach using an Illumina MiSeq instrument for sequencing, as this platform is in widespread use and is the most cost-effective sequencing option for the number of samples anticipated. Barcode design and indexing strategies can be tailored to fit the level of multiplexing, throughput, and coverage desired. We decided to employ GT-seq for the genotyping of 288 Atlantic sturgeon representing nine populations at 161 loci to generate a genomic baseline for individual assignment to DPS and river of origin. These individuals were selected from samples that were already included as part of the microsatellite baseline at LSC, and they represent individuals captured in spawning rivers as either river-resident juveniles (< 500 mm total length [TL]) or adults (≥ 1,500 mm TL).

As a novel adaptation of the GT-seq method, we sequenced three classes of markers simultaneously: nuclear SNPs, mitochondrial SNPs, and microsatellite loci. If tractable, this approach would allow inferences to be made across all three classes of markers, thus providing a comprehensive picture of population structure and maximizing the amount of information available for assignment testing.

3.1 GT-seq Methods

3.1.1 Primer Design for GT-seq

Contigs containing the 161 selected nuclear SNPs were input into the online primer design program BatchPrimer3 (You et al. 2008), which designs large numbers of PCR primers based on user defined parameters. Although this program designs primers that can be used in a multiplex, it does not explicitly evaluate interactions among all possible primer pairs in the multiplex (e.g., heterodimers). As a result, amplification success of each locus can only be assessed through actual PCR and sequencing. We sought to design primers that flanked either side of the candidate SNP marker, creating PCR amplicons of 200 bp or less. A maximum of up to 200 bp or less on each side of the SNP was excised from each contig for input into BatchPrimer (in some cases, the SNP was less than 200 bp from the one of the contig ends, in which case all of the flanking sequence was excised on that side). Settings for BatchPrimer3 were as follows: max amplicon size = 200 bp, min amplicon size = 50, min primer T_m = 55, max primer T_m = 65, max self-complementarity = 4, max 3` self-complementarity = 2, max poly X = 3. Using these settings, 153 primer pairs were successfully designed that met our design parameters (**Table 3**). Primers for the remaining eight loci were unable to be designed within the parameters we set within BatchPrimer and therefore were not included in the GT-seq panel.

In addition to the new nuclear SNP markers, we also decided to include the existing full panel of 12 microsatellite loci used for Atlantic sturgeon assignment testing (**Table 4**), as well as newly designed primers for amplifying several of the mitochondrial SNPs previously identified by T.L. King (**Table 5**) to increase the potential utility of the GT-seq panel. Because all allele sizes for these 12 microsatellite loci were already 350 bp or less, the primers for these loci were not redesigned as paired-end 250 bp or 300 bp sequencing should recover the full microsatellite alleles. Also, it should be noted that the annealing temperatures of these microsatellite primers were largely different than those specified for the primers designed in BatchPrimer, and therefore their amplification success in the multiplex was not optimized. Seven mtDNA primer pairs were designed in BatchPrimer3 using the same settings as for the nuclear SNPs to amplify most of the SNPs identified by T.L. King within *nad*1, tRNAIle, tRNA-Gln, *cox*1, *atp*6, *nad*4L, *nad*5, and the control region (**Table 5**).

In order to sequence multiple samples in one high throughput sequencing run and be able to subsequently demultiplex these reads into individual samples, each sample needs to have a unique index (or barcode) attached to its reads. We employed a two-step approach for dual indexing of each sample for multiplex sequencing on an Illumina MiSeq (Glenn et al. 2016). First, each primer pair (all nuclear SNPs, microsatellites, and mtDNA SNP primers) was synthesized with the sequence ‘CGACAGGTTCAGAGTTCTACAGTCCGACGATC’ on the 5` end of the forward primer and the sequence ‘GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT’ on the 5’ end of the reverse primer. After an initial PCR amplification step consisting of all primers in the multiplex, all amplicons had these adapter sequences incorporated into them. Next, the actual dual indices and flow cell adapters are added in another PCR cycle with the primer ‘AATGATACGGCGACCACCGAGATCTACACXXXXXXXXCGACAGGTTCAGAG*T’ as the forward primer and ‘CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCA*G’ as the reverse primer, where ‘XXXXXXXX’ denotes a unique 8 bp barcode sequence and the ‘*” denotes a phosphorothioate linkage added during primer synthesis. We used 8 unique forward indexed primers and 12 unique reverse indexed primers (**Table 6**), which results in the ability to dual index and sequence 96 samples in one sequencing run.

Table 3. Primers used in the development of a GT-seq assay for Atlantic sturgeon.

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_767	1527	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCATTACAGAACATCAGTCACA	129
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCAGATGGAGATGCTACCT	
contig_2585	1869	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCGCTGTGATGTCAGTGTTGG	158
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGCAGGTGGGAGCTAACTGA	
contig_6443	339	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGATGTGCCAATTAAAGCAA	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGAGAATGGGTATTGCACAGC	
contig_8889	1236	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAGGTTTGCTTGGGACAGC	153
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCAGAACCCTGTGTATGTGCTG	
contig_17037	1619	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCACCTGGCATGGCTGAAAG	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCGGATGAAGCGCCAATC	
contig_24799	775	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCGAAAGCAGAAGCAAGAAA	142
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGACTCCTATTGCATCGCTGTT	
contig_31935	606	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAAGGTATAGCCCATTTGCTT	181
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAATAAAGTGAGTGGAAAGTATAGGT	
contig_39444	2584	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCACAGAAAGTGAAGATTGCATGA	190
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCTTCATCAGTCGGGAAAGT	
contig_44181	571	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTTCCCAAAGGCAAGTCTC	195
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTCCTGAATTGTATATGGCACAA	
contig_48956	96	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCAATGCTGTGTTGGTTGCT	160
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGTCCAGATCAGCGTGAAAC	
contig_61910	325	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCACCACAATTCTCAGCCAAGC	175
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTACCTCTTGCTGCGATGC	
contig_74051	490	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTTTCATTCCTTGACATTTCTG	178
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTCAGGCCAGGATAATTGCT	
contig_75258	616	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGTATAACCTCTGGCGTCACT	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCATTTCCATCCCTGTTGGT	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_80054	1773	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCGGTTCCACTTTAGTTGCTTG	134
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCGGGACATTATTCTCTTGA	
contig_86226	859	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCAACTTCTATGTTCCAGGCATT	183
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGAAATGCAGATGAGGTTGG	
contig_88389	951	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGGCAGTGAGACAGCACTTT	191
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCCTCAAATACCGAACCTT	
contig_91914	6072	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGAATGGCTATCCCAGAGAA	81
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAGCGTCCGTGCTTACTGAT	
contig_93145	774	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTAGGGATACTGCAATCAATAAA	195
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCACACAAAGACAGCGATCA	
contig_93920	151	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTTGGCCTGCCTTGATGA	174
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTTCTCCCTGTACCCGTCAA	
contig_98816	288	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCTCTCTGGTTTCCCATGTA	170
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAATCCAGAATCCAACACCAGT	
contig_102467	514	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTATGTTTCAATGCGGGAAGC	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCAGAACTAAAGGACAGTAATCAG	
contig_119516	2249	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTGCTGACATGCTGAGGTG	159
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGAGAGTGGCAGAGGAGAGC	
contig_124134	2224	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCAGGGTCGTTACTGGCTTC	158
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCAGTCGGGATTAGCTCTAT	
contig_124581	2348	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTGTTATTAGCTCACCATGCAA	168
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTCAATGTATGCTCCGCTGT	
contig_128545	659	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTCCCTGCTGAAGAAGTGTG	175
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGTGCATGACTTTGCAGATTT	
contig_145014	469	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCTCCACCACAAGGGTACAA	195
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGAGCAGGCATGAAAGTGAAA	
contig_146123	3698	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGATCCCAACGATTCAGCAGT	190
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCAATCTGAAACTCCAACC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_154779	984	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAACGGTAAGGGTCATGCAAA	183
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCACCGAGATGATGAAGGAG	
contig_155051	532	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCTGATTAAAGTGTCCGATAGCA	173
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCTGCTGTGTGGAGTCTTGT	
contig_159128	3981	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGGTATCATCCATTCCCATC	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGGTACAAAGTCTCGGCAAT	
contig_167621	382	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGGCTGACTCCTCAAGCAAT	187
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCAGCTCATAGCCATTCACA	
contig_168922	1110	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTGAATCTGGCCTCTTACCC	168
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGAGGGAGCCAAGATGCTTT	
contig_176276	1833	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCTGTCAAGATGCAGAAAGA	144
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCTTTGAGAAACATGCCTTT	
contig_182703	529	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGTGGGTGAAATGGATGGAG	187
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGCAGCCCTTGTGCTACATT	
contig_195684	2875	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTTCAGAGCAATCCAACAAT	197
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCACACAGCACAGCGAGCAG	
contig_197214	688	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCACGCTCACTGCTGATAAA	111
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCCGATAACGTCAAACAGGA	
contig_198579	649	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTCTTGCGTCTGGATTACAA	164
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAACACGCCCAGGTGAGTAA	
contig_198706	212	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAATCTGAAGCTCTTTGTGCAG	196
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTACCACAAACAATCGAAGGA	
contig_204708	190	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTGGGATCTACGTCCAGCAG	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAATATCAAACCTTAACAGAGTACCA	
contig_205045	683	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTGAGACGGGACAGAGCAAC	160
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTCAGGAAGTGGTTGGTGTG	
contig_210463	240	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCATGAACGAGTCACGCCACAG	188
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCCAATCAACACCATTGACC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_211497	481	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGGAGATTTGGACAGGCAGA	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCACAAGAGACGCACGTAGGA	
contig_212828	664	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTGTGAAGAGTGCAGCCAGT	195
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTGCACACTTTGGCACAT	
contig_213642	1202	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCCTATTTGTGTTCTTTATGC	197
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTCTGTCTGTATGTATGGTAGTTTGC	
contig_215639	432	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGTCTTGTAGGCTTGCATGG	198
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCGGTTAGTGTAGCGGTTAGG	
contig_223774	1536	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCACTCTGTTCATCGGCTCT	155
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTGCCTCCATCAACAGGATT	
contig_242683	197	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCAATCCTTATGAGATGGTCTT	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCAGTGACAAGATCCCACAG	
contig_251136	85	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAATGGTTCCACCTCGTCAG	180
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCATCGCTGGGTACTGGATGTC	
contig_252163	990	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTTGGCATCAGTCCACCAAG	156
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCACACACAAACGAAACAC	
contig_252793	913	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCCTGGTACTCGTTTCACTT	121
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCAGGGTAGAAATAACTGCAT	
contig_253460	737	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCACCGTGGGCATGGATATAAA	183
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGAAGTATGAAGGAAGGACGAA	
contig_254281	573	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGTGAAAGCCGTTGCTTACC	192
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCCGTGTAGAACCATCATTCA	
contig_256737	1907	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTGCATCCTTCCCACCTCAT	161
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAAGGCTCCGTCTAATTGTTGA	
contig_260617	915	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGAGCACCACAACAGGCTTTA	153
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCATGAACCCTGAAACATCC	
contig_264486	310	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTGAGAAATGCCGCCACAC	163
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGAATCGCACAGCATCCTTTC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_266918	1436	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCGAAATCTGCAACAACAAA	183
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCTCAGCCCATTTGCAC	
contig_270080	782	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGCGTATTCCAGTTCATAGGG	197
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGAGCTGAACAAACATCTGCT	
contig_270624	353	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTGTCCCATGACTTGTGTCTTT	198
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAGGGCACTTCACGAAAGACA	
contig_273784	467	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGAACCACAAACCAACAGGT	163
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTCGAAACTTGGGATTGTG	
contig_279175	234	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTGTTCCTTTGGGTCATATC	196
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCCGTTTAGGTCAATCAGCA	
contig_292868	2246	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGCCATTCTTCTAGCATCGACA	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTGGTGAATGGAAGTGGAA	
contig_297854	170	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGAAACGCTGCTTCCTAGTG	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTGTGCTTGAGGTATTCAAA	
contig_298604	707	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCACCAAGCAAATGAAGCTA	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCATTGCATTTGAACATAACTG	
contig_299947	2425	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCATCAGCAAGGATTGATTCG	197
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTGTGGTGAGTGTAGCATCCA	
contig_300700	130	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTTGCGCGTAGTGGATTT	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCGACTCAAACAGCGCAAATA	
contig_304245	271	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTTTCAGTTCCCAGGCTTCA	199
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGTGCCAATCATCCCTTACCT	
contig_305031	243	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCAGACCAGCACGGTATTTGTT	169
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCATCAGTCAATAAACCCACCA	
contig_309381	286	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCCTTTGGATTTGCTAAGTGA	170
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGAACCACTGCAAGATTTGA	
contig_323589	1334	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCAAGTAAGCAACAAACTTAGAAAGG	185
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCACTGGCTGTACCGGCTATT	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_327571	242	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAACGTGTCTGTCTTTGCCAGT	165
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGTTCCGGTGTGAACCACT	
contig_335132	559	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGCAGCATGTCCATCACTAA	177
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCACAGGTAGTCGAGGGACAGG	
contig_338725	877	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTGCGGAGAGAAGAACATGAA	200
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCGGGTTCAAAGTTGTCGAAGG	
contig_341935	824	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTCAGCCCTACCTAATAACAAGCA	195
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCACCACGAACTCCACTGCT	
contig_343910	817	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTGATCGGCAGGAATTAAGG	178
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCTATCCGGGCTGTCCATATT	
contig_345578	752	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAAGAACAAGCAACACAAGC	183
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCTTGTTGCTCATCTTGCACCT	
contig_347490	1012	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCCCTACCCGATAACGACATGC	175
		REV: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAAAGGAACGATGCGAACTTG	
contig_349978	217	FWD: CGACAGGTTCAGAGTTCTACAGTCCGACGATCAACAGCAACCCACAAGATCA	188
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAGACAGCACGGCAGAACAC	
contig_351321	1199	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGGTTAAAGCTGGCTGTG	154
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACAAGGGTGAGGCGTGACT	
contig_353457	745	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCATTTCTTCTTTCAGAC	185
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATTCCCTTTGCAGGATCACA	
contig_361665	430	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGGAGCAGGGACACC	154
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGCACCATGCTTTGGATT	
contig_363259	920	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTGATTGCTTCCCTCTGTC	182
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAATGAACAAGCAAAGACACA	
contig_367859	97	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGAAACACTTTAGGCAACTTATT	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGATGCAGGTGTTGAAATGG	
contig_369024	1194	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAACTTTGGTTGAGTGTGC	164
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAATCTGAAACATCCCAAT	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_369587	1273	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAATAGGGAGTGCAATATGGTT	179
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGGAAACAACTCAAGTCCA	
contig_383472	225	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGCCACTGCCTGTTATAGT	196
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCAGACAAACGCCTGATTA	
contig_390428	864	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCTTTCACTGCCTTTGATT	192
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAATCGCTAGGTGCCTGCT	
contig_392395	284	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGTAAAGTGGTGATAAAGACTGTG	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCACCAGAGGAAAGGAAAT	
contig_407775	155	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCCCTTCTTTGCAGTTATG	159
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGGTGTTGGTTATGGTTG	
contig_408529	378	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTCACCAAGGAGCAGTGTTG	184
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAATTGAGCCTGCTTGCGATA	
contig_411127	3955	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCAGTGTAGCTGGTAGGTTG	123
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACTATGCAGGGACAGGTTTC	
contig_429827	739	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGGGTAAAGGCTGCTAACAA	191
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTAAAGGTGGCATGGGTCTT	
contig_433349	1601	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCATGCAGAGAGAAGGTTCA	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTTATTGGCGGGCGACCT	
contig_437923	191	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAAATACCGTGGGACAACA	198
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGCACCTCAGCACCAATAG	
contig_456933	540	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGATGATGGATGAGGCAGA	182
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGAGAACTGGAGAGGGACAGC	
contig_463504	2834	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCACAGGATTCTCCCATCA	200
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTGCCTGGTTTGTGGTGT	
contig_471552	1074	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGCAAGAGCAGGACTCATT	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGTGCCTTCAGCAGTGT	
contig_471638	1688	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGTCTGAGTGCATGTGTGTG	145
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCATCATCACAAGCTGTTC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_479703	1283	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCAGAGCTTGGGAAGATGA	195
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCCTTCTCGTTCGCTACTG	
contig_487120	1790	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAAGCTCCATCAGGTGTTCC	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACGCATGGCTTGTTAAATTG	
contig_502728	1169	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTCAAACGAAAGTCAAACGA	200
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACACCAGAGTACCAAATCAGG	
contig_509349	450	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAGGGTCAATTAGTCAAGGTG	172
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAACCTGTCTTGTGGCATTT	
contig_512906	1610	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCTGCTGACTATGGGTTCC	190
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGCATTCAAATCTCTCCA	
contig_520290	1100	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGAAGACAGGGTGCAAAGA	193
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACAGAAACCTGGAGCAGAGG	
contig_520553	4280	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGACAGCCTTGAACGAGAG	200
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACTCAACACGGCACACAAG	
contig_531481	538	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCATCCCACGCAAACTAAC	190
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTGCTGAAGCCAGTTTGTT	
contig_536471	444	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAAGTCGCGTGGCATA	197
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGAAAGCCGACATATTCCAA	
contig_538841	676	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGTTGTTTCAGATTAGTGCAT	156
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTGGATCTGTCCCAGCTT	
contig_559002	252	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTGATTCGTGCATTT	197
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCATGAAGTACAAAGAAACTGATTG	
contig_591595	343	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACGATGAGGAAGATGTGG	196
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAAAGGAGGACCCGAAGG	
contig_592686	221	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTCGTGCCCTATTCAGTTT	197
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGGACAACAACACTGCAA	
contig_593190	224	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATTCAACACCTGGGTCCAC	200
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAATCCATACATTTCCATTCAGT	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_596653	817	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCTACACCGACACACAGAA	159
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCTGCTTCTCTCTGGTAA	
contig_597294	908	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGTGCCATCGTGAACTTT	197
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTCAGCTCTCCGTGTCTCC	
contig_618918	1670	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCAGGGTGAAATCCAAACA	196
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAAGGTTTGCGAAGGGAAT	
contig_625315	831	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCACCCTCGCCTTAATCA	134
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCACTGATTAAACGGGAGT	
contig_629792	1174	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCGTCCGCGTATCTATCAC	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGTTAGGGTTCCTGGCTA	
contig_630189	273	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACGTTGTGGAATATGAG	97
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGTTTGAGATAAATACTGCACA	
contig_664819	1473	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGAACAAAGCGTTTCTGAC	195
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCTTTCTGGATGCTGTGTT	
contig_672725	213	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGAGAAGCGACAGCCAAA	183
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCACCTCCGACATCAGACC	
contig_675727	486	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGACCAGGATTGCCTACC	134
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGGGAAGTATGGTCAGTCA	
contig_682427	799	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGGTTGCATTGTTATTTGTG	171
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGAACGAACCTCCTGTTTGA	
contig_713657	688	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGAGACATGAAAGCTGCAAA	138
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAATTGCGTTGTTGGAATCA	
contig_716352	70	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTCTACACAGGCCAGAGC	199
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGCTGGAACTGGGCTAACT	
contig_718337	1120	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGGCCGATTATGAGTGTAAA	192
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCATCCACACTGCTTCTCCA	
contig_778708	635	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGCCTGACTTTGTGAAGTTT	168
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGCCATGTGGGTGTATTC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_801222	119	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTAATTGACGTGCAGCCTTCT	142
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAACACCTCCAGACAGACGTT	
contig_805784	333	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACAGAGCATAGCATGACCTT	170
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTGCTTTGCTGTTGCAGTT	
contig_843891	798	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGAGTCATCTCTGATCTGCTG	194
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGCAGACAGTGAAATGAGG	
contig_845093	112	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATTCAGGCCCGTCTGTGTC	160
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTTAAGCAGTGTAGGGCTGA	
contig_847347	356	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACACAAAGCCGCCACA	196
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGAGCAGGGCATTGTTCC	
contig_847878	183	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACTGCACACAGACCCTCGT	175
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACCCAGAGTCCCACATGGT	
contig_850177	68	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATTTAGCATTCCCAAGGTC	108
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAATAATAATAACAACAGCAATAACAA	
contig_852468	214	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGCCACCAGAGGAACC	180
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACAGTATTTGTTCTGCATTTCTG	
contig_863648	1069	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAACTCCGTGGTGTGTGGA	156
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGAAGCTGCAATAAGTGG	
contig_871803	747	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAACCTCACAAAGCATCCAAG	169
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACCATGCAGTCATTAAGCA	
contig_895412	689	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGAGAAACTCGTCAAAGTCG	163
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGAGATTCAGAGAGCAGTGATG	
contig_902317	288	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAACTGCAACATGAACAACT	200
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGCAATCCATATTGTACCCAAG	
contig_913003	698	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCTTTGTTTAGCCTGTTGA	164
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGAGAAGTGAGGGCGTCGTA	
contig_930173	363	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGACTCGGTATGGCAAAT	160
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAATGCTCTGTTTCCATTAGGC	

Contig	SNP position	Primers	Amplicon size (without adapters)
contig_932915	559	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCAGCACACAAGACATTATAGAAG	164
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGACGGAGAACTGCATTGA	
contig_947914	972	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTGGAGGTGAATGGAGATG	150
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAACAGGACCACCATCTGTAGG	
contig_959459	651	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCATGCCAATCGAGCTAATAA	197
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCATATCCAGTGTCCATGC	
contig_999448	97	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGTCAGCACTCGGTTTATG	191
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCAGACCACGAAGTTGACAC	
contig_1001037	77	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTGAGCCCTCGCCTGTC	140
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTCCTCTTGCTGGCTGAT	
contig_1018361	307	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGACGAGAAGTGAAGGTACAG	150
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTACCATAGGAGACATTCTGTATT	
contig_1032892	120	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAACTGCATAGTACCAAACTTGAAA	156
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTGATTGACCGCATCCTCT	
contig_1049854	640	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCGAAGACAGCTACTGGAC	187
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGAGCTGGAAAGGAGAGAG	
contig_1056163	186	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGCTTCTTCTCACGAGGT	171
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTTTACCAGATGCGCCACTT	
contig_1112399	367	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTGACGTGCTTGCTGTTC	147
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTCTAACCAGAAGCCCACA	
contig_1122966	470	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCACCACCCTGCGTCA	187
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGCCTGATGCTCTCTTTGAC	
contig_1143132	625	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCATCCTTTGATTGCTTCC	159
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAAAGGTTTAGTGGAACACAATGA	
contig_1215906	151	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACGGCAGCTACTTCCAAAC	198
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCGACTTGCGTGAAATAACC	
contig_1299478	95	FWD: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATGGAAAGGGTCATCTCAGC	171
		REV: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCGAGCGAGAGAGAGAGAGA	

Table 4. Primer sequences of 12 microsatellite loci currently used by the Leetown Science Center King Conservation Genetics Laboratory (LSC-KCGL) for genotyping of Atlantic sturgeon.

Locus name	Forward primer	Reverse primer	Allelic size range (min, median, max)
*Aox*12	CAGAGTCTTCCTGCAGCACTT	GTGCAAAAAGAGAGTGTGTGTGTC	163, 187, 209
*Aox*23	GTTAGCTTAACCATGAATTGTG	CAGTGTGCTAGCTTTCTCAATA	91, 100, 133
*Aox*45	TTGTCCAATAGTTTCCAACGC	TGTGCTCCTGCTTTTACTGTC	106, 127, 166
LS-19	CATCTTAGCCGTCTGTGGTAC	CAGGTCCCTAATACAATGGC	125, 146, 164
LS-39	TTCTGAAGTTCACACATTG	ATGGAGCATTATTGGAAGG	113, 125, 131
LS-54	CTCTAGTCTTTGTTGATTACAG	CAAAGGACTTGAAACTAGG	165, 185, 189
LS-68	TTATTGCATGGTGTAGCTAAAC	AGCCCAACACAGACAATATC	115, 151, 171
*Aox*D44	ACCGAGTTTCAAATCAAATAGC	TGAAACTGCTGTGCAATAAGAG	124, 156, 198
*Aox*D165	TTTGACAGCTCCTAAGTGATACC	AAAGCCCTACAACAAATGTCAC	186, 202, 234
*Aox*D170	GAACCATTTTATTGACATTCGAC	CCCTGTCTCACGTACATTTTATG	132, 148, 168
*Aox*D188	TGAAGTCATTGGTGATGTGTATG	ATGGAAATGTTTTATGGTAATGTG	261, 309, 353
*Aox*D241	TGTTCACAATATAGTCTTCCAGGTC	CACAACAAATCAAAACAGAAGC	178, 218, 298

Notes: The allelic size range includes the minimum, median, and maximum allele sizes observed among *n* = 1,658 Atlantic sturgeon genotyped from 12 populations.

Table 5. List of 49 single nucleotide polymorphisms (SNPs) identified from shotgun genomic sequences mapped to the complete mitochondrial DNA genome of the Atlantic sturgeon (GenBank Accession KP997217) by TL King.

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
12S rRNA	300	SNV	1	C	T	Hetero	83	312	0.47	34.7
ND1	3085	SNV	1	A	G	Hetero	863	2,685	0.47	35.0

Sturg_nd1F- CCCTGTCCAGTCTAGCTGTATACTC
Sturg_nd1R-GGGCTAGCAGTCAAATTGC
Amplifies bp 3212–3421

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
ND1	3250	SNV	1	G	A	Hetero	102	322	0.49	35.6
ND1	3252	SNV	1	C	T	Hetero	98	315	0.42	35.4
ND1	3279	SNV	1	T	C	Hetero	115	320	0.46	35.4

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
ND1	3291	SNV	1	C	T	Hetero	108	312	0.47	35.5
ND1	3312	SNV	1	T	C	Hetero	107	298	0.48	35.2
ND1	3318	SNV	1	C	T	Hetero	101	286	0.43	35.0
ND1	3354	SNV	1	T	C	Hetero	71	186	0.38	35.5
ND1	3364	SNV	1	G	T	Hetero	66	163	0.42	35.2
ND1	3369	SNV	1	C	T	Hetero	74	174	0.42	35.3
ND1	3372	SNV	1	C	T	Hetero	65	164	0.42	34.6
ND1	3381	SNV	1	T	C	Hetero	62	248	0.37	35.4
ND1	3393	SNV	1	T	C	Hetero	224	303	0.40	35.3

Sturg_tRNAF CAATATAGGAATCGTGCCTGAA
Sturg_tRNAR TGGTGTAGTGGAAGCACCAA
Amplifies bp 3852–3992

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
tRNA-Ile	3876	SNV	1	C	T	Homo	1,056	1,280	0.27	35.2
tRNA-Gln	3945	SNV	1	T	C	Homo	2,513	2,517	0.34	35.1
ND2	5070	SNV	1	A	G	Hetero	1,056	3,771	0.48	35.3

Sturg_CO1F CCGAAATTTAAACACCACCTTC
Sturg_CO1R GCTCATACTATTCCCATGTAGCC
Amplifies bp 6137–6328

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
COI	6186	SNV	1	A	G	Hetero	64	181	0.48	35.4
COI	6233	SNV	1	C	T	Hetero	27	93	0.22	36.0
COI	6236	SNV	1	C	T	Hetero	25	79	0.28	35.5
COI	6239	SNV	1	G	A	Hetero	27	75	0.30	35.4
COI	6248	SNV	1	G	A	Hetero	27	75	0.33	35.7

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
COI	6275	SNV	1	G	A	Hetero	25	97	0.36	35.5
COI	6674	SNV	1	G	A	Hetero	1,264	3,289	0.47	35.0
ATP6	8567	SNV	1	G	A	Hetero	43	118	0.21	35.0

Sturg_atp6F TAGCGTTAGGCGTTCGACTT
Sturg_atp6R GCTTGAATTATTGCTACGGCTA
Amplifies bp 8611–8782

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
ATP6	8708	SNV	1	G	A	Hetero	40	156	0.40	35.6
ATP6	8714	SNV	1	A	G	Hetero	45	168	0.42	35.0
ATP6	8717	SNV	1	C	T	Hetero	50	189	0.48	35.3
COIII	9266	SNV	1	T	C	Hetero	104	381	0.22	35.0
ND3	9698	SNV	1	G	A	Hetero	1,094	3,863	0.38	35.3

Sturg_nd4LF CCTCCTCTCTGCCCTTCTCT
Sturg_nd4LR GTGGTCTGTGCCGTGTGTT
Amplifies bp 10182–10574

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
ND4L	10221	SNV	1	G	A	Hetero	954	1,767	0.50	35.2
ND4L	10322	SNV	1	G	A	Hetero	77	222	0.42	35.3
ND4L	10326	SNV	1	G	A	Hetero	71	178	0.39	35.2
ND4L	10329	SNV	1	T	C	Hetero	73	170	0.49	35.3
ND4L	10332	SNV	1	A	G	Hetero	79	207	0.47	35.5

Sturg_nd5F GCCATCATTGAAGCCCTAAA
Sturg_nd5R TCAGGCAAGTCGTTTGATTG
Amplifies bp 13186–13383

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
ND5	13233	SNV	1	G	A	Hetero	69	176	0.39	35.1
ND5	13236	SNV	1	C	T	Hetero	61	174	0.46	35.9
ND5	13246	SNV	1	A	G	Hetero	62	149	0.39	35.3
ND5	13248	SNV	1	C	T	Hetero	52	147	0.42	35.3
ND5	13269	SNV	1	G	A	Hetero	56	142	0.36	34.7
ND5	13279	SNV	1	C	T	Hetero	44	125	0.36	35.4
ND5	13299	SNV	1	A	T	Hetero	36	99	0.39	35.6
CR	15708	SNV	1	G	A	Hetero	1,032	1,892	0.49	35.0
CR	15750	SNV	1	A	G	Homo	2,262	2,591	0.43	35.4
CR	16029	SNV	1	G	A	Hetero	709	2,065	0.39	34.9

Sturg_CRF ATCTGTGTCCGGCATTTGAT
Sturg_CRR TGCCATTCACTGTTTGTCCT
Amplifies bp 16105–16284

Region	Reference Position	Type	Length	Reference	Allele	Zygosity	Count	Coverage	Forward/Reverse Balance	Average Quality
CR	16126	SNV	1	A	G	Hetero	141	321	0.36	34.7
CR	16151	SNV	1	G	A	Homo	865	968	0.44	35.0
CR	16206	SNV	1	A	G	Hetero	786	1,586	0.47	35.2
CR	16236	SNV	1	T	C	Hetero	1,561	2,140	0.41	35.3

Notes: Primers designed in this study to amplify certain mtDNA SNPs are inserted into the table around the SNPs they amplify, which are shaded grey. Non-shaded SNPs are not amplified.

Table 5. PCR primers used to incorporate unique 8 bp indices and Illumina flow cell adapters into Atlantic sturgeon amplicons.

Name	Sequence (5`-3`)
D501-F	AATGATACGGCGACCACCGAGATCTACACTATAGCCTCGACAGGTTCAGAG*T
D502-F	AATGATACGGCGACCACCGAGATCTACACATAGAGGCCGACAGGTTCAGAG*T
D503-F	AATGATACGGCGACCACCGAGATCTACACCCTATCCTCGACAGGTTCAGAG*T
D504-F	AATGATACGGCGACCACCGAGATCTACACGGCTCTGACGACAGGTTCAGAG*T
D505-F	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGCGACAGGTTCAGAG*T
D506-F	AATGATACGGCGACCACCGAGATCTACACTAATCTTACGACAGGTTCAGAG*T
D507-F	AATGATACGGCGACCACCGAGATCTACACCAGGACGTCGACAGGTTCAGAG*T
D508-F	AATGATACGGCGACCACCGAGATCTACACGTACTGACCGACAGGTTCAGAG*T
D701-R	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCA*G
D702-R	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCA*G
D703-R	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCA*G
D704-R	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCA*G
D705-R	CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCA*G
D706-R	CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCA*G
D707-R	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCA*G
D708-R	CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCA*G
D709-R	CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCA*G
D710-R	CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCA*G
D711-R	CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCA*G
D712-R	CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCA*G

Notes: All amplicons were previously PCR amplified with primers containing the adapter CGACAGGTTCAGAGTTCTACAGTCCGACGATC on the forward primer and GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT on the reverse primer. Unique indices are denoted in green, 'F' and 'R' in the primer name denotes forward and reverse, and a '*' indicates a phosphorothioate bond.

3.1.2 Optimization Run

An initial trial sequence run of one sturgeon individual and the entire multiplex panel was performed to gauge amplification success and evenness of representation among loci. Aliquots of all 172 loci primer pairs (153 nuclear SNPs + 7 mtDNA SNPs + 12 microsatellite loci) were combined into a stock solution of 5 µM each primer. A 10 µl first step PCR reaction was set up following the protocol of Aykanat et al. (2016) consisting of 1X Qiagen multiplex master mix, equal amounts of forward and reverse primers at a final concentration of 0.054 µM, and 50 ng of genomic DNA. Thermal cycling conditions were as follows: initial denaturation at 95 ºC for 15 min; seven cycles of 95-58-72 for 30-60-45 s, respectively; and 15 cycles of 95-62-72 ºC for 30-60-45 s, respectively. The multiplex PCR was cleaned with an equal volume of AmPure beads (Agencourt) and washed twice with 200 µl of 80% ethanol. The cleaned PCR products were eluted from the beads with 12.5 µl 10 mM Tris-HCl ph 8.5, and 10 µl was used for the subsequent PCR. For the second step PCR to incorporate the dual indices and Illumina sequencing flow cell adapters, a 10.8-µl reaction was set up with 1 µl of 1:10 diluted PCR product from the first PCR, forward and reverse indices at a final concentration of 0.3 µM, and 1X Kapa Hifi PCR ready-mix. Thermal cycling conditions were as follows: initial denaturation at 95 ºC for 4 min; 15 cycles of 98-60-72 ºC for 20-15-20 s, respectively; followed by a final extension step at 72 ºC for 3 minutes. The library was subjected to a final clean up with an equal volume of AmPure XP beads, washed twice with 80% ethanol, and eluted with 32.5ul 10mM Tris-HCl pH 8.5. The cleaned PCR product was quantified on a Qubit fluorometer using the high sensitivity DS DNA assay and diluted to a concentration of 6 nM for MiSeq sequencing. The sample was 300 bp paired-end sequenced alongside several other unrelated libraries on a 600 cycle V2 cartridge.

3.1.3 Production Runs

After the initial optimization run to assess amplification of loci within the multiplex PCR, we identified 288 Atlantic sturgeon for construction of an SNP genotype baseline, with the plan that 96 individuals would be sequenced among three separate MiSeq runs (**Table 7**). Samples were chosen such that nine populations of $n = 32$ each were represented. All samples were collected from river-resident juveniles (< 500 mm TL) or adults (≥ 1,500 mm TL) captured in their natal river by permitted researchers from other agencies. DNA of each set of 96 samples was aliquoted into a 96-well plate at a standardized concentration of 20 ng/µl. The first and second PCRs followed the protocols for the optimization run, with the exception that some primer concentrations were adjusted between runs to try and influence over- or under-representation of certain loci (see results). Each run utilized the same set of dual indexed primers, allowing all 96 individuals within a single run to be demultiplexed.

To have even sequence representation of each individual within a MiSeq run, the amount of PCR product loaded on the flow cell needs to be standardized among individuals. To streamline this quantification and reduce the hands-on time, the indexed PCR product was processed through a 96 well SequalPrep normalization plate following the manufacturer's instructions. This step resulted in all samples being standardized to a similar concentration and volume. The final dilution of the pooled libraries was to a concentration of 6 nM. Each set of 96 samples was paired-end sequenced (250 bp x 2) on an Illumina MiSeq utilizing a 500 cycle V2 kit (the 600 cycle sequencing kit was unnecessary for these amplicons, most of which had sizes of approximately 200 bp or less).

Table 6. Individual Atlantic sturgeon samples and their corresponding source river run to develop a genomic baseline for Atlantic sturgeon.

Sample ID	Source River
AOK-051	Kennebec
AOK-052	Kennebec
AOK-053	Kennebec
AOK-054	Kennebec
AOK-055	Kennebec
AOK-056	Kennebec
AOK-057	Kennebec
AOK-058	Kennebec
AOK-059	Kennebec
AOK-061	Kennebec
AOK-062	Kennebec
AOK-063	Kennebec
AOK-065	Kennebec
AOK-066	Kennebec
AOK-067	Kennebec
AOK-068	Kennebec
AOK-069	Kennebec
AOK-070	Kennebec
AOK-071	Kennebec
AOK-072	Kennebec
AOK-073	Kennebec
AOK-074	Kennebec
AOK-075	Kennebec
AOK-076	Kennebec
AOK-077	Kennebec
AOK-078	Kennebec
AOK-079	Kennebec
AOK-080	Kennebec
AOK-081	Kennebec
AOK-082	Kennebec
AOK-083	Kennebec
AOK-084	Kennebec
Aoxy-03951	Delaware
Aoxy-03953	Delaware
Aoxy-03955	Delaware
Aoxy-03957	Delaware
Aoxy-03962	Delaware
Aoxy-03963	Delaware
Aoxy-03983	Delaware
Aoxy-03994	Delaware
Aoxy-04831	Delaware
Aoxy-04832	Delaware
Aoxy-04834	Delaware
Aoxy-04835	Delaware
Aoxy-04836	Delaware
Aoxy-09638	James
Aoxy-09639	James
Aoxy-09640	James
Aoxy-09641	James
Aoxy-09642	James
Aoxy-09643	James
Aoxy-09644	James
Aoxy-09645	James
Aoxy-09666	James
Aoxy-09726	James
Aoxy-09727	James
Aoxy-09728	James
Aoxy-09729	James
Aoxy-09730	James
Aoxy-09731	James
Aoxy-09914	James
Aoxy-09915	James
Aoxy-09916	James
Aoxy-09917	James
Aoxy-09918	James
Aoxy-09919	James
Aoxy-09921	James
Aoxy-09927	James
Aoxy-09928	James
Aoxy-09931	James
Aoxy-09935	James
Aoxy-09939	James
Aoxy-09945	James
Aoxy-09946	James

Sample ID	Source River
Aoxy-09947	James
Aoxy-09948	James
Aoxy02915	Edisto
Aoxy02916	Edisto
Aoxy02917	Edisto
Aoxy02923	Edisto
Aoxy02924	Edisto
Aoxy02927	Edisto
Aoxy02932	Edisto
Aoxy02933	Edisto
Aoxy02934	Edisto
Aoxy02940	Edisto
Aoxy02944	Edisto
Aoxy02947	Edisto
Aoxy02949	Edisto
Aoxy02950	Edisto
Aoxy02960	Edisto
Aoxy02961	Edisto
Aoxy02964	Edisto
Aoxy02968	Edisto
Aoxy02969	Edisto
Aoxy02973	Edisto
Aoxy02974	Edisto
Aoxy02975	Edisto
Aoxy02976	Edisto
Aoxy02979	Edisto
Aoxy02982	Edisto
Aoxy02984	Edisto
Aoxy02986	Edisto
Aoxy02989	Edisto
Aoxy02990	Edisto
Aoxy03003	Edisto
Aoxy03004	Edisto
Aoxy03005	Edisto
Aoxy06575	Altamaha
Aoxy06576	Altamaha
Aoxy06577	Altamaha
Aoxy06578	Altamaha
Aoxy06579	Altamaha
Aoxy06580	Altamaha

Sample ID	Source River
Aoxy06581	Altamaha
Aoxy06582	Altamaha
Aoxy06583	Altamaha
Aoxy06584	Altamaha
Aoxy06585	Altamaha
Aoxy06586	Altamaha
Aoxy06587	Altamaha
Aoxy06588	Altamaha
Aoxy06589	Altamaha
Aoxy06590	Altamaha
Aoxy06591	Altamaha
Aoxy06592	Altamaha
Aoxy06593	Altamaha
Aoxy06594	Altamaha
Aoxy06595	Altamaha
Aoxy06596	Altamaha
Aoxy06597	Altamaha
Aoxy06598	Altamaha
Aoxy06599	Altamaha
Aoxy06600	Altamaha
Aoxy06609	Altamaha
Aoxy06610	Altamaha
Aoxy06611	Altamaha
Aoxy06612	Altamaha
Aoxy06613	Altamaha
Aoxy06614	Altamaha
Aoxy09163	Hudson
Aoxy09164	Hudson
Aoxy09165	Hudson
Aoxy09166	Hudson
Aoxy09167	Hudson
Aoxy09168	Hudson
Aoxy09169	Hudson
Aoxy09170	Hudson
Aoxy09177	Hudson
Aoxy09179	Hudson
Aoxy09180	Hudson
Aoxy09181	Hudson
Aoxy09182	Hudson
Aoxy09183	Hudson

Sample ID	Source River
Aoxy09184	Hudson
Aoxy09185	Hudson
Aoxy09186	Hudson
Aoxy09187	Hudson
Aoxy09189	Hudson
Aoxy09192	Hudson
Aoxy09195	Hudson
Aoxy09196	Hudson
Aoxy09197	Hudson
Aoxy09198	Hudson
Aoxy09200	Hudson
Aoxy09201	Hudson
Aoxy09202	Hudson
Aoxy09203	Hudson
Aoxy09204	Hudson
Aoxy09206	Hudson
Aoxy09299	Hudson
Aoxy09632	James
Aoxy09716	Hudson
AoxyDe08-04881	Delaware
AoxyDe08-04902	Delaware
AoxyDe09-05066	Delaware
AoxyDe09-05067	Delaware
AoxyDe09-05079	Delaware
AoxyDe09-05081	Delaware
AoxyDe09-05082	Delaware
AoxyDe09-05088	Delaware
AoxyDe09-05089	Delaware
AoxyDe09-05095	Delaware
AoxyDe09-05096	Delaware
AoxyDe09-05097	Delaware
AoxyDe09-05098	Delaware
AoxyDe09-05099	Delaware
AoxyDe09-05100	Delaware
AoxyDe09-05101	Delaware
AoxyDe09-05105	Delaware
AoxyDe09-05106	Delaware
AoxyDe09-05107	Delaware
AoxyOOG14-09969	Ogeechee
AoxyOOG14-09970	Ogeechee

Sample ID	Source River
AoxyOOG14-09971	Ogeechee
AoxyOOG14-09972	Ogeechee
AoxyOOG14-09973	Ogeechee
AoxyOOG14-09975	Ogeechee
AoxyOOG14-09976	Ogeechee
AoxyOOG14-09977	Ogeechee
AoxyOOG14-09978	Ogeechee
AoxyOOG14-09979	Ogeechee
AoxyOOG14-09980	Ogeechee
AoxyOOG14-09981	Ogeechee
AoxyOOG14-09983	Ogeechee
AoxyOOG14-09984	Ogeechee
AoxyOOG14-09985	Ogeechee
AoxyOOG14-09986	Ogeechee
AoxyOOG15-10004	Ogeechee
AoxyOOG15-10006	Ogeechee
AoxyOOG15-10007	Ogeechee
AoxyOOG15-10008	Ogeechee
AoxyOOG15-10010	Ogeechee
AoxyOOG15-10011	Ogeechee
AoxyOOG15-10013	Ogeechee
AoxyOOG15-10014	Ogeechee
AoxyOOG15-10016	Ogeechee
AoxyOOG15-10017	Ogeechee
AoxyOOG15-10018	Ogeechee
AoxyOOG15-10019	Ogeechee
AoxyOOG15-10020	Ogeechee
AoxyOOG15-10021	Ogeechee
AoxyOOG15-10022	Ogeechee
AoxyOOG15-10023	Ogeechee
AoxySav08-05386	Savannah
AoxySav08-05387	Savannah
AoxySav08-05390	Savannah
AoxySav08-05391	Savannah
AoxySav08-05392	Savannah
AoxySav08-05403	Savannah
AoxySav08-05426	Savannah
AoxySav08-05427	Savannah
AoxySav08-05440	Savannah
AoxySav08-05461	Savannah

Sample ID	Source River
AoxySav13-11210	Savannah
AoxySav13-11214	Savannah
AoxySav13-11227	Savannah
AoxySav13-11246	Savannah
AoxySav13-11248	Savannah
AoxySav13-11249	Savannah
AoxySav13-11251	Savannah
AoxySav13-11258	Savannah
AoxySav13-11263	Savannah
AoxySav13-11275	Savannah
AoxySav13-11278	Savannah
AoxySav13-11279	Savannah
AoxySav13-11280	Savannah
AoxySav13-11286	Savannah
AoxySav13-11296	Savannah
AoxySav13-11298	Savannah
AoxySav13-11299	Savannah
AoxySav13-11301	Savannah
AoxySav13-11320	Savannah
AoxySav13-11347	Savannah
AoxySav13-11349	Savannah
AoxySav13-11361	Savannah
AOY-13-03	York
AOY-13-04	York
AOY-13-05	York
AOY-13-06	York
AOY-13-07	York
AOY-13-08	York
AOY-13-09	York
AOY-13-10	York
AOY-14-19	York
AOY-14-20	York
AOY-14-21	York
AOY-14-22	York
AOY-14-44	York
AOY-14-45	York
AOY-14-46	York
AOY-14-47	York
AOY-15-57	York
AOY-15-58	York

Sample ID	Source River
AOY-15-59	York
AOY-15-60	York
AOY-15-61	York
AOY-15-62	York
AOY-15-63	York
AOY-15-64	York
AOY-17-01	York
AOY-17-02	York
AOY-17-03	York
AOY-17-04	York
AOY-17-05	York
AOY-17-06	York
AOY-17-07	York
AOY-17-08	York

3.1.4 Bioinformatic Analyses

Each set of sequences obtained for an individual represent multiple SNP loci. In order to identify the genotypes of each SNP, these sequences need to be mapped against the reference dataset of contigs containing the SNPs (i.e., the 153 nuclear SNP loci reads are mapped against the corresponding 153 contigs that contain them from the *de novo* reference assembly). Then, the genotype at each locus for each individual can be determined by looking at the nucleotide distribution at the pre-defined SNP location within the contig. Due to inefficient amplification of the microsatellite loci in production runs (see results), we only concentrate here on the bioinformatic steps used to obtain genotypes of the SNP loci.

To evaluate the optimization run, the demultiplexed sequences were trimmed with a quality threshold of Q10 using bbduk.sh, overlapping read pairs were merged using bbmerge.sh (default parameters) of the bbmap package (B. Bushnell, https://sourceforge.net/projects/bbmap). To assess amplification success, a custom perl script was used to search for and enumerate the number of sequences matching each expected amplicon. The script required the forward and reverse primers to be exact matches to be counted. These counts were input into a spreadsheet and visualized to see which loci amplified and in what proportion. Loci with exceedingly high or low coverage were identified, so that their concentration in subsequent production runs could be changed to result in more even amplification.

For production runs, bioinformatic processing was adjusted to improve speed and retain fastq quality information, as map quality was no longer relevant (due to the much reduced search space, and high distinctiveness of the final targets), but base quality scores were more relevant for genotyping individuals at the low end of coverage. Processing of each production run followed the same basic pipeline. Sequence reads of each sample were demultiplexed based on the assigned dual index and placed into individual folders. All reads were screened for the presence of residual Illumina Truseq adapters on both ends using bbduk.sh (option flags included k = 29, mink = 11, hdist = 1, tpe, tbo). Read pairs were then merged using the default settings of bbmerge.sh, with no quality trimming performed. A fasta formatted file of the 153 contigs was indexed for the mapping program bbmap, using the bbmap.sh script (option flags included build = 1, k = 13). In addition, a list of targeted SNP positions to genotype was created in the form of a tab delimited text file, with two columns corresponding to the contig name and SNP base position within the contig respectively. Merged read pairs from each sample were then mapped against the reference file using bbmap.sh to create a *.bam file (option flags with bbmap included build = 1, minid = 0.90). The resulting *.bam file was indexed and sorted using ‘samtools sort’ and ‘samtools view’. The indexed *.bam file used ‘samtools mpileup’ to provide a summary of the coverage of mapped reads and the data in the right format for bcftools (option flags included -d 100, -C 50, -I, ‘-t DP,INFO/ADF,INFO/ADR’, -Rvf). Finally, ‘bcftools call’ was used to output the genomic variants into a vCard file (VCF file). For the mitogenomic amplicons within the GT-seq panel, SNPs were processed identically as for the nuclear SNPs, with the exception of changing the ploidy flag to ‘--ploidy 1’ for the bcftools call portion of the script. The bash shell scripts used to process these data can be provided by the authors upon request.

For each individual sturgeon, we generated a VCF file that contains the SNP genotypes. All VCF files were placed in one directory, compressed with bgzip, and indexed using bcftols index. All indexed VCF files were then merged into one large combined VCF file of all samples and variants using bcftools merge. This final VCF file, and though it contains the number of samples and genotypes of variants, is not initially filtered to require any set depth of coverage, number of genotyped loci per population, etc. Filtering of the VCF file was accomplished with vcftools (Danecek et al. 2011).

3.1.5 Assessment of the Utility of Mitogenomic SNPs for MSA

When primers were designed for the mitogenomic SNPs identified earlier by T.L. King (**Table 5**), it was assumed that the SNPs would be phylogeographically informative. However, examination of the SNPs we

were able to genotype in our production runs through GT-seq revealed that these SNPs identified by T.L. King were largely invariant (see results). Due to the sudden loss of T.L. King and the inability to fully reproduce his earlier analyses, we performed a reanalysis of the mitogenomic data from among the Atlantic sturgeon used for shotgun genomic sequencing and construction of a reference genome. Because this reanalysis was performed after testing of the GT-seq panel, none of the SNPs identified by our reanalysis had amplicons designed or sequenced. Our inclusion of these data here is to assess the potential utility of a mitogenomic variation for future mixed stock analyses and assignment testing efforts.

Quality trimmed reads from the same 71 Atlantic sturgeon used for the *de novo* reference assembly were used for mapping against an Atlantic sturgeon mitogenomic reference from GenBank with the accession number KP997217. Reference indexing, read mapping, and variant calling were performed identically to the GT-seq dataset, with the exception that the bcftools call '--ploidy' flag was set to 1.

All VCF files were placed in the same directory, indexed, and merged. The resulting merged VCF file was filtered using vcftools and the following parameters: minDP = 5, maf = 0.05, minQ = 30, max-missing 0.1, remove any individual with > 50% missing data. The filtered VCF file was imported into R as a genind object using the adegenet package (Jombart 2008, Jombart 2011), and the genind file ploidy was set to 1. A Discriminant Analysis of Principal Components plot (DAPC; Jombart et al. 2010) was constructed from this genind file in the adegenet package to visualize population structure. The number of principal components retained was 12, and the number of discriminant functions retained was 9.

3.2 Results

3.2.1 GT-seq Optimization Run

The initial GT-seq optimization run, which consisted of just one individual amplified with each of the *n* = 151 nuclear SNP, *n* = 12 microsatellites, and *n* = 7 mtDNA SNP primer pairs at 0.54 µM final concentration resulted in 610,572 merged read pairs, of which 577,097 reads (~95%) could be assigned to a locus (**Table 8**; **Figure 1**). Eight of the 151 nuclear SNP loci did not amplify, whereas the remainder had coverage ranging from 1–24,196 reads. Among the seven mtDNA SNP loci, all loci were represented, though one locus had only two reads, while the next least abundant mtDNA amplicon had 5,053 reads. Overall, the mtDNA loci had the highest average number of reads, with one locus containing 188,075 reads accounting for 33% of the total reads in the entire run. Amplification of the microsatellite loci was the least successful among the marker types, with four loci having no amplification, and four loci with one to nine reads. The remaining four loci had 28–656 reads.

3.2.2 Production Runs

We sought to target a minimum coverage level of 10X per locus for SNP calling, and approximately 100,000 reads per sample, which should be attainable with 96 individuals in one paired-end sequencing MiSeq run. It was evident from the optimization run that when the number of reads obtained was scaled down to estimate what level of coverage would be expected per locus at 100,000 total reads with the current primer cocktail, 28 out of the 143 amplifying nuclear and one of the mtDNA loci would not meet this 10X threshold. Therefore, we adjusted the concentrations of primers in the multiplex mastermix in an attempt to make coverage more even for the production runs. First, any locus with zero reads was removed from the multiplex under the assumption that they were never going to amplify in the master mix (*n* = 8 nuclear SNPs, *n* = 4 microsatellites; **Table 8**). Next, we examined the actual sequences of any locus with less than 10 reads and omitted any locus with off target reads not matching the expected amplicon (*n* = 3 nuclear SNPs, *n* = 1 microsatellites, *n* = 1 mtDNA SNP). After omission of these loci, the remaining loci were grouped into three somewhat arbitrary categories: loci with greater than 5,000 reads, loci with less than 500 reads, and loci between 501 and 4,999 reads. Loci with more than 5,000 reads

were adjusted to have a multiplex concentration of 0.025 nM, loci with less than 500 reads were increased to 0.1 nM, and loci with reads between 501 and 4,999 remained at a concentration of 0.054. Locus 'contig_618918' was accidentally omitted from the primer cocktail in the first production run.

The first production run of n = 96 Atlantic sturgeon with the adjusted primer concentrations resulted in an average of 143,822 reads per fish across all marker types (range 51,617–184,768, standard deviation 22,230). The adjustment of primer concentrations had a pronounced effect on the number of reads obtained for each locus (**Table 8**; **Figure 2**). The loci in **Figure 2** are in the same order as within **Figure 1**, and each bar in **Figure 2** represents the average number of reads at that locus across the 96 sequenced individuals. Although **Figure 1** represents actual counts from a single individual at each locus and **Figure 2** represents the average number of reads across 96 individuals at each locus, the general pattern of amplification success can still be assessed. Overall, there was a general flattening of the read distribution, indicating that adjustment of the primer concentrations did have the anticipated impact of changing locus representation. However, some nuclear loci completely ceased to be represented (contig_273784, contig_270080, contig_266918, contig_264486; **Table 8**), and this could not be explained by the lower number of reads per individual versus the optimization run (i.e., adjusting the number of reads expected per locus in the original optimization run of 577,097 total reads to the relative abundance expected at ~100,000 should not result in these loci falling below zero reads). Four of the mtDNA loci no longer amplified in the adjusted primer cocktail (sturg_CR, sturg_nd5, sturg_nd1, sturg_nd4L), and the two that still amplified were among the most highly represented loci (sturg_co1, sturg_atp6). One microsatellite locus did not amplify in the adjusted primer cocktail (*Aox*19).

The primer concentrations in second production run of the next n = 96 Atlantic sturgeon were again slightly adjusted to try and further even out representation of a few loci. Three nuclear SNP loci (contig_155051, contig_1056163, contig_863648), and five mtDNA SNP loci were adjusted (sturg_CR, sturg_nd5, sturg_nd1, sturg_atp6, sturg_nd4l; **Table 8**; **Figure 3**). These libraries were run twice due to inefficient clustering in the first library, resulting in a low number of reads. Consequently, the numbers presented in **Table 8** and **Figure 3** represent the read counts across the two runs combined. Adjustment of these three nuclear loci did not appreciably change their representation. For the adjusted mtDNA loci, four of them continued to not amplify, while one representation of one locus did decrease as intended (sturg_atp6). Sequencing of the third production run utilized the same primer concentrations as for the second run (**Table 8**; **Figure 4**). Overall, the results were very similar to the second run.

Table 7. Summary of sequencing results for each locus that was considered for inclusion within the GT-seq panel.

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_252163	Nuc	0.054	0	0	-	0	-	0	-
contig_252793	Nuc	0.054	0	0	-	0	-	0	-
contig_253460	Nuc	0.054	0	0	-	0	-	0	-
contig_254281	Nuc	0.054	0	0	-	0	-	0	-
contig_256737	Nuc	0.054	0	0	-	0	-	0	-
contig_260617	Nuc	0.054	0	0	-	0	-	0	-
contig_592686	Nuc	0.054	0	0	-	0	-	0	-
contig_593190	Nuc	0.054	0	0	-	0	-	0	-
contig_305031	Nuc	0.054	1	0	-	0	-	0	-
contig_801222	Nuc	0.054	3	0.1	11.1 (5.1)	0.1	11.8 (7.3)	0.1	6.7 (5.1)
contig_323589	Nuc	0.054	4	0.1	41.3 (16.5)	0.1	3.7 (3.2)	0.1	32.3 (21.3)
contig_902317	Nuc	0.054	4	0	-	0	-	0	-
contig_349978	Nuc	0.054	6	0.1	19.5 (10)	0.1	5.1 (4.5)	0.1	11.1 (6.6)
contig_270624	Nuc	0.054	8	0	-	0	-	0	-
contig_512906	Nuc	0.054	9	0.1	78.8 (21.6)	0.1	21.2 (9.6)	0.1	74.2 (22.8)
contig_61910	Nuc	0.054	12	0.1	28.8 (10.8)	0.1	181.1 (72.6)	0.1	22 (11.3)
contig_154779	Nuc	0.054	14	0.1	40.2 (10.4)	0.1	161.4 (51.9)	0.1	36.5 (15.7)
contig_1112399	Nuc	0.054	14	0.1	232.8 (68.3)	0.1	78.3 (34.3)	0.1	159 (91)
contig_17037	Nuc	0.054	17	0.1	13.4 (5.7)	0.1	338.4 (151.3)	0.1	10.9 (6.4)
contig_80054	Nuc	0.054	18	0.1	102.6 (40.4)	0.1	509.2 (244.1)	0.1	71.1 (58)
contig_597294	Nuc	0.054	19	0.1	71.6 (38.4)	0.1	113.9 (66.4)	0.1	49.8 (33.9)
contig_1032892	Nuc	0.054	22	0.1	2.7 (3)	0.1	6.6 (9.2)	0.1	0.5 (0.8)
contig_204708	Nuc	0.054	24	0.1	2,564.1 (806.3)	0.1	4,102.3 (1510.2)	0.1	3,119.5 (1,638.7)
contig_273784	Nuc	0.054	26	0.1	0 (0)	0.1	0 (0)	0.1	0 (0)

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_223774	Nuc	0.054	28	0.1	1,266.7 (359.8)	0.1	211.5 (91.1)	0.1	808.7 (273)
contig_75258	Nuc	0.054	29	0.1	36.8 (13.3)	0.1	936.4 (354.2)	0.1	28.3 (17.3)
contig_124134	Nuc	0.054	29	0.1	9.1 (4.2)	0.1	10.4 (6.2)	0.1	2.8 (2.5)
contig_1299478	Nuc	0.054	29	0.1	900.8 (1,383.1)	0.1	770.7 (1,026.2)	0.1	832.1 (1,436.8)
contig_48956	Nuc	0.054	38	0.1	30.6 (10.4)	0.1	182.6 (65.7)	0.1	23.6 (15)
contig_119516	Nuc	0.054	40	0.1	84.6 (20.8)	0.1	1,028.5 (437.3)	0.1	83.1 (32.6)
contig_672725	Nuc	0.054	40	0.1	226.9 (68.9)	0.1	0.1 (0.2)	0.1	1.1 (1.1)
contig_852468	Nuc	0.054	44	0.1	143.6 (36.4)	0.1	232.5 (82.2)	0.1	92.8 (28)
contig_102467	Nuc	0.054	46	0.1	160 (36.9)	0.1	1,395.5 (388.8)	0.1	102.6 (35.1)
contig_383472	Nuc	0.054	52	0.1	14.1 (6.1)	0.1	124.2 (59.3)	0.1	63.8 (36.6)
contig_176276	Nuc	0.054	55	0.1	59.1 (20)	0.1	711.3 (271)	0.1	32.9 (18.4)
contig_347490	Nuc	0.054	56	0.1	191.9 (282.6)	0.1	40.4 (45.9)	0.1	45.7 (77.3)
contig_351321	Nuc	0.054	78	0.1	791.7 (660.2)	0.1	325.6 (230.8)	0.1	738 (636.4)
contig_98816	Nuc	0.054	81	0.1	302.6 (88)	0.1	2,875.2 (996.5)	0.1	256.3 (98.5)
contig_1049854	Nuc	0.054	83	0.1	413.2 (94.7)	0.1	40.2 (16)	0.1	220 (62.9)
contig_300700	Nuc	0.054	87	0.1	491 (115.1)	0.1	97.1 (36.2)	0.1	331.6 (100.3)
contig_625315	Nuc	0.054	91	0.1	827.9 (222.1)	0.1	72.7 (36.2)	0.1	357 (149.8)
contig_215639	Nuc	0.054	102	0.1	906.1 (206)	0.1	744.8 (219.6)	0.1	457.6 (133.1)
contig_86226	Nuc	0.054	104	0.1	799.3 (196)	0.1	1,085.8 (353.1)	0.1	688.6 (224.6)
contig_805784	Nuc	0.054	112	0.1	166.4 (40.4)	0.1	392.3 (121.7)	0.1	141.4 (47.2)
contig_195684	Nuc	0.054	121	0.1	464.9 (96.5)	0.1	811.2 (244.5)	0.1	328.7 (120.2)
contig_31935	Nuc	0.054	126	0.1	179.6 (63.5)	0.1	1,078.7 (422.9)	0.1	129.6 (58.4)
contig_211497	Nuc	0.054	127	0.1	553 (110.8)	0.1	261.6 (71.2)	0.1	355.6 (96.8)
contig_999448	Nuc	0.054	129	0.1	1,169.4 (264.6)	0.1	387 (114.8)	0.1	689.7 (213.6)
contig_895412	Nuc	0.054	142	0.1	83.6 (21.2)	0.1	388.7 (134)	0.1	72.6 (26.3)
contig_88389	Nuc	0.054	144	0.1	3,837.8 (773.3)	0.1	5,165.1 (1,345.3)	0.1	3,759.2 (1,020.4)

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_675727	Nuc	0.054	152	0.1	424.9 (151.1)	0.1	159.9 (53.5)	0.1	198.3 (76.9)
contig_44181	Nuc	0.054	153	0.1	164.5 (51.5)	0.1	3,875.3 (1,216.9)	0.1	139.5 (45)
contig_407775	Nuc	0.054	157	0.1	881.9 (499.5)	0.1	41.5 (31.2)	0.1	215.6 (128.5)
contig_8889	Nuc	0.054	173	0.1	257.3 (113.4)	0.1	2,776.3 (1,263.2)	0.1	236.5 (105.2)
contig_682427	Nuc	0.054	176	0.1	2,418.8 (550.1)	0.1	1,497.6 (464.1)	0.1	2,007.7 (628.5)
contig_91914	Nuc	0.054	177	0.1	651.9 (336.5)	0.1	151.2 (84.5)	0.1	310 (173.4)
contig_168922	Nuc	0.054	191	0.1	5,351.5 (2,232)	0.1	5,061.2 (1,915.7)	0.1	4,033.7 (1,192.6)
contig_279175	Nuc	0.054	201	0.1	474.4 (107.6)	0.1	473.1 (165)	0.1	383.6 (113.5)
contig_270080	Nuc	0.054	203	0.1	0 (0)	0.1	0 (0)	0.1	0 (0)
contig_664819	Nuc	0.054	212	0.1	3,160.3 (829.5)	0.1	170.8 (176.7)	0.1	456.7 (410.1)
contig_713657	Nuc	0.054	213	0.1	425.1 (134.5)	0.1	149.7 (60.4)	0.1	190.3 (97.2)
contig_767	Nuc	0.054	219	0.1	5,013.2 (1,370.2)	0.1	5,909.2 (1,922.7)	0.1	3,948 (1,210.6)
contig_128545	Nuc	0.054	222	0.1	1,840.8 (532.8)	0.1	1,958.8 (647.1)	0.1	1,501.3 (501)
contig_845093	Nuc	0.054	238	0.1	2,219.6 (520.4)	0.1	133.3 (53.4)	0.1	607.6 (240.7)
contig_630189	Nuc	0.054	243	0.1	2,124.3 (1,078.7)	0.1	340.3 (164.2)	0.1	1,482.2 (756.4)
contig_335132	Nuc	0.054	264	0.1	721.4 (144)	0.1	422.8 (167.1)	0.1	623.5 (214.1)
contig_298604	Nuc	0.054	281	0.1	897.6 (596.5)	0.1	126.9 (81)	0.1	635.9 (405)
contig_146123	Nuc	0.054	354	0.1	886 (183.8)	0.1	6,609.7 (1,936.9)	0.1	827.3 (213)
contig_718337	Nuc	0.054	355	0.1	131.3 (35.8)	0.1	41.1 (19.6)	0.1	56.6 (28.6)
contig_629792	Nuc	0.054	360	0.1	279.8 (67.3)	0.1	720.4 (284.2)	0.1	269.9 (86.4)
contig_338725	Nuc	0.054	376	0.1	596.8 (206.4)	0.1	124.5 (58.7)	0.1	267.8 (116)
contig_408529	Nuc	0.054	396	0.1	197.4 (51)	0.1	237.9 (127.2)	0.1	153.8 (70.4)
contig_145014	Nuc	0.054	406	0.1	2,503.3 (501)	0.1	5,028.4 (1,586.8)	0.1	1,528.6 (455.7)
contig_198579	Nuc	0.054	409	0.1	4,563.3 (1,006.1)	0.1	2,644.5 (1,049.5)	0.1	3,444.2 (1,103.8)
contig_242683	Nuc	0.054	423	0.1	1,981 (511.5)	0.1	303.3 (128.3)	0.1	453.7 (217.5)
contig_392395	Nuc	0.054	470	0.1	928.3 (338)	0.1	60.9 (29.5)	0.1	328.7 (185.9)

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_850177	Nuc	0.054	470	0.1	1,721.9 (711.6)	0.1	482.5 (273.6)	0.1	1,141 (515.8)
contig_341935	Nuc	0.054	481	0.1	2,017.8 (415.4)	0.1	43.9 (17.1)	0.1	378 (110.4)
contig_198706	Nuc	0.054	509	0.054	127.7 (27.1)	0.054	831.9 (267.3)	0.054	77.1 (26.1)
contig_369587	Nuc	0.054	560	0.054	235.6 (58.4)	0.054	347.5 (122.3)	0.054	134.3 (46.7)
contig_266918	Nuc	0.054	566	0.054	0 (0)	0.054	0 (0)	0.054	0 (0)
contig_1001037	Nuc	0.054	579	0.054	64.3 (18.7)	0.054	61.9 (26.8)	0.054	35.6 (16.2)
contig_93145	Nuc	0.054	618	0.054	122.6 (30.5)	0.054	198.5 (64.9)	0.054	68.4 (31.5)
contig_1122966	Nuc	0.054	675	0.054	504.5 (99.6)	0.054	43.6 (22)	0.054	87.5 (31.7)
contig_124581	Nuc	0.054	683	0.054	88 (21.6)	0.054	509.1 (152.4)	0.054	46.7 (19.5)
contig_778708	Nuc	0.054	684	0.054	336.1 (141.6)	0.054	599.6 (219.4)	0.054	231.8 (99.8)
contig_509349	Nuc	0.054	702	0.054	49.9 (12.1)	0.054	61 (25.2)	0.054	28.6 (10.4)
contig_343910	Nuc	0.054	723	0.054	42.4 (11.4)	0.054	81.4 (26.9)	0.054	22.5 (9.7)
contig_361665	Nuc	0.054	782	0.054	289.4 (91.7)	0.054	164.6 (73.8)	0.054	114 (55.4)
contig_847347	Nuc	0.054	785	0.054	8.3 (3.2)	0.054	26.6 (11.4)	0.054	5.1 (3.1)
contig_251136	Nuc	0.054	787	0.054	412 (216.6)	0.054	935.7 (462.8)	0.054	290.7 (158.2)
contig_159128	Nuc	0.054	788	0.054	129.9 (33.4)	0.054	178.7 (66.3)	0.054	57.3 (27.8)
contig_213642	Nuc	0.054	838	0.054	1,507.5 (346.6)	0.054	1,256.2 (566.5)	0.054	941.4 (349.5)
contig_913003	Nuc	0.054	858	0.054	23.3 (8.3)	0.054	33.7 (10)	0.054	11.8 (5.9)
contig_471552	Nuc	0.054	924	0.054	152.6 (31.6)	0.054	178.1 (65.1)	0.054	87.7 (26.7)
contig_429827	Nuc	0.054	1047	0.054	160.1 (41.1)	0.054	284.9 (98)	0.054	88.9 (36.6)
contig_463504	Nuc	0.054	1049	0.054	393.2 (128.8)	0.054	514.6 (216.1)	0.054	235.6 (81.5)
contig_182703	Nuc	0.054	1118	0.054	1,683.4 (371.4)	0.054	3,653.8 (1,203.3)	0.054	1,112 (338.9)
contig_591595	Nuc	0.054	1141	0.054	147.7 (36.2)	0.054	253.4 (94.8)	0.054	101.2 (40)
contig_456933	Nuc	0.054	1168	0.054	511.1 (127)	0.054	556.5 (190.6)	0.054	307.3 (113.5)
contig_2585	Nuc	0.054	1174	0.054	189.4 (43.2)	0.054	382 (148.1)	0.054	100.7 (35.9)
contig_304245	Nuc	0.054	1266	0.054	263.6 (55.4)	0.054	223.8 (80.2)	0.054	124.4 (35.1)

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_39444	Nuc	0.054	1275	0.054	111.8 (31.5)	0.054	1,985.5 (566)	0.054	90.8 (27.6)
contig_536471	Nuc	0.054	1287	0.054	1,030.8 (244.8)	0.054	1,461.4 (539.7)	0.054	636.3 (204.8)
contig_843891	Nuc	0.054	1308	0.054	199.7 (48.2)	0.054	48.9 (24.2)	0.054	42.3 (18.9)
contig_871803	Nuc	0.054	1356	0.054	58.9 (16.3)	0.054	100.6 (36.4)	0.054	33.3 (13.5)
contig_369024	Nuc	0.054	1513	0.054	107.2 (57.1)	0.054	170.1 (91.4)	0.054	86.8 (34.7)
contig_197214	Nuc	0.054	1518	0.054	577.4 (235.5)	0.054	1,292.4 (552.2)	0.054	328.8 (175.6)
contig_367859	Nuc	0.054	1539	0.054	1,704 (647.7)	0.054	2,184.7 (915.2)	0.054	796.9 (356.8)
contig_299947	Nuc	0.054	1584	0.054	1,518 (388.6)	0.054	1,413.1 (638)	0.054	1,175.9 (428.7)
contig_947914	Nuc	0.054	1664	0.054	145.1 (37.3)	0.054	183.8 (94.5)	0.054	90.9 (38.8)
contig_264486	Nuc	0.054	1737	0.054	0 (0)	0.054	0 (0)	0.054	0 (0)
contig_345578	Nuc	0.054	1799	0.054	1,127.4 (227.7)	0.054	1,167.9 (378.5)	0.054	640.7 (161.8)
contig_297854	Nuc	0.054	1816	0.054	636.8 (134.7)	0.054	1,162.6 (341.3)	0.054	392.5 (130.4)
contig_1018361	Nuc	0.054	1937	0.054	2,304.1 (601.8)	0.054	1,809.5 (704.8)	0.054	1,107 (376.5)
contig_363259	Nuc	0.054	1992	0.054	1,240.1 (312.1)	0.054	1,096.8 (439.1)	0.054	683.1 (215.9)
contig_167621	Nuc	0.054	2020	0.054	901.4 (184.8)	0.054	1,315.5 (499.4)	0.054	72 (39.7)
contig_74051	Nuc	0.054	2050	0.054	72.3 (17.5)	0.054	339.1 (117.4)	0.054	47.4 (14.9)
contig_520290	Nuc	0.054	2072	0.054	987.9 (187.4)	0.054	952 (378.8)	0.054	684.2 (191.8)
contig_210463	Nuc	0.054	2186	0.054	1,847.4 (476.8)	0.054	529.4 (223.8)	0.054	344.4 (129)
contig_502728	Nuc	0.054	2189	0.054	1,142.9 (296.7)	0.054	1,570.6 (656)	0.054	563 (240.8)
contig_411127	Nuc	0.054	2233	0.054	1,807 (548.2)	0.054	1,363.8 (685.4)	0.054	1,115 (405.7)
contig_390428	Nuc	0.054	2302	0.054	453.6 (192.8)	0.054	47.5 (23.2)	0.054	44.4 (17.7)
contig_520553	Nuc	0.054	2314	0.054	1,325.4 (256.8)	0.054	1,506.1 (559.5)	0.054	893.2 (283.7)
contig_471638	Nuc	0.054	2325	0.054	2,062.2 (421.3)	0.054	989.9 (520.5)	0.054	1,319.2 (423.9)
contig_212828	Nuc	0.054	2392	0.054	3,096 (878.6)	0.054	521.7 (194.5)	0.054	684.5 (209.3)
contig_559002	Nuc	0.054	2410	0.054	853.2 (205.4)	0.054	1,076.8 (386.8)	0.054	579.6 (169.2)
contig_932915	Nuc	0.054	2608	0.054	2,267.6 (397.4)	0.054	854.6 (329.7)	0.054	1,086.1 (377)

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
contig_433349	Nuc	0.054	2650	0.054	1,844.6 (392.9)	0.054	2,192.5 (893.5)	0.054	1,406.8 (498.6)
contig_205045	Nuc	0.054	2664	0.054	631 (153.9)	0.054	3,964.4 (1,020.5)	0.054	371.4 (122.4)
contig_6443	Nuc	0.054	2706	0.054	1,018.2 (210.2)	0.054	5,691.5 (1,555.2)	0.054	668.1 (193.2)
contig_596653	Nuc	0.054	2966	0.054	2,745.7 (596.2)	0.054	3,525.5 (1,060.8)	0.054	1,877.1 (503.6)
contig_479703	Nuc	0.054	3038	0.054	1,147 (262.3)	0.054	1,347.1 (360.9)	0.054	607.3 (174.3)
contig_487120	Nuc	0.054	3093	0.054	1,302.8 (356.4)	0.054	1,630.6 (556)	0.054	887.5 (324.1)
contig_93920	Nuc	0.054	3333	0.054	1,507.2 (323.2)	0.054	3,629.8 (1,070.1)	0.054	779.5 (236.2)
contig_538841	Nuc	0.054	3394	0.054	1,281.7 (303.2)	0.054	1,365.6 (588.9)	0.054	906 (256.2)
contig_716352	Nuc	0.054	4290	0.054	1,298.1 (333.6)	0.054	1,102.9 (466.2)	0.054	681.5 (239.1)
contig_959459	Nuc	0.054	4321	0.054	2,142.9 (663.3)	0.054	25,36.2 (1,030.4)	0.054	1,227.3 (556)
contig_24799	Nuc	0.054	4553	0.054	854.4 (191.6)	0.054	2,173.9 (779.9)	0.054	475.2 (143.1)
contig_292868	Nuc	0.054	4556	0.054	2,860.5 (592.1)	0.054	1,162.9 (676.2)	0.054	1,634.4 (623.4)
contig_1143132	Nuc	0.054	4702	0.054	3,414.6 (853)	0.054	3,394.9 (1,320.2)	0.054	2,474.3 (750.3)
contig_155051	Nuc	0.054	5047	0.025	3.7 (2.1)	0.035	3.1 (1.9)	0.035	0.6 (0.8)
contig_327571	Nuc	0.054	5198	0.025	791.7 (186)	0.025	4.5 (2.9)	0.025	2.1 (1.5)
contig_618918	Nuc	0.054	5683	0*	-	0.054	187.1 (71.6)	0.054	83.3 (26.8)
contig_437923	Nuc	0.054	6065	0.025	91.8 (25.3)	0.025	0.5 (0.7)	0.025	0.4 (0.6)
contig_309381	Nuc	0.054	6084	0.025	1,241.9 (226)	0.025	2.4 (1.9)	0.025	1.5 (1.6)
contig_847878	Nuc	0.054	6872	0.025	146.6 (38)	0.025	4.4 (2.8)	0.025	2.2 (1.7)
contig_1215906	Nuc	0.054	6925	0.025	495.5 (116)	0.025	2.5 (1.9)	0.025	1.1 (1.2)
contig_531481	Nuc	0.054	9208	0.025	781 (272.5)	0.025	4.5 (3)	0.025	2.2 (1.6)
contig_1056163	Nuc	0.054	9929	0.025	1 (0.8)	0.035	0.8 (0.9)	0.035	0.2 (0.6)
contig_353457	Nuc	0.054	15713	0.025	271.6 (82.8)	0.025	7.3 (3.4)	0.025	2.5 (1.8)
contig_930173	Nuc	0.054	17437	0.025	6.7 (3.1)	0.025	7.1 (4.4)	0.025	3.4 (2.4)
contig_863648	Nuc	0.054	24196	0.025	12.2 (6.4)	0.035	19.5 (10.5)	0.035	4.3 (4.4)
sturg_tRNA	mt	0.054	2	0	-	0	-	0	-

		Optimization Run		Production Run 1		Production Run 2		Production Run 3	
Locus	Class	Primer Conc. [µM]	Count	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)	Primer Conc. [µM]	Average Count (SD)
sturg_CR	mt	0.054	5053	0.025	0 (0)	0.035	0 (0)	0.035	0 (0)
sturg_nd5	mt	0.054	6669	0.025	0 (0)	0.035	0 (0)	0.035	0 (0)
sturg_nd1	mt	0.054	9883	0.025	0 (0)	0.035	0 (0)	0.035	0 (0)
sturg_co1	mt	0.054	31545	0.025	2,703.3 (3,743)	0.025	116.8 (63.5)	0.025	10.3 (8.8)
sturg_atp6	mt	0.054	90253	0.025	1,8475.3 (1,0631.1)	0.005	246.8 (141.1)	0.005	957.7 (494.1)
sturg_nd4L	mt	0.054	188075	0.025	0 (0.1)	0.035	0 (0)	0.035	0 (0)
AOX23	Msat	0.054	0	0	-	0	-	0	-
AOXLS39	Msat	0.054	0	0	-	0	-	0	-
AOXLS54	Msat	0.054	0	0	-	0	-	0	-
AOXD170	Msat	0.054	0	0	-	0	-	0	-
AOXLS68	Msat	0.054	1	0.1	0.2 (0.4)	0.1	0.1 (0.4)	0.1	2.4 (15.4)
AOXD188	Msat	0.054	1	0	-	0	-	0	-
AOXD44	Msat	0.054	6	0.1	380.3 (98.9)	0.1	108.4 (36.6)	0.1	256.2 (121.9)
AOX12	Msat	0.054	9	0.1	5.6 (3.4)	0.1	0 (0)	0.1	0 (0.1)
AOXD241	Msat	0.054	28	0.1	448.8 (435.6)	0.1	79.2 (75.2)	0.1	262.4 (240.5)
AOXLS19	Msat	0.054	84	0.1	0 (0)	0.1	0 (0)	0.1	0 (0)
AOXD165	Msat	0.054	570	0.054	561.2 (159.8)	0.054	45.7 (22)	0.054	75.1 (41.6)
AOX45	Msat	0.054	656	0.054	180.5 (97.8)	0.054	0.3 (0.5)	0.054	1.4 (1.5)

Notes: The marker classes are abbreviated as follows: “Nuc” = nuclear, “mt” = mitochondrial, and “Msat” = microsatellite. The optimization run was based on a single individual, whereas each of the three production runs used 96 unique individuals.

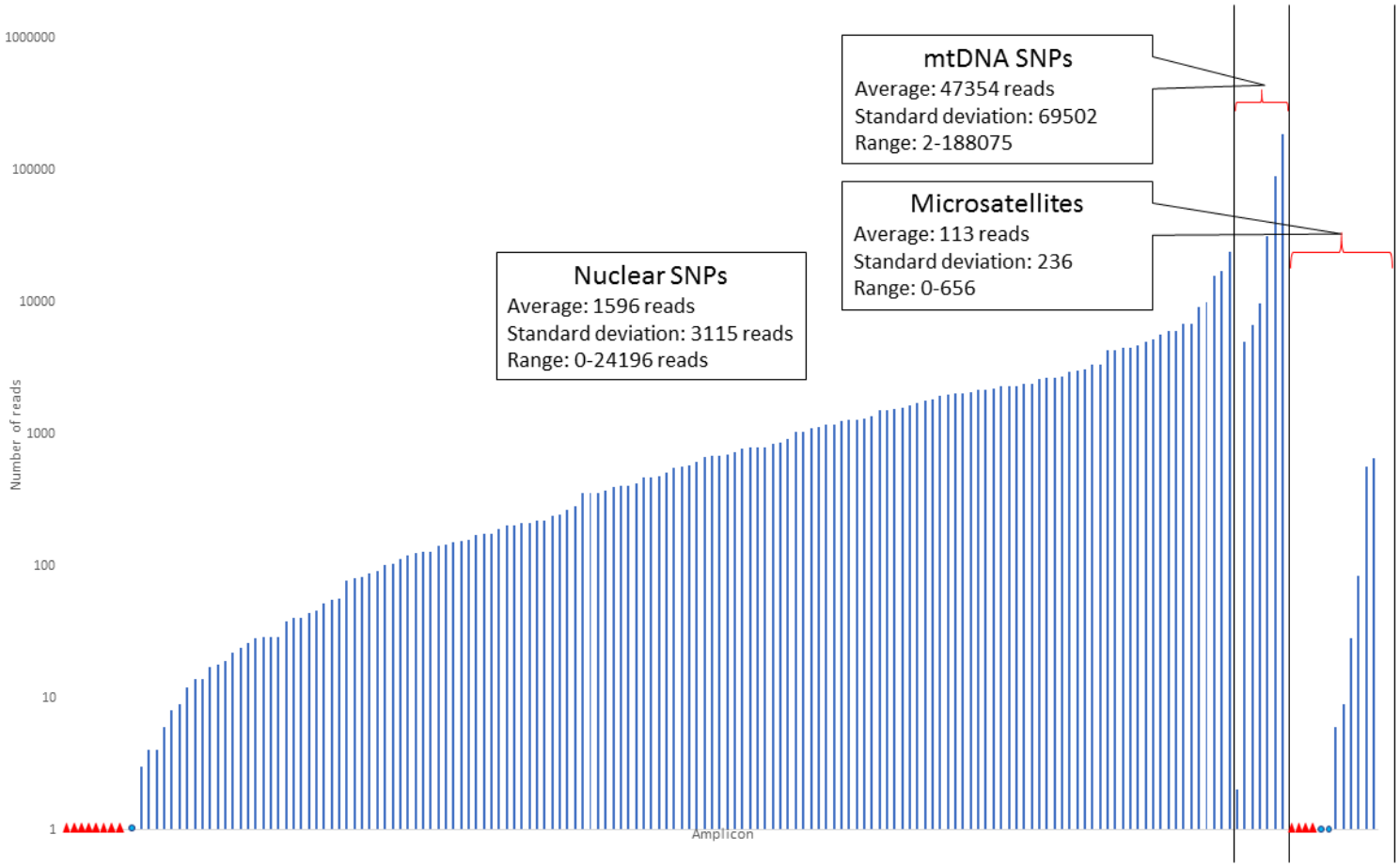


Figure 1. Number of reads partitioned by nuclear SNP containing amplicons, mtDNA SNP containing amplicons, and microsatellites for an initial GT-seq optimization run.
The total number of reads is 577,097, which were amplified from a single individual and sequenced on a MiSeq. See text for details and Table 8 for a summary of data used to make the plot. Red triangles represent loci that had zero reads, and blue circles represent loci with 1 read. Note the log scale on the y-axis.

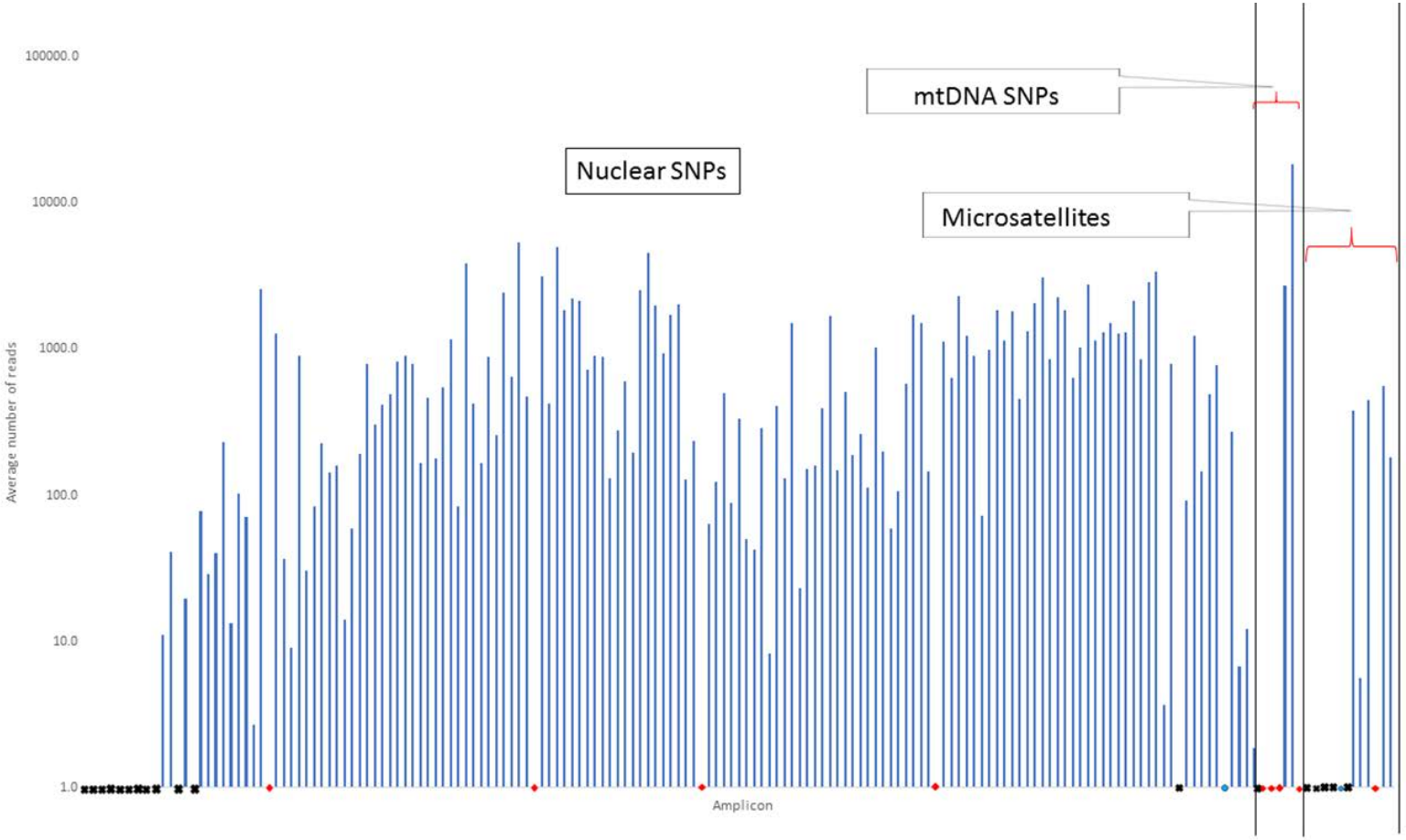


Figure 2. Number of reads partitioned by nuclear SNP amplicons, mtDNA SNP amplicons, and microsatellites for the first GT-seq run of n = 96 individuals.

The order of loci is the same as within Figure 1. The y-axis represents the average number of reads across all 96 individuals at each locus. See text for details and Table 8 for data used to make the plot. Red triangles represent loci that had zero reads, blue circles represent loci with an average of one read or less, and black "X"s represent primer pairs that were removed from the multiplex prior to amplification. Note the log scale on the y-axis.

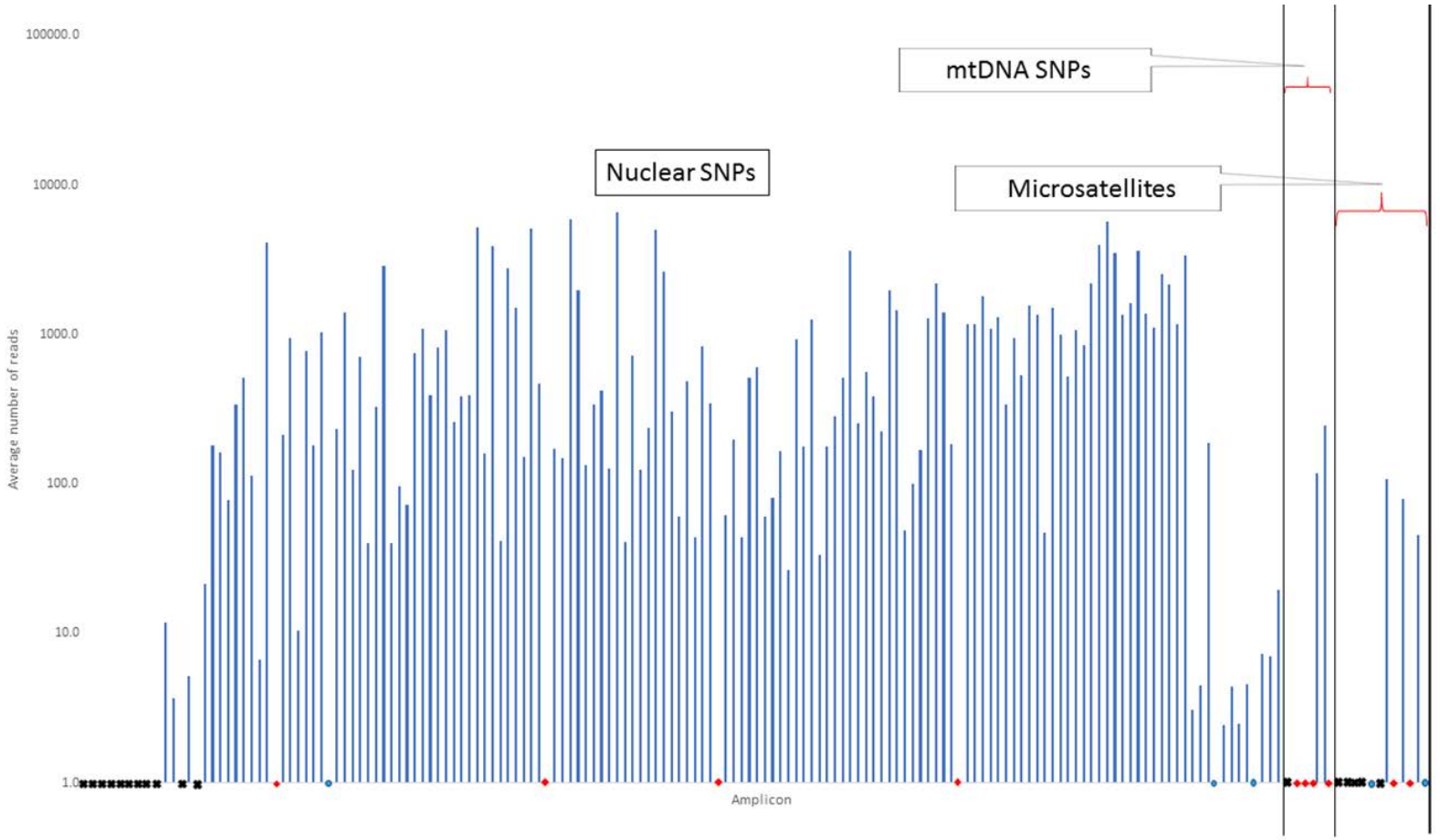


Figure 3. Number of reads partitioned by nuclear SNP amplicons, mtDNA SNP amplicons, and microsatellites for the second GT-seq run of n = 96 individuals.

The order of loci is the same as within Figure 1. The y-axis represents the average number of reads across all 96 individuals at each locus. See text for details and Table 8 for data used to make the plot. Red triangles represent loci that had zero reads, blue circles represent loci with an average of one read or less, and black "X"s represent primer pairs that were removed from the multiplex prior to amplification. Note the log scale on the y-axis.

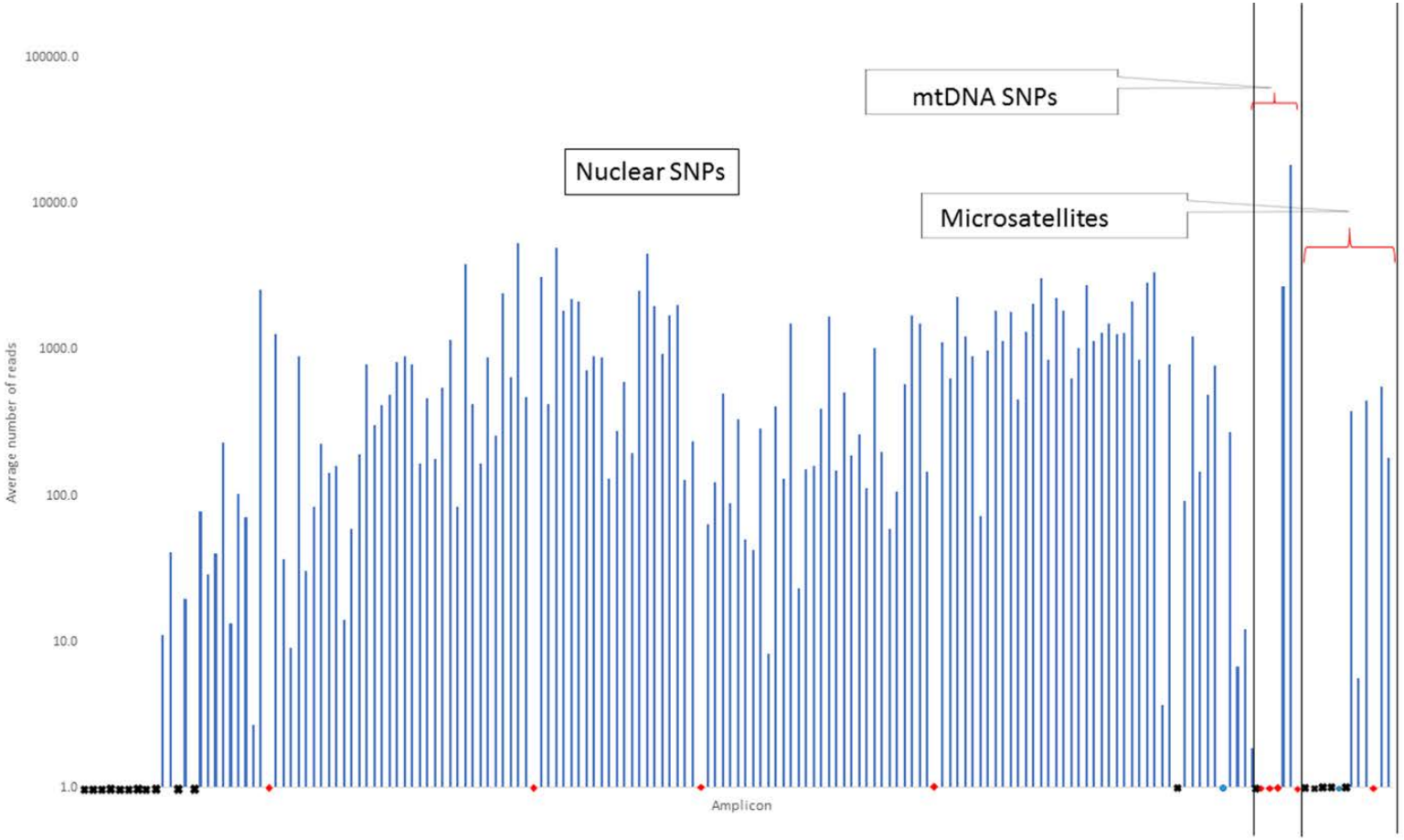


Figure 4. Number of reads partitioned by nuclear SNP amplicons, mtDNA SNP amplicons, and microsatellites for the third GT-seq run of n = 96 individuals.

The order of loci is the same as within Figure 1. The y-axis represents the average number of reads across all 96 individuals at each locus. See text for details and Table 8 for data used to make the plot. Red triangles represent loci that had zero reads, blue circles represent loci with an average of one read or less, and black "X"s represent primer pairs that were removed from the multiplex prior to amplification. Note the log scale on the y-axis.

3.2.3 GT-seq Run Summary

3.2.3.1 Nuclear SNPs

The final merged VCF file combining data from all the nuclear SNP loci across *n* = 288 individuals was filtered to retain genotypes with a minimum coverage depth of 10 and loci with a maximum of 70% missing data; any loci with indels were removed. Application of these filters resulted in a final VCF file containing all *n* = 288 individuals and *n* = 118 loci.

3.2.3.2 mtDNA SNPs

Sequencing of the candidate mitogenomic SNPs identified previously by TL King within the GT-seq panel was largely unsuccessful. Only two of the primer pairs (Sturg_cox1, and Sturg_atp6) ultimately amplified consistently among most individuals with greater than 10X coverage. However, the putative SNPs identified by TL King within these two amplicons were invariant.

3.2.3.3 Microsatellites

Only two of the microsatellite loci (*Aox*D44 and *Aox*D165) amplified consistently among the three GT-seq runs. Their inclusion was largely exploratory, and they were not examined further.

3.2.3.4 Assessment of the Utility of Mitogenomic SNPs for MSA

Assessment of the utility of mitogenomic SNPs for MSA – A reanalysis of the shotgun genomic data for mitogenomic variation yielded a VCF file that was filtered to retain loci with a minor allele frequency of 0.05 or above, minimum coverage depth of 5, and minimum mapping quality of 30. This final VCF file contained 65 individuals and 45 sites. Comparison of these 45 sites against those in **Table 5** revealed the following 10 shared variant sites with the earlier analyses of T.L. King: 3085, 3393, 6674, 9698, 10221, 15750, 16029, 16206, 16126, and 16236.

Twenty-one of the SNPs we identified were in the control region (within bp 15708–16565 of reference KP997217). An earlier assessment of variation within a 203 bp portion of the control region among 322 Atlantic sturgeon corresponding to bp 15680–15883 of the reference KP997217 by Wirgin et al. (2000) and Grunwald et al. (2008) were in agreement with our results. Within this 203 bp region, Wirgin et al. (2000) and Grunwald et al. (2008) identified 27 polymorphic nucleotide sites, and within our sample of 65 Atlantic sturgeon, we identified seven of the same variant sites. This result makes sense given that our sample size was almost five times smaller than that of Wirgin et al. (2000), and therefore we failed to characterize all the same sites that Wirgin et al. (2000) did. However, our sequencing of the remainder of the control region identified 14 SNPs previously not identified by Wirgin et al. (2000), as well as multiple SNPs within other genes. With a larger sample size, it is probable that multiple additional SNP loci may be present in the mitogenome.

DAPC analysis of the mitochondrial variation largely supported the population structure observed with the microsatellite and SNP-based analyses (**Figure 5**). Three primary groupings were evident, including a northern group (Kennebec, James, Saint Lawrence, Hudson, and Delaware), a southern group (Altamaha, Edisto, Savannah, Ogeechee), and the York River as a separate cluster.

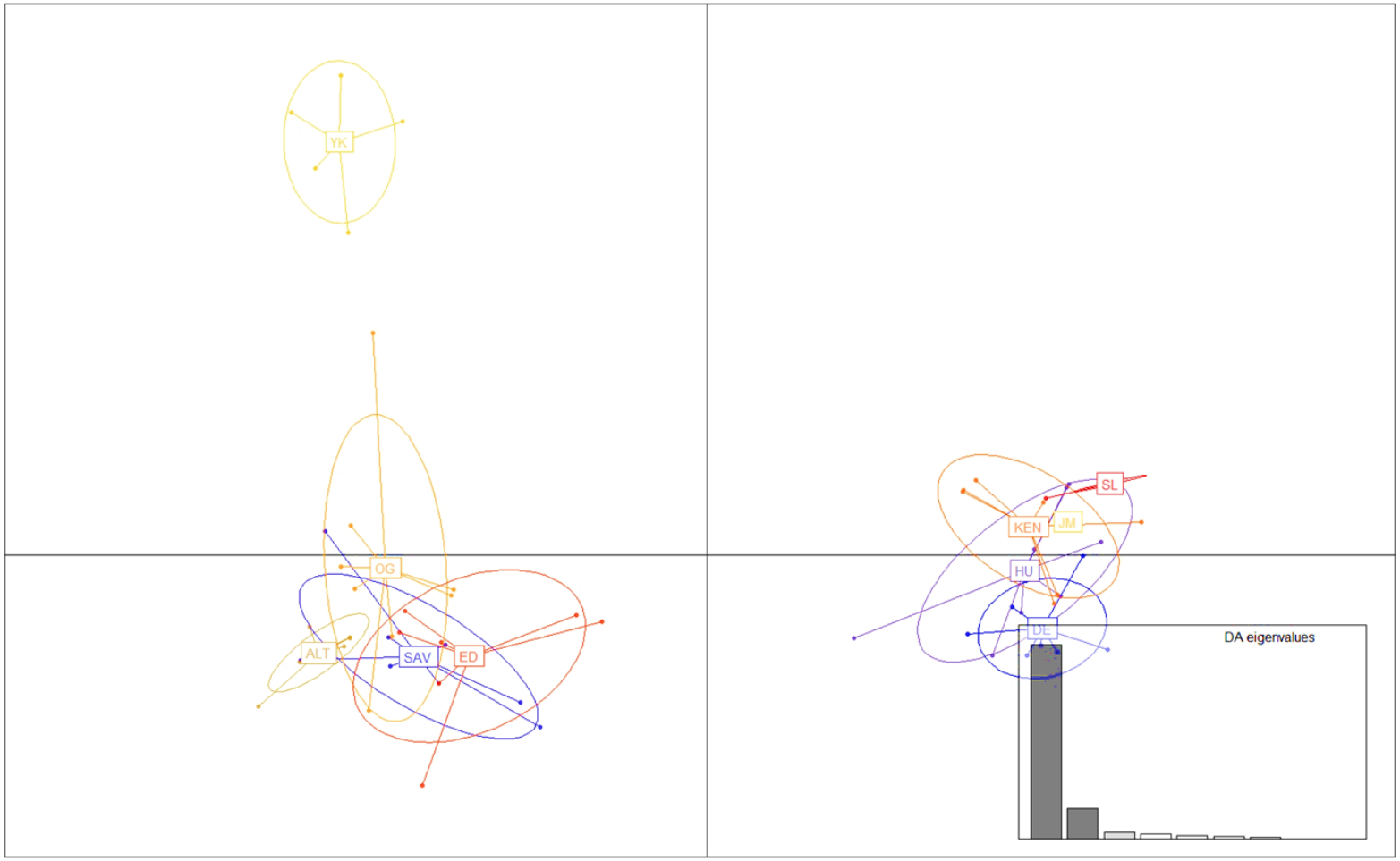


Figure 5. DAPC plot illustrating the relationships among n = 65 Atlantic sturgeon from among 10 Atlantic coastal rivers genotyped at n = 40 mitochondrial SNP loci.
Rivers are abbreviated as YK = York, OG = Ogeechee, SAV = Savannah, ALT = Altamaha, ED = Edisto, DE = Delaware, HU = Hudson, KEN = Kennebec, JM = James, SL = Saint Lawrence. The plot was constructed from in the adegenet package (Jombart 2008, Jombart 2011) within R. The number of principal components retained was 12, and the number of discriminant functions retained was 9.

4 Population Genomics and Assessment of Marker Utility

Our GT-seq panel provided genotype data of Atlantic sturgeon river-resident juveniles (< 500 mm TL) or adults (≥ 1,500 mm TL) captured in their natal river from representative populations from across their range (**Table 9**; **Figure 6**). Following bioinformatic analyses, our genomic baseline for Atlantic sturgeon included 288 individuals genotyped at 118 nuclear SNP loci. On average, 99.1% of loci were able to be genotyped for each individual across the entire data set. In addition, the majority of these individuals had previously been genotyped using 12 microsatellite loci by the USGS LSC as part of their genetic baseline, which enabled a comprehensive comparison between the marker types.

Table 8. Collections of Atlantic sturgeon used to develop a genomic baseline.

River	Samples	Collection Year(s)	Age Class
Kennebec	32	2010–2011	Adult
Hudson	32	2015	Adult
Delaware	32	2008–2009	RRJ
York	32	2013–2015, 2017	Adult
James	32	2013–2015	Adult
Edisto	32	2005	RRJ
Savannah	32	2008, 2013	RRJ
Ogeechee	32	2014–2015	RRJ
Altamaha	32	2011	Adult

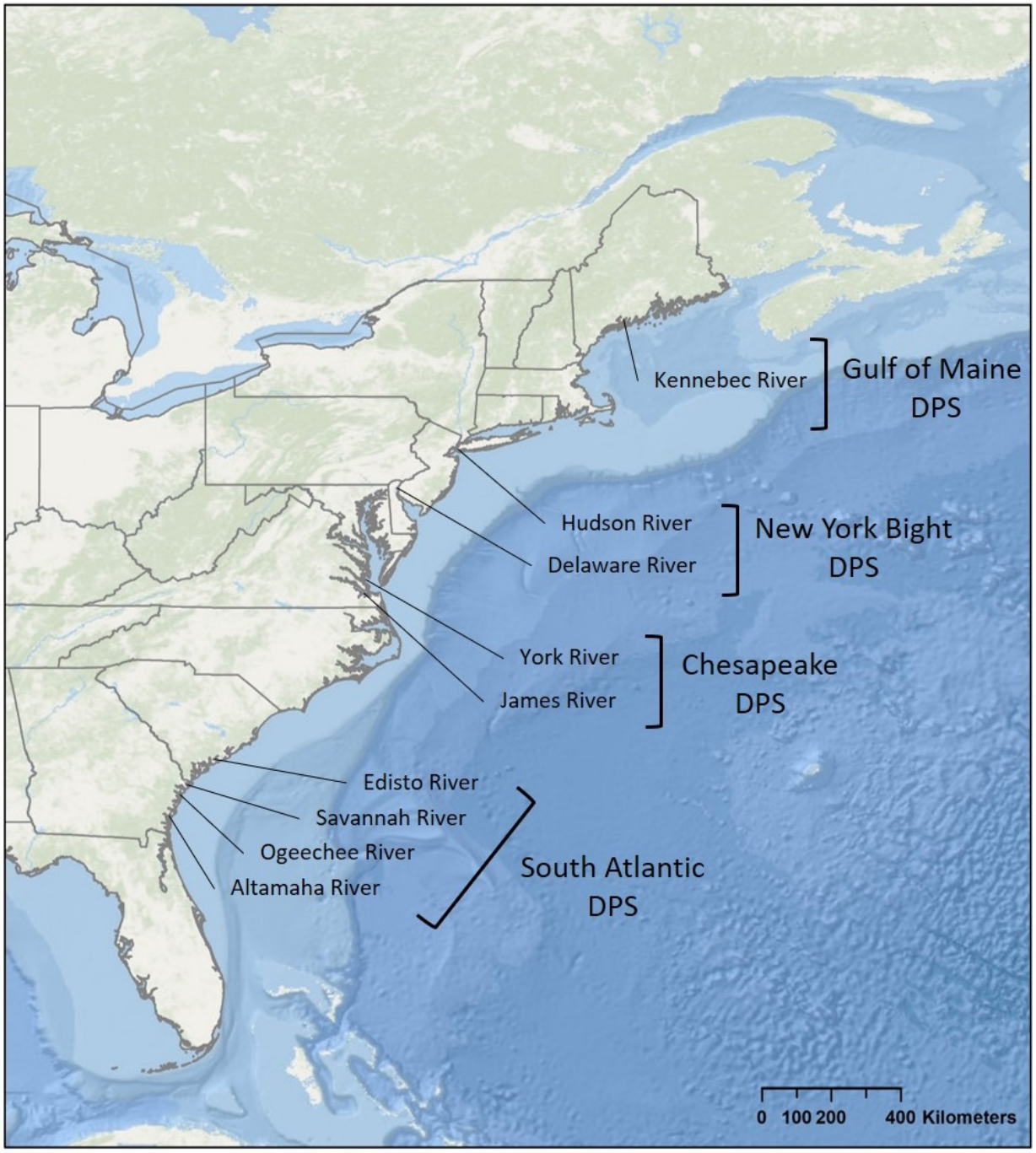


Figure 6. Natal rivers of nine populations of Atlantic sturgeon represented in our genomic baseline.

Thirty-two individuals were chosen to represent each population, consisting of river-resident juveniles (< 500 mm TL) and/or adults (≥ 1,500 mm TL).

4.1 Within-population Diversity

Diversity metrics were calculated for each population using the divBasic function from the diveRsity package in R (Keenan et al. 2013). Allelic richness averaged 1.77 alleles per locus across all populations (range: 1.71–1.82; **Table 10**). Although allelic richness was similar for all populations, there was a general trend of more northern populations containing greater levels of allelic richness. Observed heterozygosity averaged 0.32 (range: 0.31–0.34) across populations and was always greater than expected heterozygosity (mean: 0.25, range: 0.23–0.27). A chi-square test for goodness of fit to Hardy-Weinberg equilibrium expectations (HWE) was run at each locus within each population. Using a simple Bonferroni-adjustment for multiple comparisons (0.05/118) within each population, a critical test statistic of $\alpha = 0.000424$ was applied to identify significant deviations from HWE. Within each population, an average of 17.1 loci (range = 15–19 loci) exhibited significant deviations from HWE. Upon closer examination, seven loci were found to deviate from HWE in all nine populations (**Figure 7**). We suspect these deviations reflect relatively small sample sizes, as well as small effective population sizes, and may be further confounded by and ongoing shift from tetraploidy to diploidy in this taxa. In particular, the overall pattern of heterozygote excess suggests that some loci may be paralogous and fixed heterozygotes—an inherent challenge when working with species that have a history of polyploidy. Future work may wish to consider filtering these loci.

Table 9. Within-population diversity metrics for nine populations of Atlantic sturgeon based on 118 nuclear SNP markers.

Population	Average Number of Individuals Genotyped per Locus	Total Alleles Observed	Proportion of Total Alleles Observed in Population	Allelic Richness	Observed Heterozygosity (H_O)	Expected Heterozygosity (H_E)
Kennebec	31.83	219	95.76	1.82	0.33	0.26
Hudson	31.76	217	95.20	1.82	0.34	0.27
Delaware	31.57	214	93.79	1.78	0.32	0.25
York	31.99	210	92.09	1.75	0.31	0.23
James	31.90	214	94.07	1.78	0.33	0.25
Edisto	30.81	204	89.69	1.71	0.31	0.23
Savannah	31.73	210	92.37	1.74	0.31	0.23
Ogeechee	31.83	213	93.22	1.77	0.32	0.25
Altamaha	31.85	211	92.66	1.75	0.31	0.24

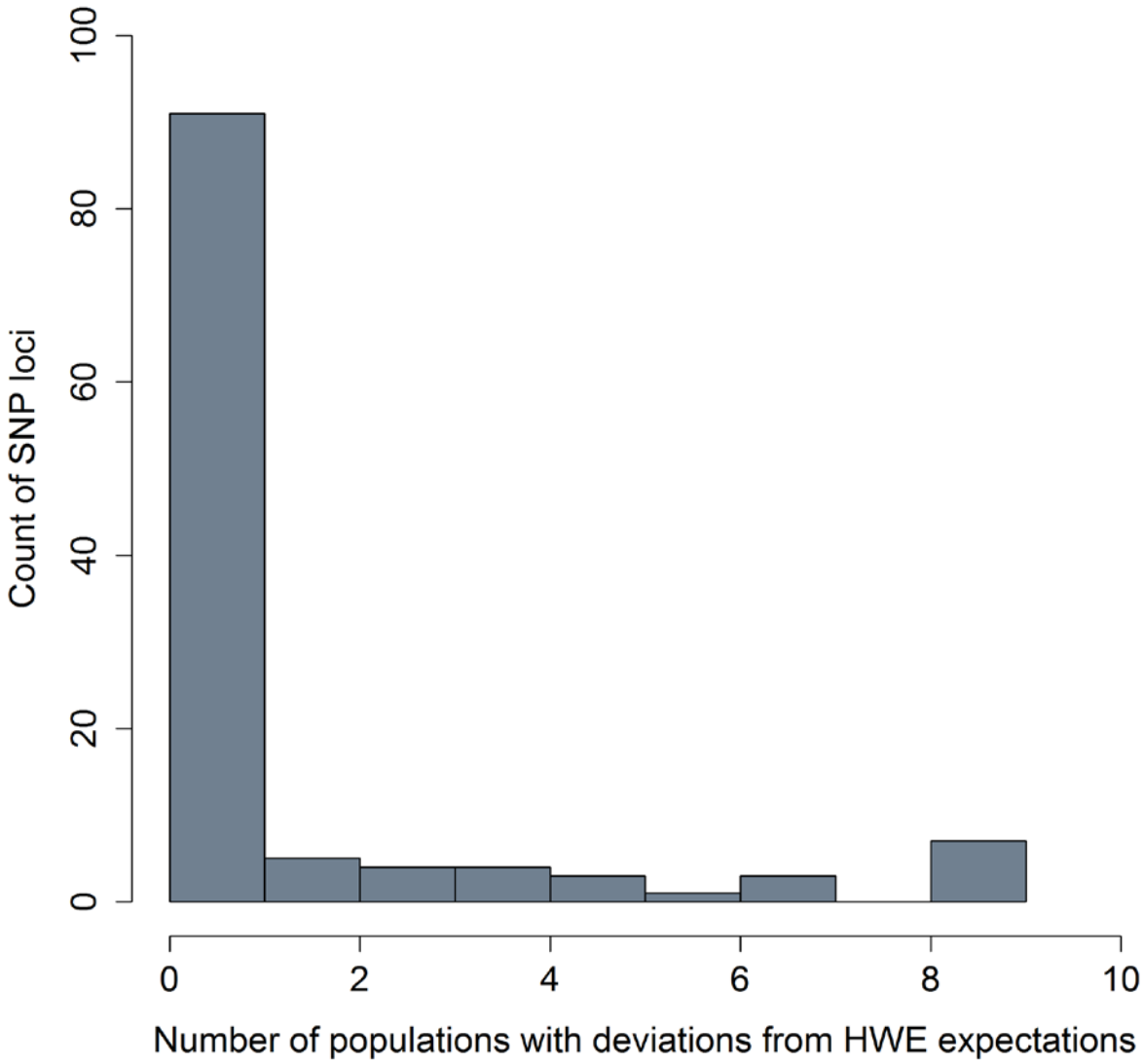


Figure 7. Histogram showing the frequency of deviations (P < 0.000262) from HWE expectations for 118 nuclear SNP loci surveyed across nine populations of Atlantic sturgeon.

We used GENEPOP version 4.2.1 (Raymond and Rousset 1995) to evaluate if there was linkage disequilibrium between loci within each population. Using a critical test statistic of $\alpha = 0.05$, we found fewer instances of linkage disequilibrium between pairs of loci in each of the nine populations examined than would be expected by random chance (**Table 11**), so no further adjustments for multiple comparisons were implemented. In the future, a more complete and thoroughly annotated sturgeon genome would be helpful to ascertain the physical proximity of our SNP loci. Sampling of a larger number of individuals from a larger number of populations would also provide greater statistical power to detect linkage between loci. Based on the available data, we assumed our markers were independent.

Table 10. Results of tests for linkage disequilibrium between 118 pairs of nuclear SNP loci conducted in GENEPOP. A critical test statistic of α = 0.05 was applied to each test.

Population	Number of Loci Pairs Tested	Number of Significant Comparisons
Kennebec	4,362	124
Edisto	3,121	63
Altamaha	3,652	91
Hudson	4,265	105
James	4,093	88
Delaware	3,881	112
Ogeechee	3,821	146
Savannah	3,474	93
York	3,570	100

4.2 Among-population Diversity

Pairwise differentiation (F_{ST}) scores were calculated using the diffCalc function from the diveRsity package (Kennan et al. 2013) in R and are presented in **Table 12**. Mean differentiation between populations was 0.0421 (range: 0.010–0.0975). There was a general pattern of isolation by distance, as populations with natal rivers that were in closer proximity typically exhibited less differentiation. The four populations in the South Atlantic DPS were particularly similar to one another (F_{ST} range: 0.0146–0.0327). In contrast, pairwise comparisons between the Kennebec River and Hudson River populations to populations in the South Atlantic DPS showed the greatest levels of differentiation (range: 0.0442–0.0975).

Table 11. Pairwise differentiation scores (F_{ST}) for nine populations of Atlantic sturgeon based on 118 nuclear SNP markers.

	Kennebec	Hudson	Delaware	York	James	Edisto	Savannah	Ogeechee	Altamaha
Kennebec	0.0000	-	-	-	-	-	-	-	-
Hudson	0.0201	0.0000	-	-	-	-	-	-	-
Delaware	0.0223	0.0125	0.0000	-	-	-	-	-	-
York	0.0568	0.0604	0.0321	0.0000	-	-	-	-	-
James	0.0381	0.0304	0.0114	0.0376	0.0000	-	-	-	-
Edisto	0.0975	0.0854	0.0594	0.0378	0.0454	0.0000	-	-	-
Savannah	0.0830	0.0704	0.0424	0.0281	0.0375	0.0182	0.0000	-	-
Ogeechee	0.0643	0.0442	0.0260	0.0353	0.0264	0.0327	0.0146	0.0000	-
Altamaha	0.0890	0.0768	0.0529	0.0353	0.0475	0.0171	0.0100	0.0159	0.0000

These overall patterns of genetic differentiation among populations were very similar to what has been previously reported among the same nine populations using microsatellite markers (**Table 13**). A mantel test comparing pairwise F_{ST} values based on SNPs and microsatellites for these populations showed a strong ($r = 0.771$) and statistically significant ($P < 0.001$) correlation. This result suggests that the microsatellite panel is a reasonable proxy for genome-wide patterns of differentiation. However, the pairwise values were generally smaller for SNPs than those reported with microsatellite markers. This is an expected result as microsatellites exhibit high mutation rates and diverge much more rapidly among populations with restricted gene flow when compared to nuclear SNPs (Fischer et al 2017).

Table 12. Pairwise F_{ST} scores for nine populations of Atlantic sturgeon based on 12 microsatellite markers.

	Kennebec	Hudson	Delaware	York	James	Edisto	Savannah	Ogeechee	Altamaha
Kennebec	0.0000	-	-	-	-	-	-	-	-
Hudson	0.0443	0.0000	-	-	-	-	-	-	-
Delaware	0.0566	0.0211	0.0000	-	-	-	-	-	-
York	0.1094	0.1176	0.1096	0.0000	-	-	-	-	-
James	0.0687	0.0539	0.0488	0.0955	0.0000	-	-	-	-
Edisto	0.1192	0.1097	0.1139	0.1061	0.0788	0.0000	-	-	-
Savannah	0.1028	0.1004	0.1041	0.0900	0.0591	0.0248	0.0000	-	-
Ogeechee	0.0837	0.0833	0.0943	0.1079	0.0531	0.0391	0.0286	0.0000	-
Altamaha	0.1206	0.1201	0.1180	0.1015	0.0769	0.0308	0.0120	0.0453	0.0000

We used a DAPC (Jombart et al. 2010) implemented in the adegenet package (Jombart 2008, Jombart 2011) in R to visualize the relationships among populations (**Figures 8**–**9**). In our analysis, we retained 60 principal components and eight discriminant functions, based on the guidance provided in an online tutorial (http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf) written by the developer of the adegenet R package. This analysis found results that were largely concordant with past inferences based on microsatellite markers and the existing DPS management system, predicated on genetic variation across the coast with geographic clusters of genetically similar populations. The Kennebec River population formed a cluster that was nearly discrete from other populations. The Hudson and Delaware River populations were located in close proximity to one another, with some overlap with the James River population. In contrast, the York River population formed another nearly discrete cluster. The uniqueness of the York River populations could be the result of recent genetic bottlenecks and genetic drift, or perhaps might reflect a unique evolutionary lineage that has persisted for longer periods. The four populations we examined from the South Atlantic DPS clustered together with substantial overlap among the groups. This is consistent with inferences drawn from microsatellite markers and may reflect greater connectivity among these geographically proximate populations.

We ran STRUCTURE 2.3.4 (Falush et al. 2003), a Bayesian clustering program, to provide an additional insight into Atlantic sturgeon population structure (**Figure 10**). The number of clusters tested ranged from $K = 1$ through $K = 10$. The model we used allowed admixture to occur among clusters, used correlated allele frequencies, and did not include informative priors based on collection location. Each run consisted of a burn-in period of 100,000 steps followed by 100,000 steps that were used for data collection. Five iterations of each level of *K* were run, and each was started using a sequential random seed. We evaluated the model outputs using STRUCTURE HARVESTER (Earl and vonHoldt 2012) and Structure Selector (Li and Liu 2018). We used the default settings of CLUMPAK (Kopelman et al. 2015) implemented using StructureSelector (Li and Liu 2018) to visualize the consensus solution for each candidate *K*.

The most supported number of clusters varied among the different diagnostic tools considered. Evanno's delta *K* most strongly supported two genetic clusters (Evanno et al. 2005). The maximum likelihood approach LnP(*K*) found the highest level of support was associated with four genetic clusters. The estimators published by Puechmaille (2016) supported either four (MedMed*K*) or five (MedMea*K*, MaxMed*K*, and MaxMea*K*) genetic clusters within the data set.

At $K = 2$, two genetic clusters were identified that largely reflected northern and southern populations with a break around the Chesapeake Bay. Interestingly, the James River showed stronger affinity for the northern cluster than the York River, despite their relative locations on the coast. The Edisto, Savannah, and Altamaha River populations showed strong affinity to the same cluster, but the Ogeechee River was

not as clearly resolved. At $K = 3$, an additional cluster was identified that was represented primarily by fish in the James and Ogeechee River populations, with a smaller component in the Delaware River. For the STRUCTURE runs that assumed $K = 4$, the York River population clearly formed a distinct cluster relative to the other populations we considered. This is consistent with our DAPC analysis, which also found the York River population to be unusual relative to the other populations we evaluated. At $K = 5$ through $K = 10$, all additional clusters appeared to be spurious (Puechmaille 2016).

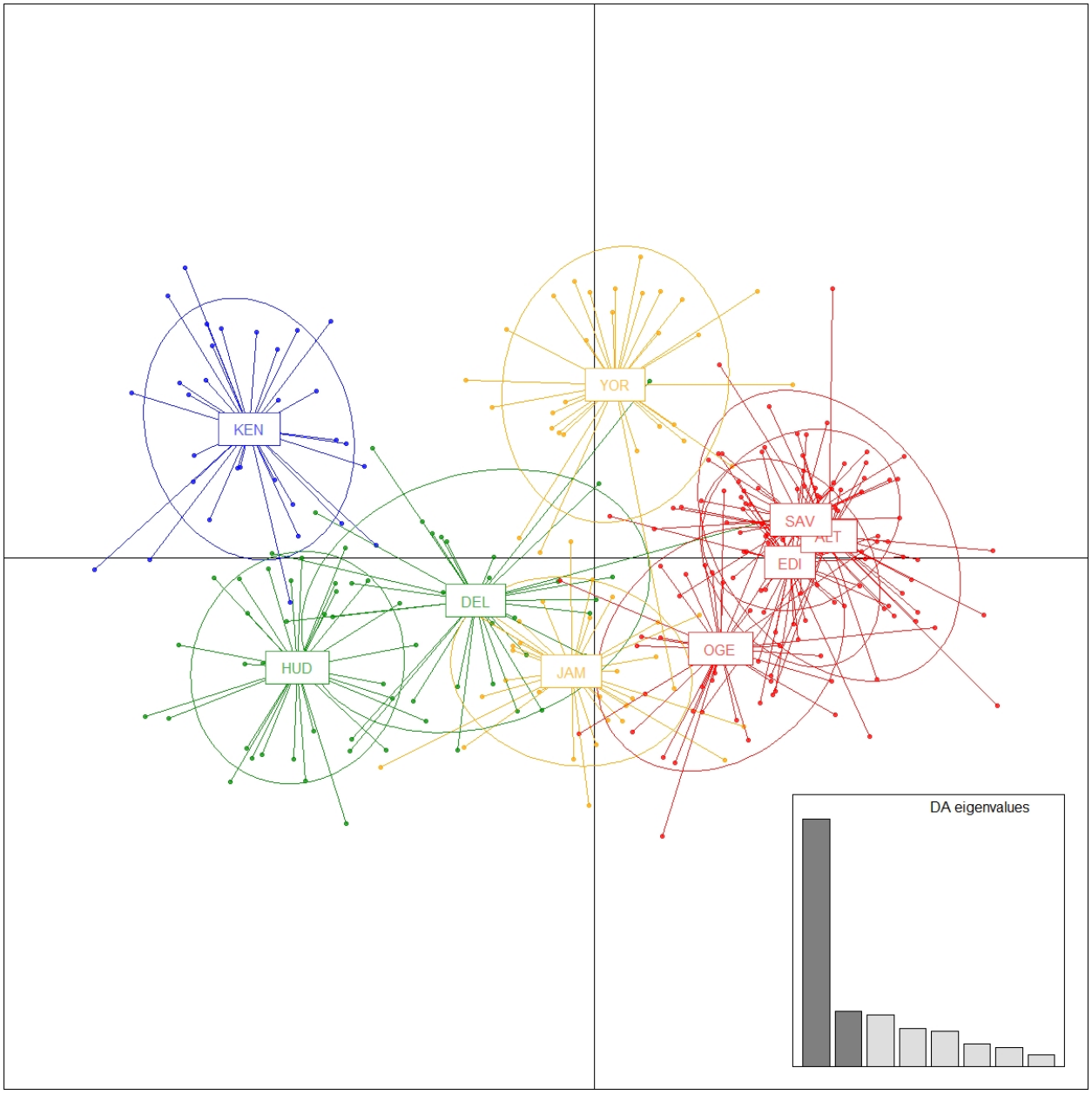


Figure 8. Scatterplot showing the results of discriminant analysis of principle components (axis 1 and axis 2).
Populations are denoted with the following abbreviations: KEN – Kennebec River, HUD – Hudson River, DEL – Delaware River, YOR – York River, JAM – James River, EDI – Edisto River, SAV – Savannah River, OGE – Ogeechee River, ALT – Altamaha River. Each DPS is shown in a different color: Gulf of Maine – blue, New York Bight – green, Chesapeake – yellow, South Atlantic – red.

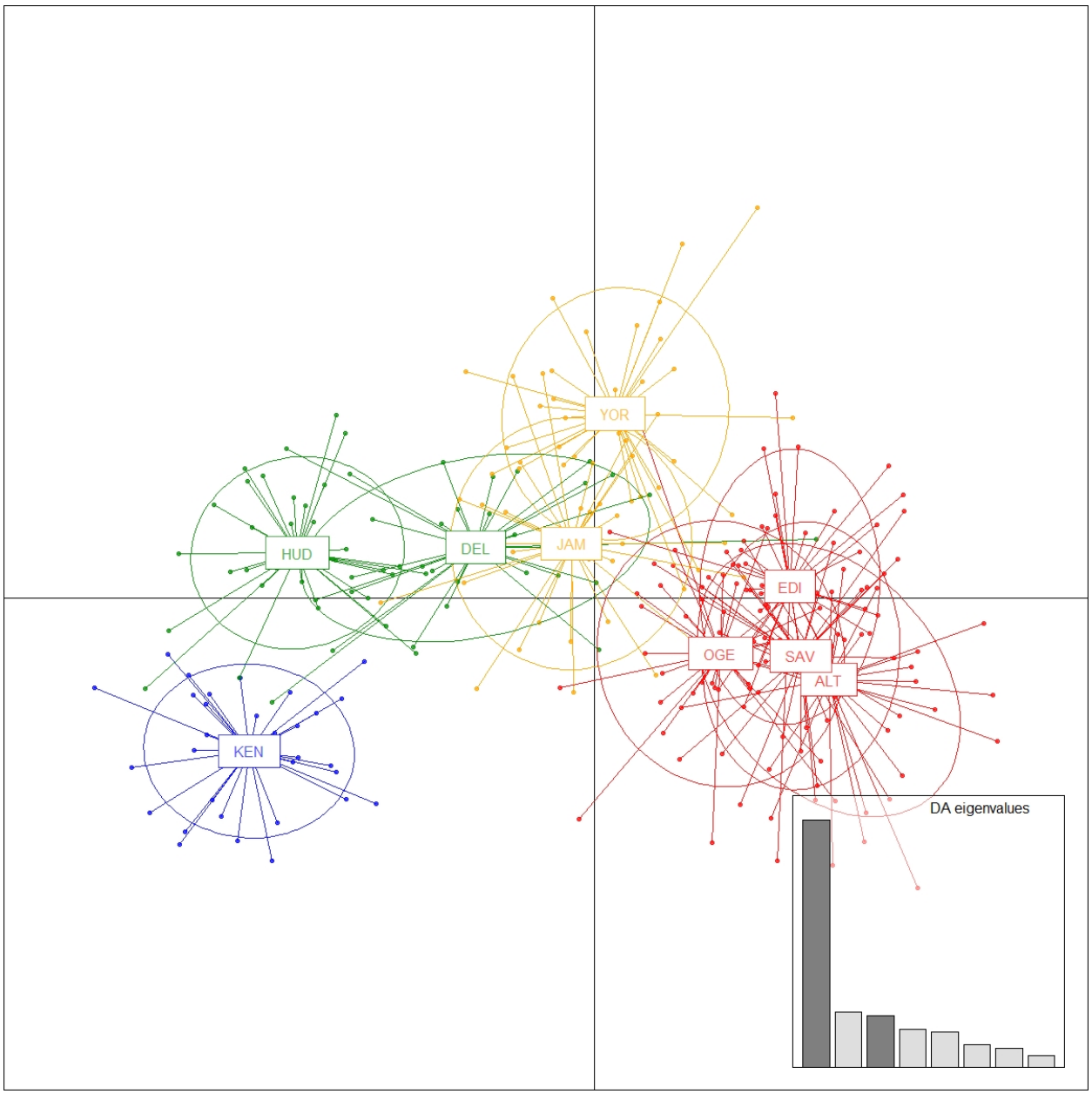


Figure 9. Scatterplot showing the results of discriminant analysis of principle components (axis 1 and axis 3).
Populations are denoted with the following abbreviations: KEN – Kennebec River, HUD – Hudson River, DEL – Delaware River, YOR – York River, JAM – James River, EDI – Edisto River, SAV – Savannah River, OGE – Ogeechee River, ALT – Altamaha River. Each DPS is shown in a different color: Gulf of Maine – blue, New York Bight – green, Chesapeake – yellow, South Atlantic – red.

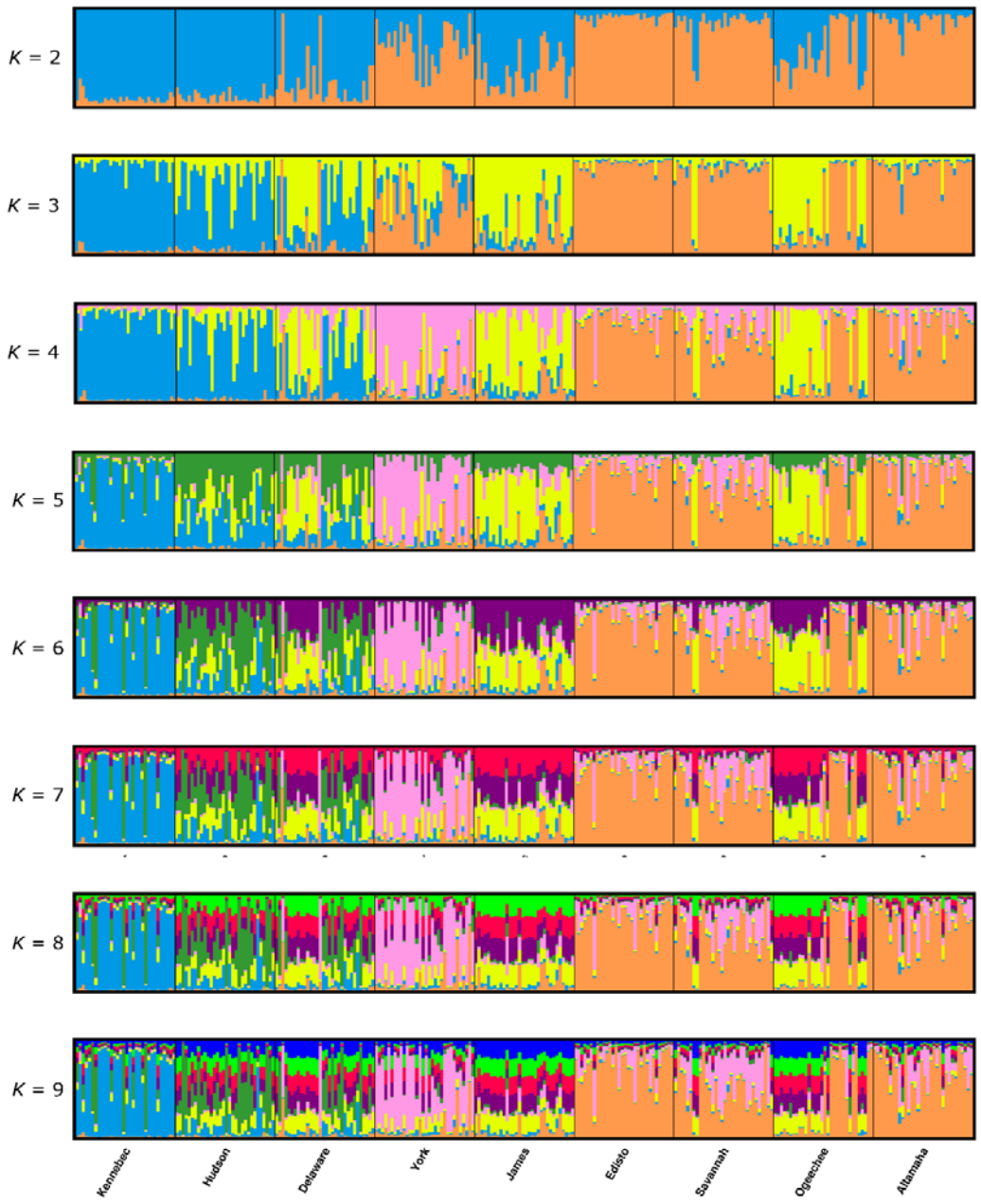


Figure 10. Population clusters of Atlantic sturgeon inferred using STRUCTURE (based on an assumed K = 2 through K = 9).
We ran five iterations of each potential *K* using 100,000 burn-in steps followed by 100,000 steps that were recorded for data collection. We used the default settings of CLUMPAK (Kopelman et al. 2015) implemented using StructureSelector (Li and Liu 2018) to visualize the consensus solution for each candidate *K*. Each individual is represented as a vertical bar, and the proportion of its genome assigned to each of the inferred clusters is represented by separate colors.

4.4 Assignment Testing

Genetic assignment testing was conducted using GeneClass2 (Piry et al. 2004) using the criterion of Rannala and Mountain (1997) with the newly established genomic baseline as a reference. We tested the efficiency of this approach by assigning known samples to rivers within the baseline using a leave-one-out approach. Overall, 66.3% of individuals were correctly assigned to their river of origin, (**Table 14**). Individuals from the Kennebec and James rivers were most clearly resolved (81.3%). In contrast, fewer than half of the individuals collected in the Delaware and York rivers were correctly assigned (40.6% and 46.9%, respectively). Assignment efficiencies in the South Atlantic averaged 63.2% accuracy to river but were much stronger when aggregated at the DPS level. Many of the individuals that were assigned to the wrong natal river were at least assigned to a river within the same DPS. The Chesapeake DPS was a notable exception, as misclassified individuals from the James and York rivers were often incorrectly assigned to the New York Bight or South Atlantic DPS.

Overall, 86.5% of individuals were correctly assigned to their DPS of origin (**Table 15**). Because Federal management of Atlantic sturgeon is based on the five DPSs, this performance metric warrants discussion. Prior to the current study, assignment testing for Atlantic sturgeon was primarily conducted with microsatellite markers and there was a desire to improve assignment efficiency to reduce misassignments. However, the microsatellite genetic baseline maintained by LSC-KCGL was substantially improved through the addition of many more individuals within reference populations in recent years. The current microsatellite genetic baseline maintained by LSC-KCGL correctly assigns 84.9% of individuals to their natal river and 95.8% of individuals to their DPS, based on 12 reference populations. Consequently, the current genomic baseline is several times more likely to incorrectly assign the origin of an Atlantic sturgeon.

The performance of our genomic baseline may be inhibited by several factors. First, we used a relatively modest number of SNP markers for assignment testing (n = 118). There is a recent trend in genomics of moving towards a much larger number of loci for stock assignments (thousands of loci or more). This trend has been driven by rapidly decreasing sequencing costs and the development of new methods (e.g., Ali et al. 2016). Our genomic baseline was developed from a modest (n = 32) number of individuals per population, and as a result our characterization of baseline allele frequencies is imperfect. This is especially important when biallelic SNPs with subtle differences in allele frequencies are considered. Although we were not able to genotype every individual at every loci, our high overall coverage (99.1%) suggests that this was not a major factor influencing our assignment efficiencies.

To explore how the number of markers we used may have influenced our assignment success, we conducted an *ad hoc* sensitivity analysis using varying numbers of microsatellite and SNP markers from our existing assays. We reran GeneClass2 and omitted varying numbers of loci from the baseline assignment tests based on the order they were included in the input file, which was assumed to be random. There results were plotted to show the rate of improvement as additional loci were considered with each class of marker (**Figure 11**). Our *ad hoc* sensitivity suggested that greater increases in assignment performance could be realized by adding a few additional microsatellites relative to larger numbers of SNPs. Still, a much larger SNP panel (which may have to be genotyped using a different method than GT-seq) could offer enhanced accuracy over the current microsatellite markers.

Table 13. Classification confusion matrix for genomic population assignments to river based on 118 nuclear SNPs using the Atlantic sturgeon baseline.

	Assigned River								
Collection Location	Kennebec	Hudson	Delaware	York	James	Edisto	Savannah	Ogeechee	Altamaha
Kennebec	**26**	1	4	-	-	-	-	1	-
Hudson	1	**21**	9	-	1	-	-	-	-
Delaware	1	9	**13**	2	4	1	2	-	-
York	-	-	1	**24**	3	-	2	-	2
James	-	-	4	-	**26**	-	-	2	-
Edisto	-	-	-	-	1	**25**	3	-	3
Savannah	-	-	1	1	-	2	**20**	3	5
Ogeechee	-	-	3	-	2	6	2	**15**	4
Altamaha	-	-	-	1	1	5	2	2	**21**
Grand Total	28	31	35	28	38	39	31	23	35

Notes: Overall, 66.3% of individuals within the baseline were correctly assigned to their river of origin using GeneClass2. Correct assignments are denoted in bold text.

Table 14. Classification confusion matrix for genomic population assignments to river based on 118 nuclear SNPs using the Atlantic sturgeon baseline.

	Assigned DPS			
Collection Location	Gulf of Maine	New York Bight	Chesapeake Bay	South Atlantic
Gulf of Maine	**26**	5	0	1
New York Bight	2	**52**	7	3
Chesapeake Bay	0	5	**53**	6
South Atlantic	0	4	6	**118**

Notes: Overall, 86.5% of individuals within the baseline were correctly assigned to their river of origin using GeneClass2. Correct assignments are denoted in bold text.

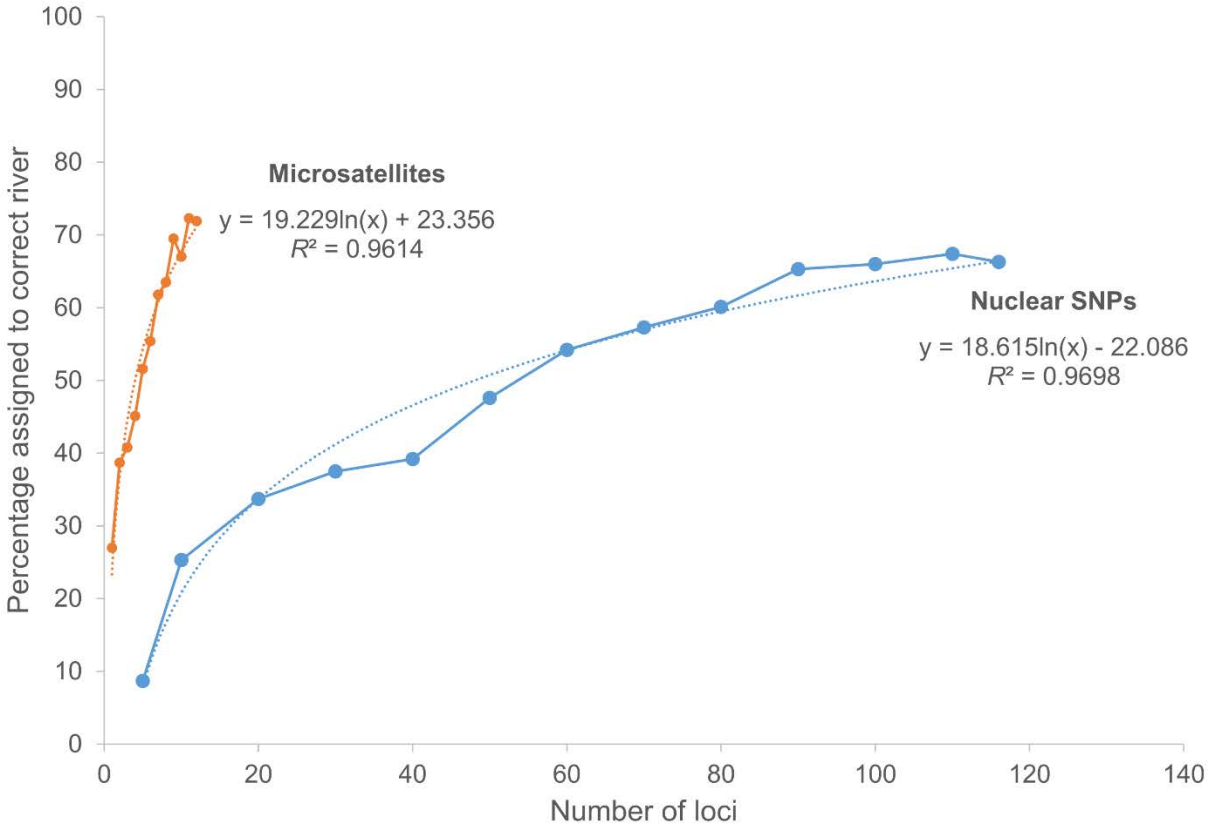


Figure 11. Results of an ad hoc sensitivity analysis comparing the performance of varying numbers of nuclear SNP loci and microsatellite loci for assigning Atlantic sturgeon their natal river using GeneClass2.

A maximum of 118 nuclear SNP loci and 12 microsatellite loci were available for assignments. Across the range of loci numbers considered, it required approximately an order of magnitude more nuclear SNP loci to achieve comparable assignment testing performance relative to microsatellite loci.

5 Application of Molecular Markers to Sturgeon of Unknown Origin

The funded proposal indicated the USGS LSC would perform population genomics techniques on samples collected by the U.S. Navy under a separate BOEM-funded project. However, the U.S. Navy project did not capture Atlantic sturgeon in offshore habitats. As a result, we coordinated with researchers working under a different BOEM contract (cooperative agreement M16AC00003, titled "Monitoring Endangered Atlantic Sturgeon and Commercial Finfish Habitat Use Offshore New York"; co-PIs Michael Frisk and Brian Dunton) to obtain samples from Atlantic sturgeon collected in coastal waters.

Sample collections were authorized by the National Marine Fisheries Services (Endangered Species Permits 16422 and 20351), New York State Department of Environmental Conservation (Endangered/Threatened Species Scientific License 336), and Stony Brook University's Institutional Animal Care and Use Committee (IRB-1022451-4). The Atlantic sturgeon considered herein were collected during targeted research trawls aboard the RV Seawolf. Tows occurred in known marine aggregation sites located off the Rockaway Peninsula, New York, and sampling gear used was similar to that described by Dunton et al. (2015). The Atlantic sturgeon we genotyped and assigned were on average 998 mm TL (range 715–2,320 mm TL) and weighed 6.70 kg (range 1.72–77.00 kg).

5.1 Use of Microsatellites

Because assignment efficiency was lower using the genomic approach developed in the current study when compared to existing microsatellite markers, we consulted with our project officer at BOEM and chose to use our microsatellite markers to provide genetic assignments for these fish to maximize the value of the information generated to BOEM, our colleagues that collected the samples, and the management community.

5.2 DNA Extractions and Microsatellite Genotyping

Molecular analyses were performed by the USGS LSC, Kearneysville, WV. Genomic DNA was isolated from fish tissue using the Puregene Tissue Kit (Gentra Systems, Minneapolis, MN). DNA concentrations were evaluated using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and stock DNA was diluted prior to PCR.

All samples were screened for 12 Atlantic sturgeon microsatellite loci (*LS*19, *LS*39, *LS*54, *LS*68, *Aox*12, *Aox*23, *Aox*45, *Aox*D170, *Aox*D188, *Aox*D165, *Aox*D44, *Aox*D241; described in May et al. 1997, King et al. 2001, and Henderson-Arzapalo and King 2002). Seven multiplexed or single PCR reactions were generated to genotype the 12 microsatellite DNA markers. Amplifications were carried out on either a PTC-200 thermal cycler (MJ Research) or T100 thermal cycler (BioRad) using the following procedure: initial denaturing at 94 °C for 2 min; 35 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 2 min; and a final extension at 72 °C for 10 min. The annealing temperature for loci *LS*19, *LS*54, *LS*68, *Aox*23, and *Aox*45 was lower (53 °C), and the annealing temperature for locus *LS*39 was even lower (48 °C).

PCR products were combined, diluted, and run in six separate reactions on an Applied Biosystems (Foster City, CA, USA) 3130XL Genetic Analyzer using an internal size standard (LIZ-500 or ROX-500, Applied Biosystems). Genemapper 4.0 fragment analysis software was used to bin and output allelic data for each sample. All microsatellite scoring was automated and then checked by eye.

A positive control sample (of known multi-locus genotype) was included on each PCR plate for checking success of PCR amplifications and for correct binning success in the Genemapper analysis software. A negative control sample (containing all the ingredients for PCR amplification except DNA) was included on each PCR plate to make sure there was no contamination in the PCR. Redo PCR was performed on all

samples with missing data due to weak or unamplified alleles. Redo PCR amplifications were done with single loci and not multiplexed PCR as the original samples were done. All Genemapper files were double-checked for scoring errors. Duplicate samples were checked for using the excel add-in Microsatellite Toolkit.

Two matching samples were found within the dataset (Aoxy21255 and Aoxy21258), which is highly unlikely to happen by chance. These two samples were re-extracted using new tissue received and amplified at all loci. After re-checking for matches, Aoxy21255 was found to match with its re-extracted tissue sample, but the new Aoxy21258 now matched another sample Aoxy21259. After re-extracting and amplifying a new sample of Aoxy21259, the new sample did not match the old sample but matched Aoxy21257. Results for all three of these individuals have been omitted from the results presented herein. It is recommended that new tissue samples be extracted, amplified, and genotyped for these individuals.

5.3 Assignment Testing

Genetic assignment testing was conducted using GeneClass2 (Piry et al. 2004) using the criterion of Rannala and Mountain (1997) with the previously described genetic baseline as a reference. We tested the efficiency of this approach by assigning known samples to rivers within the baseline. Overall, 84.9% of individuals were correctly assigned to their river of origin, and 95.8% of individuals were correctly assigned to their DPS of origin. (**Tables 16–17**). Most of the individuals that were assigned to an incorrect river were collected within the South Atlantic DPS and misclassified to another river within the South Atlantic DPS, which primarily reflects lower levels of differentiation in this area (and potentially more recent and/or ongoing gene flow) relative to other areas. No clear patterns were apparent in misclassifications at the DPS level, other than that misclassified individuals were typically assigned to a neighboring DPS. This reflects that neighboring DPSs are generally more similar than those that are further apart.

Next, we used GeneClass2 to assess the origin of genotyped individuals collected in the Atlantic Ocean off of the Rockaway Peninsula in New York. The river with the highest assignment probability was assumed to be the river of origin. Most assignment probabilities were strong (mean: 90.6% and range: 30.7–100.0%). However, results should be interpreted with appropriate caution as the assignments are based solely on allele frequencies, and it is possible that by chance an individual has a genotype that is more common in another population.

The majority of Atlantic sturgeon were assigned to the Hudson River population (87.1%, **Table 18 and Figure 12**). This result is generally congruent with previous mixed stock analyses in the New York Bight, which have also found that most Atlantic sturgeon sampled in this area are from the Hudson River population (Dunton et al. 2012, Waldman et al. 1996, O'Leary et al. 2014, Wirgin et al. 2015). A smaller proportion of the individuals were assigned to the Delaware River population (7.53%). In total, 94.6% of the Atlantic sturgeon we analyzed were assigned to the New York Bight DPS. Also consistent with previous studies, a small proportion of the sturgeon represented other populations from across the coast (**Table 18 and Figure 12**), including individuals assigned to the Kennebec, York, James, Edisto, and Ogeechee rivers. Four of the five ESA-listed DPSs were represented in the Stony Brook samples. This highlights the propensity of Atlantic sturgeon to roam widely from their natal rivers and form mixed stock aggregations and is also supported by telemetry data (see review by Hilton et al. 2016). Furthermore, it suggests that development activities in this region may impact Atlantic sturgeon stocks across the coast. These assignment testing results were provided to Stony Brook University so that they can be integrated with tagging and telemetry data to provide a more comprehensive perspective of Atlantic sturgeon ecology. Future genetic assignment testing is necessary to evaluate the impacts of proposed and permitted anthropogenic activities on endangered Atlantic sturgeon populations, and to track their recovery.

Table 1615. Classification confusion matrix for genetic population assignments to river using the Atlantic sturgeon baseline.

	Assigned River												
Collection Location	St Lawrence	St John	Kennebec	Hudson	Delaware	York	James	Albemarle	Edisto	Savannah	Ogeechee	Altamaha	Grand Total
St Lawrence	**30**	-	-	-	-	-	-	-	-	-	-	-	30
St John	-	**30**	1	-	-	-	-	-	-	-	-	-	31
Kennebec	1	1	**49**	-	-	-	1	-	-	-	-	-	52
Hudson	-	-	7	**305**	23	-	2	-	-	-	-	-	337
Delaware	-	1	1	22	**149**	-	8	-	-	-	-	-	181
York	-	-	-	1	-	**125**	8	1	-	1	-	-	136
James	1	-	-	9	9	-	**312**	-	1	4	5	5	346
Albemarle	-	-	-	-	-	-	-	**30**	-	5	1	1	37
Edisto	-	-	-	-	-	-	-	-	**85**	9	7	8	109
Savannah	-	-	-	-	-	-	1	1	5	**64**	9	18	98
Ogeechee	-	-	-	-	-	-	-	-	12	11	**80**	12	115
Altamaha	-	-	-	-	-	-	-	-	7	29	2	**148**	186
Grand Total	32	32	58	337	181	125	332	32	110	123	104	192	1,658

Notes: Overall, 84.9% of individuals within the baseline were correctly assigned to their river of origin using GeneClass2. Correct assignments are denoted in bold text.

Table 16. Classification confusion matrix for genetic population assignments to DPS using the Atlantic sturgeon baseline.

	Assigned DPS						
Collection Location	Canadian Rivers	Gulf of Maine	New York Bight	Chesapeake	Carolina	South Atlantic	Grand Total
Canadian Rivers	**60**	1	0	0	0	0	61
Gulf of Maine DPS	2	**49**	0	1	0	0	52
New York Bight DPS	1	8	**499**	10	0	0	518
Chesapeake DPS	1	0	19	**445**	1	16	482
Carolina DPS	0	0	0	0	**30**	7	37
South Atlantic DPS	0	0	0	1	1	**506**	508
Grand Total	64	58	518	457	32	529	1,658

Notes: Overall, 95.8% of individuals within the baseline were correctly assigned to their DPS of origin using GeneClass2. Correct assignments are denoted in bold text.

Table 1817. Stock composition of 186 Atlantic sturgeon sampled in coastal waters off the Rockaway Peninsula, New York as part of the BOEM-funded project titled "Monitoring Endangered Atlantic Sturgeon and Commercial Finfish Habitat Use Offshore New York."

Distinct Population Segment	Population	Number of Assigned Individuals	Percentage of Assigned Individuals
Canadian Rivers	St Lawrence	0	0.00%
Canadian Rivers	St John	1	0.54%
Gulf of Maine	Kennebec	1	0.54%
New York Bight	Hudson	162	87.10%
New York Bight	Delaware	14	7.53%
Chesapeake	York	2	1.08%
Chesapeake	James	4	2.15%
Carolina	Albemarle	0	0.00%
South Atlantic	Edisto	1	0.54%
South Atlantic	Savannah	0	0.00%
South Atlantic	Ogeechee	1	0.54%
South Atlantic	Altamaha	0	0.00%

Notes: Assignments are based on genotypes at 12 microsatellite markers and were run in GeneClass2 using a genetic baseline developed by USGS-LSC. The most probable stock of origin is reported for each individual sample.

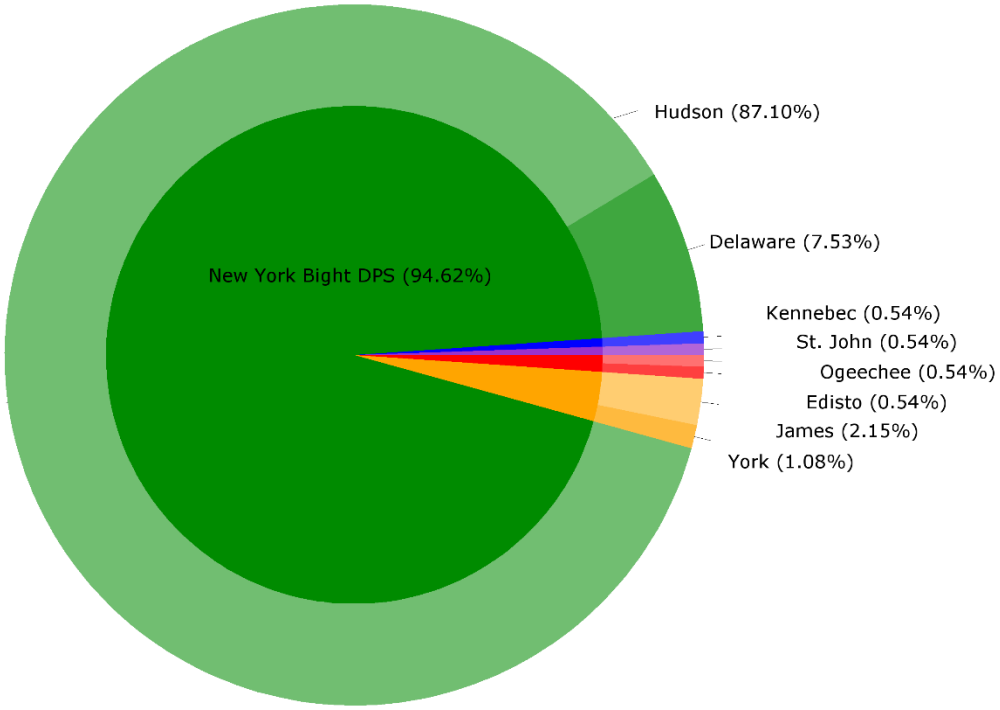


Figure 12. Visualization of stock composition of 186 Atlantic sturgeon sampled in coastal waters off the Rockaway Peninsula, New York.
Sturgeon were collected as part of the BOEM-funded project titled, "Monitoring Endangered Atlantic Sturgeon and Commercial Finfish Habitat Use Offshore New York." Assignments are based on genotypes at 12 microsatellite markers and were run in GeneClass2 using a genetic baseline developed by USGS-LSC. The most probable stock of origin is reported for each individual sample. The inner circle shows assignments grouped by DPS, whereas the outer ring shows assignments to specific spawning rivers within each DPS.

6 General Conclusions

Our results from SNP genotyping were largely concordant with previous insights provided by microsatellite markers. Based on the observed differences in allele frequencies among populations, our genomic baseline supports previous assertations that Atlantic sturgeon show homing to natal rivers, despite mixing extensively in marine waters during non-breeding periods. These differences in allele frequencies provide the foundation for genetic assignment testing to determine the natal river and DPS of Atlantic sturgeon captured in offshore habitats.

Overall, there was a general pattern of genetic isolation by distance, as populations that spawn in rivers that were further apart tended to be more differentiated than those that were geographically proximate. The observed levels of differentiation among populations were somewhat lower using nuclear SNPs relative to those derived from microsatellites, but this is an expected result as microsatellites are known to mutate much more rapidly than other regions of the genome. Lower levels of differentiation (based on SNPs) between populations in the South Atlantic DPS suggest that populations in this region may have greater levels of gene flow relative to their more northerly conspecifics, which has also previously been suggested based on microsatellite data. Population clustering and ordination techniques using the new genomic data both support an overall population structure that is similar to the current DPS management units that were developed primarily based on using microsatellite genetic data.

The GT-seq panel we developed represents an advance in our ability to study the population genomics of Atlantic sturgeon. We were able to efficiently genotype 118 nuclear SNP loci with a cost-effective assay using tools that are available to most genetic laboratories. Our approach allows multiple classes of markers to be sequenced simultaneously (mitochondrial and nuclear SNPs, as well as microsatellites). In the future, our GT-seq approach could be modified to include markers for specific traits (as they become available), such as sex or disease resistance. If implemented, this would provide a truly comprehensive genomics approach—allowing genotypic data from microsatellites, nuclear SNPs, mitochondrial SNPs, and specific coding regions to be generated simultaneously and at reasonable cost.

Although GT-seq is an attractive approach for high throughput genotyping, it can be challenging to get balanced representation of loci. Adjustments of primer concentrations, both among the nuclear loci as well as the mtDNA and microsatellite loci, had somewhat unpredictable effects. For example, some loci ceased to amplify after adjustment of the concentration of unrelated locus primers. This suggests that the interactions of primers in the multiplex is unpredictable, and that small changes of the concentration of some loci can have unanticipated effects on others. Considering just the nuclear loci, we were ultimately able to reliably genotype 118/151 loci for a success rate of ~78%, whereas Campbell et al. (2014) reported a genotyping success rate of 187/192 loci (~97%). However, our success rate is aligned well with that of Aykanat et al. (2016), who successfully amplified 217 of 282 loci (~75%). Our results suggest that when designing a GT-seq assay, it is prudent to over-estimate the number of loci needed for the research problem at hand, with the knowledge that ~3–25% will likely fail to amplify or be represented at a coverage threshold insufficient for high confidence genotyping.

Although we chose the Illumina MiSeq for sequencing, we are aware that other higher throughput platforms exist such as the Illumina HiSeq that can allow for more cost-effective sequencing of a larger number of individuals (Campbell et al. 2014). For our study, given the exploratory nature of a relatively small number of individuals and loci, as well as the potential need in the future to quickly sequence modest numbers of individuals, the MiSeq was an appropriate choice for sequencing. Nevertheless, studies seeking to genotype many hundreds or thousands of individuals would be best suited to follow the approach of Campbell et al. (2014) and incorporate more barcoding indices and sequence libraries on a HiSeq.

In order to improve assignment success, we recommend developing a genomic assay that includes a much larger number of informative SNP markers and/or successfully genotypes a larger number of microsatellite markers. One molecular marker class that deserves additional investigation is the mitochondrial genome, as it harbors a number of SNPs that were previously identified by Wirgin et al. (2000) as phylogeographically informative as well as additional SNP loci identified by this study. Our assessment of variation in the mitogenome located most of the variability within the control region (n = 21 SNPs), which is not surprising because it should be under less selective pressure than other protein coding genes within the mtDNA. Nevertheless, we did find an additional n = 24 SNPs distributed among protein coding genes such as *nad*1, *nad*3, and *cox*1. This suggests additional sequencing of more individuals should be performed to see if any of these nucleotides are population specific and if they provide increased resolution over previous population structure estimates using just a 203 bp section of the control region. Results of Wirgin et al. (2012) suggest that inclusion of the 203 bp region of the control region combined with microsatellite results from 11 loci only had a small effect on the results of a mixed stock analysis, but no results were presented of the analysis of the mtDNA SNPs alone. Our study found additional SNPs within part of the control region not sequenced by Wirgin et al. (2000) and Grunwald et al. (2008). We hypothesize a large full mtDNA dataset of ~700 fish as in Grunwald et al. (2008) could recover additional loci with phylogeographic utility. Given that mtDNA is circular and more resistant to degradation than nuclear DNA, and at a higher copy number in genomic DNA samples, a high throughput amplicon sequencing panel targeting mtDNA SNPs could be a valuable tool. Such an approach would likely work on degraded DNA samples and potentially offer good stock resolution. In the absence of further research and development of additional SNP markers for Atlantic sturgeon (possibly using an approach other than GT-seq), the existing microsatellite loci are the most effective means available to determine the natal river and DPS of Atlantic sturgeon encountered in offshore waters.

Although our study provides a foundation for additional genomic work on Atlantic sturgeon, we encountered several challenges that are likely to impact future studies. First, Atlantic sturgeon are undergoing functional diploidization from tetraploidy (Rajkov et al. 2014). This process has likely produced many paralogous loci, which if undetected could skew inferences made from SNP markers. Although we attempted to screen out paralogous loci bioinformatically, these loci could present a source of error if treated as orthologous. At the present, there is not a published genome for Atlantic sturgeon or any closely related taxa that can be leveraged to help resolve this issue. In addition, our laboratory has found that quality of DNA extracted from Atlantic sturgeon samples tends to be lower than for other species we have studied, which may present challenges if genomics approaches that rely on long sections of high-quality DNA are to be implemented.

In recent years, there has been increasing interest in the development of offshore energy sources to meet the needs of an expanding U.S. economy. Our genetic assignments of Atlantic sturgeon captured off the coast of New York indicates that spawning populations (and their corresponding DPS) from distant locations may potentially be impacted by offshore activities. In fact, activities in this region of the New York Bight could negatively impact Atlantic sturgeon population from at least four different DPSs. Consequently, there continues to be a need to understand the impacts of offshore activities on Atlantic sturgeon, so as to help support the management and conservation of this federally listed species across the vast coastal areas where it occurs. In particular, genetic and genomic assignment testing can link sturgeon encountered in marine aggregations to their natal habitats and help partition the impacts of permitted activities on specific management units (DPS or population). To develop effective conservation strateogies, we recommend that genetic and genomic insights be integrated with the vast amounts of acoustic telemetry data that have been collected over the last decade to improve our understanding of the life history of Atlantic sturgeon and the risks of human activities.

7 References

Aykanat T, Lindqvist M, Pritchard VL, Primmer CR. 2016. From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon *Salmo salar*. Journal of Fish Biology. 89:2658–2679.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. 2016. RAD capture (Rapture): flexible and efficient sequence based genotyping. Genetics. 202:389–400.

[ASSRT] Atlantic Sturgeon Status Review Team, 2007: Status review of Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus*). Prepared by the ASSRT for the National Marine Fisheries Service, National Oceanic and Atmospheric Administration Report, July 27, 2007. 174 p.

Balazik MT, Garman GC, Fine ML, Hager CH, McIninch SP. 2010. Changes in age composition and growth characteristics of Atlantic Sturgeon (*Acipenser oxyrinchus oxyrinchus*) over 400 years. Biology Letters .6:708–710.

Bernatchez L, Wellenreuther M, Araneda C, Ashton DT, Barth JMI, Beacham TD, Maes GE, Martinsohn JT, Miller KM, Naish KA, et al. 2017. Harnessing the power of genomics to secure the future of seafood. Trends in Ecology and Evolution. 32:665–680.

Bowen J, Andrews ST. 2000. The starving time at Jamestown. Faunal Analysis of Pit 1, Pit 3, the Bulwark Ditch, Ditch 6, Ditch 7, and Midden 1 James City County, Virginia. Report to Association for the Preservation of Virginia Antiquities. 150 p.

Campbell NR, Harmon SA, Narum SR. 2014. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources. 15:855–867.

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. 2015. Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. Molecular Plant. 8:831–846.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158.

Dunton KJ, Chapman D, Jordaan A, Feldheim K, O'Leary SJ, McKown KA, Frisk MG. 2012. Genetic mixed-stock analysis of Atlantic Sturgeon *Acipenser oxyrinchus oxyrinchus* in a heavily exploited marine habitat indicates the need for routine genetic monitoring. Journal of Fish Biology. 80:207–217.

Dunton KJ, Jordaan A, Conover DO, McKown KA, Bonacci LA, Frisk MG. 2015. Marine distribution and habitat use of Atlantic sturgeon in New York lead to fisheries interactions and bycatch. Marine and Coastal Fisheries. 7:18–32.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetic Resources. 4:359–361. doi: 10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology. 14:2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164:1567–1587.

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderregger R, Widmer A. 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. BMC Genomics. 18:69.

Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vasquez NJ, Finger JW, Pierson TW, Bentley KE, Hoffberg SL, Louha S, et al. 2016. Adapterama 1: universal stubs and primers for thousands of dual indexed Illumina libraries (iTru & iNext). bioRxiv. doi: http://dx.doi.org/10.1101/049114.

Grunwald C, Maceda L, Waldman J, Stabile J, Wirgin I. 2008. Conservation of Atlantic sturgeon *Acipenser oxyrinchus oxyrinchus*: delineation of stock structure and distinct population segments. Conservation Genetics 9:1111–1124.

Henderson-Arzapalo A, King TL. 2002. Novel microsatellite markers for Atlantic sturgeon (*Acipenser oxyrhinchus*) population delineation and broodstock management. Molecular Ecology Notes. 2:437–439.

Hilton EJ, Kynard B, Balazik MT, Horodysky AZ, Dillman CB. 2016. Review of the biology, fisheries, and conservation status of the Atlantic sturgeon, (*Acipenser oxyrinchus oxyrinchus* Mitchell 1815). Journal of Applied Ichthyology. 32:30–66.

Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, Talbot R, Bishop SC, Archibald AL, Bron JE, et al. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics. 15:90.

Ingram EC, Cerrato RM, Dunton KJ, Frisk MG. 2019. Endangered Atlantic sturgeon in the New York Wind Energy Area: implications of future development in an offshore wind energy site. Scientific Reports. 9:12432.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24:1403-1405.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genomics. 11:94.

Jombart T. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. doi: 10.1093/bioinformatics/btr521

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA.. DiveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution. 4: 782-788

King TL, Lubinski BA, Spidle AP. 2001. Microsatellite DNA variation in Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus*) and cross-species amplification in the Acipenseridae. Conservation Genetics. 2:103–119.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering nodes and packaging population structure inferences across K. Molecular Ecology Resources. 15:1179–1191.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9(4):357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078–2079.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein of nucleotide sequences. Bioinformatics. 22(13):1658–1659.

Li YL, Liu JX. 2018. StructureSelector: a web based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources. 18:176–177.

Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser, Huso*, and *Scaphirhynchus*). Genetics. 158(3):1203–1215.

May B, Krueger CC, Kincaid HL. 1997. Genetic variation at microsatellite loci in sturgeon: primer sequence homology in *Acipenser* and *Scaphirhynchus*. Canadian Journal of Fisheries and Aquatic Sciences. 54: 1542–1547.

Nei M. 1972. Genetic distance between populations. The American Naturalist. 106:283–292.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a smaller number of individuals. Genetics. 76:379–390.

[NMFS] National Marine Fisheries Service. 2010. Endangered and threatened wildlife and plants; proposed listings for two Distinct Population Segments of Atlantic Sturgeon (*Acipenser oxyrinchus oxyrinchus*) in the Southeast. Federal Register. 61:61904–61929.

O'Leary SJ, Dunton KJ, King TL, Frisk MG, Chapman DD. 2014. Genetic diversity and effective size of Atlantic Sturgeon, *Acipenser oxyrhinchus oxyrhinchus* river spawning populations estimated from the microsatellite genotypes of marine-captured juveniles. Conservation Genetics. 15(5):1173–1181.

Puechmaille SJ. 2016. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources. 16:608–627.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A. 2004. GeneClass2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity. 95:536–539.

Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Science. 94: 9197-9201

Rajkov J, Shao Z, Berrebi P. 2014. Evolution of polyploidy and functional diploidization in sturgeons: microsatellite analysis in 10 sturgeon species. Journal of Heredity. 105:521–531.

Raymond M, and Rousset F. 1995. GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. Vsearch: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Secor DH. 2002. Atlantic sturgeon fisheries and stock abundances during the late nineteenth century. American Fisheries Society Symposium. 28:89–98.

Stein AB, Friedland KD, Sutherland M. 2004. Atlantic sturgeon marine distribution and habitat use along the northeastern coast of the United States. Transactions of the American Fisheries Society. 133:527–537.

Supple MA, Shapiro B. 2018. Conservation of biodiversity in the genomics era. Genome Biology. 19:131.

[USFWS and NMFS] U.S. Fish and Wildlife Service, National Marine Fisheries Service. 1996. Policy regarding the recognition of distinct vertebrate population segments under the Endangered Species Act. Federal Register. 61:4722–4725.

Waldman JR, Hart JT, Wirgin II. 1996. Stock composition of the New York Bright Atlantic sturgeon fishery based on analysis of mitochondrial DNA. Transactions of the American Fisheries Society. 125:364–371.

Waldman JR, King T, Savoy T, Maceda L, Grunwald C, Wirgin I. 2013. Stock origins of subadult and adult Atlantic sturgeon, *Acipenser oxyrinchus*, in a non-natal estuary, Long Island Sound. Estuaries and Coasts. 36:257–267.

Wirgin I, Breece MW, Fox DA, Maceda L, Wark KW, King T. 2015. Origin of Atlantic sturgeon collected off the Delaware coast during spring months. North American Journal of Fisheries Management. 35:20–30.

Wirgin I, Waldman JR, Rosko J, Gross R, Collins MR, Rogers SG, Stabile J. 2000. Genetic structure of mitochondrial DNA control region sequences. Transactions of the American Fisheries Society. 129:476–478.

You FM, Huo N, Gu YQ, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics. 9:253.

Department of the Interior (DOI)

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors the Nation’s trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.

Bureau of Ocean Energy Management (BOEM)

The mission of the Bureau of Ocean Energy Management is to manage development of U.S. Outer Continental Shelf energy and mineral resources in an environmentally and economically responsible way.

BOEM Environmental Studies Program

The mission of the Environmental Studies Program is to provide the information needed to predict, assess, and manage impacts from offshore energy and marine mineral exploration, development, and production activities on human, marine, and coastal environments. The proposal, selection, research, review, collaboration, production, and dissemination of each of BOEM’s Environmental Studies follows the DOI Code of Scientific and Scholarly Conduct, in support of a culture of scientific and professional integrity, as set out in the DOI Departmental Manual (305 DM 3).