学校编码：10384　　密级
学号：22320120153434

厦门大学

博士学位论文

美国特拉华湾美洲狼鲈群体遗传结构及系统进化研究

Population genetic structure and phylogentic analysis for *Morone americana* in Delaware Bay of United States

边力

指导教师姓名：　苏永全　教授
　　Patrick M. Gaffney　教授
专　业　名　称：海洋生物技术
论文提交日期：2016 年 5 月
论文答辩时间：2016 年 8 月

2016 年 8 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人(签名)：

年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于年月日解密，解密后适用上述授权。

(　)2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年　月　日

目录

CONTENTS

缩略语中英文对照表

缩写	英文	中文
16S	16S ribosomal RNA	16S 核糖体 RNA
AFLP	Amplified fragment length polymorphisms	扩增片段长度多态性
AMOVA	Analysis of molecular variance	分子方差分析法
ATP6,8	Adenosine triphosphate synthase subunits 6,8	三磷酸腺苷合成酶 F0 亚单位 6，亚单位 8
BLAST	Basic local alignment search tool	基本局域联配搜寻工具
bp	Base pair	碱基对
CCD	Central conserved domain	中央保守区
cDNA	Complementary DNA	互补脱氧核糖核酸
cM	Centimorgan	厘摩
COI-III	Cytochrome *c* oxidase I-III	细胞色素 *c* 氧化酶亚基 I-III
CRoPS	Complexity reduction of polymorphic sequences	简化多态序列复杂度
CSB	Conserved sequence blocks	保守序列区
Cyt *b*	Cytochrome *b*	细胞色素 *b*
DAPC	Discriminant analysis of principal components	主成分判别分析
ddRAD	Double digest restriction site-associated DNA sequencing	双酶切限制性酶切位点相关 DNA 测序
DGGE	Denaturing gradient gel electrophoresis	变性梯度凝胶电泳
DHU	Dihydrouridine	二氢尿嘧啶核苷
DNREC	Delaware department of natural resources and environmental control	特拉华自然资源环保局
GBS	Genotyping by sequencing	基于测序的基因分型
GTR	General time reversible	广义时间可逆
HMM	Hidden Markov model	隐马尔可夫模型
LGM	Last glacial maximum	末次冰盛期
ML	Maximum likelihood	最大似然
MSG	Multiplexed shotgun genotyping	多元鸟枪法基因分型
mtDNA	Mitochondrial DNA	线粒体 DNA
Mya	Million years ago	百万年前
NCBI	National center for biotechnology information	美国国家生物信息中心
ND1-6,4L	Nicotinamide adenine dinucleotide dehydrogenase subunits 1-6 and 4L	烟酰胺腺嘌呤二核苷酸脱氢酶亚基 1-6 及 4L
NGS	Next-generation sequencing	新一代测序
PAGE	Polyacrylanide gel electrophoresis	聚丙烯酰胺凝胶电泳
PCB	Polychlorinated biphenyls	多氯联苯
PCR	Polymerase chain reaction	聚合酶链式反应
ppt	Part per thousand	千分之一（盐度单位）
QTL	Quantitative trait locus	数量性状基因座
RAD-seq	Restriction site-associated DNA sequencing	限制性酶切位点相关 DNA 测序

RAPD	Random amplified polymorphic DNA	随机扩增多态性 DNA
RFLP	Restriction fragment length polymorphism	限制性片段长度多态性
RRLs	Reduced-representation libraries	简化代表库
rRNA	Ribosomal RNA	核糖体 RNA
SNP	Single nucleotide polymorphism	单核苷酸多态性
SSCP	Single-strand conformation polymorphism	单链构象多态性
SSR	Simple sequence repeat	简单序列重复
TAS	Terminal associated sequences	终止相关序列区
TGGE	Temperature gradient gel electrophoresis	温度梯度凝胶电泳
tRNA	Transfer RNA	转运 RNA
Ψ	Pseudouridine	假尿嘧啶核苷

摘要

美洲狼鲈(*Morone americana*，Gmelin, 1789)隶属于鲈形目(Perciformes)，狼鲈科(Moronidae)，狼鲈属(*Morone*)，别名美洲白鲈、白石鮨。美洲狼鲈是在美国东海岸浅海区域十分常见的一种溯河洄游鱼类，每逢春末夏初，生活于近海和河口区域的美洲狼鲈便溯河洄游至河流中上游产卵繁衍。美洲狼鲈也经常被用作海区污染程度的指示种。鉴于美洲狼鲈广泛的分布范围以及巨大的种群数量，其十分适合作为近岸洄游性鱼类群体研究的模式种类。同时，其在检测环境污染方面也具有主要作用。目前有关美洲狼鲈的遗传标记仅仅局限于少量线粒体限制性片段长度多态性(RFLP)标记，需要大量高质量的线粒体以及基因组标记对其匮乏的遗传标记数据库进行补充。本研究首先扩增出美洲狼鲈的线粒体全基因组序列，随后利用 3 个线粒体基因(16S、COI、ND2)和简化基因组技术对美国东海岸 4 个群体进行群体遗传分析，研究其遗传多样性，并探讨了群体演化历史分布，为了解特拉华湾美洲狼鲈的群体资源特征提供了理论依据，使人们可以更好地利用其作为污染指示种。

(1)采用引物步移 PCR(primer-walking PCR)技术对美洲狼鲈的线粒体全基因组进行扩增。美洲狼鲈的线粒体基因组总长为 17966 bp，由 13 个编码基因、22 个转运 RNA、2 个核糖体 RNA 及 1 个控制区组成。除了 ND6 基因外，其余基因的排列顺序与脊椎动物线粒体基因的经典排列顺序相同。而 ND6 基因与其他狼鲈科(Moronidae)的种类相同，均位于控制区的中间位置，而经典的排列顺序中 ND6 位于 ND5 和 Cyt*b* 基因之间。在控制区中识别出了 1 个终止相关序列区、1 个中央保守区、2 个保守序列区以及 1 段长串联重复区，长串联重复区中包括 8 个 121 bp 的完整重复单元和 1 个不完全重复单元。基于线粒体基因构建美洲狼鲈与其他 17 个鲈形目鱼类的进化树，美洲狼鲈被准确地划分到狼鲈属(*Morone*)下，证实了本研究得到线粒体全基因组的有效性，同时进化树结果也验证了前期研究中美洲狼鲈与密西西比狼鲈(*Morone mississippiensis*)的近缘关系。

(2)由美国特拉华湾 3 个群体(Sta92、Bro_R、Mur_R)和切萨皮克湾 1 个群体(Bro_Cr)共计采集 96 尾美洲狼鲈样品，利用线粒体 16S、COI、ND2 这 3 个基因分析了 4 个群体的遗传结构。分析结果时发现，单个基因所得结果的多态性较低，估测群体结构较为困难。因而又将 3 个基因相串联，利用串联后的序列

进行群体分析，得到 4 个群体的的单倍型多样性为 0.507-0.902，核苷酸多样性为 0.00034-0.00085，即单倍型多样性(h)处于较高水平，而核苷酸多样性(π)处于较低水平；群体间的遗传分化指数 Fst 显示除了 Mur_R 与 Bro_R 之间遗传分化不显著外，其余群体间的遗传分化均达到显著水平；群体间遗传距离的计算显示出各群体间的 Kimura-2p 遗传距离较小，AMOVA 的结果也表明大部分的遗传变异来源于群体内。线粒体基因分析得到的结论为各群体间存在着一定的遗传分化，但分化程度仍然不足以将各个群体区分开来，群体差异并不明显。认为盐度可能是造成遗传分化的主要原因，而冰期后较短的分化时间不足以使各群体形成明显的遗传差异。

(3)利用简化基因组 ddRAD 技术对美州狼鲈同样的 4 个群体 96 尾个体进行群体遗传分析。选择 EcoRI 和 NlaIII 作为构建 ddRAD 测序文库的 2 种限制性内切酶，由 96 个个体中共得到 438683 个位点，利用较为严苛的 SNP 数据筛选条件，得到 696 个多态性位点。在哈迪-温伯格平衡检验和连锁不平衡检验中又排除了 22 个位点，随后检测发现受到平衡选择作用的位点为 63 个，受到正向选择作用的位点为 32 个，中性位点为 579 个，利用这 3 类位点分别对 4 个群体进行遗传分析。结果显示，利用受平衡选择的位点未能区分出群体结构，而另外两类位点得到了相同的遗传结构，且受正向选择作用的位点得出的群体间遗传分化更大。利用中性位点分析的结果显示，各个群体间均存在着显著的遗传分化，且可以将 4 个群体分为 3 个大的类群，特拉华湾中部的 Mur_R 和南部的 Bro_R 形成 1 个类群，而北部的 Sta92 与切萨皮克湾的 Bro_Cr 分别形成另 2 个类群。另外，在遗传距离上，Sta92 类群与 Bro_Cr 类群的遗传距离较近，而与另一类群距离较远。认为盐度及人类活动可能是形成该遗传结构的主要原因。基于贝叶斯算法对 4 个群体的历史动态进行估测，最终得到的最佳动态模型是将 Sta92 与 Bro_Cr 聚为一支，Bro_R 与 Mur_R 聚为另一支，首先由一个起源地分化出这两大支，随后再次细分出各自的 2 个群体，推测在切萨皮克湾以南的区域可能存在的冰期庇护所是 4 个群体的发源地。

关键词**:** 美洲狼鲈；线粒体基因组；线粒体基因；限制性酶切位点相关 DNA 测序；群体遗传

Abstract

White perch (*Morone americana*， Gmelin, 1789) is an abundant semi-anadromous species found along the east coast of United States. During late spring to early summer, *M. americana* moves upstream from brackish water or estuary into fresh waters to spawn. It is also considered as a good indicator species of contaminant. Given its abundance and widespread distribution, *M. americana* may serve as a good model for investigating the population genomics of a partially migrating estuarine species. Moreover. It plays important roles on monitoring pollutions.To date, available genetic markers for *M. americana* are limited to mitochondrial DNA RFLP polymorphisms. In order to obtain a high-resolution profile of population structure, larger numbers of nuclear as well as mitochondrial genetic markers are needed.In this study, the complete mitochondrial genome of *M. americana* was first amplified, then three mitochondrial genes (16S、COI、ND2) and double digest restriction site-associated DNA sequencing (ddRAD) technique were applied to analyze the population structure of four populations for *M. americana* in US eastern coast. Our study provided valuable tools for investigation into population resource of this species. In addition, our results are also promising to assist the decision making from department of environment protection.

(1)Primer-walking PCR strategy was used to obtain the complete mitochondrial genome of *M. americana*. The total length of *M. americana* mitogenome is 17966 bp, consisting of 13 protein coding genes, 22 transfer RNAs, two ribosomal RNA genes, and a non-coding control region. As with several other species in Moronidae, the ND6 gene in *M. americana* is found within the control region rather than at the canonical position between the ND5 and Cyt*b* genes. In control region, we identified one termination-associated sequence (TAS), one central conserved sequence block (CSB-D) and two conserved sequence blocks (CSB-1,CSB-2). We also found a 121 bp tandem repeat sequence with eight complete repeats and one truncated repeat. Phylogenetic analysis based on mitochondrial gene places *M. americana* within Moronidae and confirms its close relationship with yellow perch (*M. mississippiensis*).

(2) In total, 96 samples were collected from US eastern coast. Three sampling sites were in Delaware Bay (Sta92、Bro_R、Mur_R) and one sampling site(Bro_Cr) was in Chesapeake Bay. We applied three mitochondrial genes (16S,

COI, ND2) to evaluate the genetic structure of four populations. The preliminary results showed that single mitochondrial gene is not polymorphic enough to resolve the population structure. Then we concatenated three partial genes to obtain a new sequence. This sequence was used to calculate the corresponding genetic parameters. The haplotype diversity for four populations ranged from 0.507 to 0.902, the nucleotide diversity ranged from 0.00034 to 0.00085. A high haplotype diversity (h) and a low nucleotide diversity (π) were shown. The pair-wise Fst among populations were all significant except the Fst between Mur_R and Bro_R. The pair-wise Kimura-2p genetic distances among populations were low. Analysis of molecular variance (AMOVA) showed that the majority variance is within populations. In general, the mitochondrial data indicated that there are significant genetic differentiations among populations. However, the differentiations were not high enough to resolve the genetic structure of four populations. Salinity is a possible reason to cause the differentiation, but the short period after glaciations was not long enough to attain a clear population structure.

(3) ddRAD technique was utilized to analyze the genetic structure for the four populations. We chose EcoRI and NlaIII as the restriction enzymes to construct the ddRAD library. In total, 438683 loci were identified from 96 individuals. After a stringent SNP filter, 696 loci were left. Subsequent Hardy-weiberg test and Linkage Disequilibrium test exclude another 22 loci. Then a selection detection test divided the remaining loci into three categories: neutral loci (579), loci under natural selection (32), loci under banlancing selection (63). We used these three categories of loci to evaluate population structure, separately. The results showed that except the loci under banlancing selection, both the other two categories of loci can obtain a similar genetic structure for four populations. Moreover, the genetic differentiations calculated from loci under positive selection were the highest. The results based on neutral loci indicated that four populations can be identified as three groups, Mur_R and Bro_R clustered into one group, the other two populations formed another two groups, respectively. The genetic distance results revealed that Sta92 was closer to Bro_Cr than to the other two populations. This population structure was probably caused by salinity and anthropic activities. The demography for four populations was evaluated based on Bayesian method. The results showed that a hypothetical origin was first divided into two clades, then two clades divided into the four populations.

Sta92 and Bro_Cr formed one clade, Mur_R and Bro_R formed another clade. This hypothetical origin was probably a southern shelter during glacial period.

Key Words: *Morone americana*； mitochondrial genome； mitochondrial gene； restriction site-associated DNA sequencing； population genetics