

# Analysis of Population Genetic Diversity of *Cynoglossus cynoglossus* Based on 2b-RAD Simplified Genome Sequencing

Lei Jia, Bo Zhang\*, Kefeng Liu, Debin Zheng, Xiaoyu Wang

Tianjin Bohai Sea Fisheries Research Institute, Bohai Sea Fisheries Research Center of Chinese Academy of Fishery Sciences, Tianjin  
Email: tianjinbohaisuo@163.com, \*zb611273@163.com

Received: Oct. 21<sup>st</sup>, 2017; accepted: Nov. 3<sup>rd</sup>, 2017; published: Nov. 16<sup>th</sup>, 2017

## Abstract

Taking the cultured and wild populations of *Cynoglossus semilaevis* in Tianjin coast and cultured populations in Haiyang as study objects, there were 32,746 SNP loci screened by 2b-RAD sequencing. The genetic relationship, genetic structure, genetic diversity in these three groups were also analyzed by SNP markers. The genetic variation within populations (Fis), the genetic variation between groups (Fst), and total variation (Fit) in each SNP locus in these three groups were estimated, which shown that Fst value was ranged from 0.0731 - 0.1635. The genetic differentiation between Tianjin culture population and Tianjin wild population is the smallest, and the genetic differentiation between Tianjin and Haiyang culture populations, Tianjin wild population and Haiyang culture population is great. The results of selective sweep showed that the polymorphism of wild population was highest in Tianjin, followed by Tianjin culture population, and lowest in Haiyang breeding population. The polymorphism information content (PIC) statistics of Haiyang, Tianjin wild and Tianjin cultured populations respectively in each SNP site shown that Tianjin wild was up to 0.220, followed by Haiyang cultured and Tianjin breeding; the lowest is 0.175. The differentiation between Tianjin wild and Tianjin culture populations is not significant. Principal component analysis was carried out by using software GCTA, and PCA showed that the contribution rates of the first two principal components PC1 and PC2 were 16.25% and 9.54%, respectively, and the cumulative contribution was 25.79%. Two dimensional clustering results show that a small part of the region in Tianjin wild population had a small intervals with that of Tianjin cultured population, but the two principal components generally can be clearly divided into two categories, and these two populations had a clear boundaries with Haiyang cultured one. Based on the above SNP locus diversity, polymorphism information content, genetic diversity comparison, the genetic diversity of Tianjin wild population is higher than Haiyang farming groups, but not significant difference between wild and cultured populations in Tianjin. It revealed that Tianjin populations may have a certain extent mixture, which means the possibility of population mixture between wild and culture individuals may exist.

## Keywords

*Cynoglossus semilaevis*, 2b-RAD, SNP, Genetic Diversity

\*通讯作者。

# 基于2b-RAD简化基因组测序的半滑舌鳎群体遗传多样性分析

贾磊, 张博\*, 刘克奉, 郑德斌, 汪笑宇

中国水产科学研究院渤海水产研究中心, 天津渤海水产研究所, 天津  
Email: tianjinbohaisuo@163.com, \*zb611273@163.com

收稿日期: 2017年10月21日; 录用日期: 2017年11月3日; 发布日期: 2017年11月16日

## 摘要

以天津野生、天津养殖和海阳养殖群体三个半滑舌鳎群体为材料, 利用2b-RAD测序筛选得到的32,746个SNP位点, 利用SNP标记, 进行群体遗传结构、遗传多样性和类群间遗传关系分析。每个SNP位点在所有群体内遗传变异(Fis), 群体间的遗传变异(Fst), 总变异(Fit)以及综合多位点的Fis, Fst和Fit估计值, 表明3个群体两两群体间的Fst值从0.0731~0.1635不等。天津养殖群体和天津野生群体的遗传分化最小, 天津养殖群体和海阳养殖群体以及天津野生群体和海阳养殖群体的遗传分化大。选择消除分析结果显示3个群体的 $\theta\pi$ 在0.118~0.206之间, Tajima'D值在0.448~1.457之间: 天津野生群体种群多态程度最高, 天津养殖群体次之, 海阳养殖群体最低。分别统计海阳、天津、天津野生三个群体中每个SNP位点的多态信息含量(PIC), 天津野生最高为0.220, 天津养殖次之, 海阳养殖最低, 为0.175, 天津野生和天津养殖差异并不显著。利用软件GCTA进行主成分分析, PCA结果显示, 前两个主成分PC1和PC2贡献率分别为16.25%和9.54%, 累积贡献25.79%。二维聚类结果显示, 天津野生少部分区域与天津养殖间隔较小, 但是两个主成分大体上还是可以明显区分为两大类, 且二者与海阳养殖群体分类界限明显。综合以上SNP位点多样性、多态性信息含量、群体遗传多样性比较结果, 所取的半滑舌鳎天津群体较海阳群体具有较高的遗传多样性和种群多态程度, 其中天津野生群体和天津养殖群体虽较海阳养殖群体有差异, 但前两者之间差异并不显著, 揭示天津野生和养殖群体可能有一定程度的混杂, 不排除野生群体中混杂有经养殖放流的养殖个体的可能性。

## 关键词

半滑舌鳎, 2b-RAD, SNP, 遗传多样性

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

针对鱼类养殖业发展现状和可持续发展要求, 通过开展鱼类生长和抗性等主要经济性状的功能基因组研究, 高通量地筛选进行分子设计育种的关键基因和分子标记, 建立分子标记辅助育种的技术是近几年鱼类分子育种研究领域的焦点和热点。随着生物技术的发展, RPAD、RFLP、SSR 和 SNP 等不同分子

标记手段被用来辅助分子育种, 开发标记的手段也在进一步发展。全基因组测序的完成带来了海量的可利用的分子标记, 利用这些分子标记对目标研究群体多样性进行分析、群体结构分析、亲缘关系以及连锁不平衡程度估计, 掌握物种的群体遗传结构、遗传多样性和种群间遗传关系, 能使育种思路更加清晰, 育种目标更易实现[1]。

2b-RAD 技术是指基于 IIB 型限制性核酸内切酶的简化基因组测序技术。通过对 IIB 型内切酶酶切基因组产生的等长 tag 进行高通量测序, 可以大幅降低基因组的复杂度, 同时不受有无参考基因组的限制, 快速进行全基因组范围内大规模 SNP 标记的开发与分型。2b-RAD 技术获得的标记数目更多、分型准确率更高, 可用于群体遗传多样性分析、群体进化研究、辅助基因组的组装、全基因组选择育种等领域[2] [3]。本研究利用该技术在半滑舌鳎的三个群体中筛选到 32,746 个分型结果, 通过群体遗传多样性分析和 SNP 多态性等分析, 揭示了半滑舌鳎天津群体较海阳群体具有较高的遗传多样性和种群多态程度, 其中天津野生群体和天津养殖群体虽较海阳养殖群体有差异, 但是差异不显著, 不排除野生群体中混杂有经养殖放流的养殖个体的可能性。

## 2. 材料与方法

### 2.1. 实验材料

半滑舌鳎 3 个群体, 天津自然群体取于 2015 年 10 月, 共 10 尾, 体长 12.3~14.6 cm, 山东海阳(HY)养殖群体取于 2015 年 10 月, 共 10 尾, 体长 12.6~15.1 cm。天津养殖群体取于 2015 年 9 月, 共 10 尾, 体长 11.2~15.2 cm。所有样品均为活体, 充氧加冰运送至实验室, 立即存于-80°C 冰箱保存。

### 2.2. DNA 提取与基因型检测

DNA 提取。采用 TIANamp Marine Animals DNA Kit 提取半滑舌鳎雌核发育鱼苗基因组 DNA, DNA 产物置于-20°C 保存备用, 通过琼脂糖检测和 Nanodrop2000 进行质量和浓度检测。简并基因组测定由上海欧易生物技术有限公司 Illumina Hiseq Xten 平台完成, 基因分型与质量控制由本实验室完成。

### 2.3. 测序及测序质量分析

利用 2b-RAD 技术构建舌鳎 30 个个体的标签测序文库, 30 个样品在 HiSeq X-Ten 平台进行 Paired-end 测序, 采用标准型 NNN 接头建库。将原始 reads 按照以下条件进行过滤:

剔除不含有 BsaXI 酶切识别位点的序列, 剔除低质量序列(低质量序列定义: 大于 10 个碱基的质量分数小于 20), 剔除有 10 个以上连续相同碱基的序列, 将原始 reads 进行过滤, 结果表 1 可见:

30 个测序文库中含有酶切位点的高质量 reads 与测序原始 reads 之比平均在 75%以上, 标签深度在 14~25 之间, 平均深度是 20, 表明本次舌鳎文库的测序质量较好。

### 2.4. 全基因组范围 SNP 筛查和分型分析

对于有参照基因组的标记分型, 依据 RAD-typing 的分型策略(Fu 等, 2013), 标记分型如下构建参考序列: 从舌鳎参考基因组序列中提取包含 BsaXI 酶切位点的标签作为参考序列。

半滑舌鳎(*Cynoglossus semilaevis*)参考基因组:

<https://www.ncbi.nlm.nih.gov/genome/?term=Cynoglossus+semilaevis> 个体 SNP 标记分型: 将个体的高质量 reads 利用 SOAP 软件(Li 等, 2009) (参数设置为: -M 4-v 2-r 0) mapping 到参考序列上。利用最大似然法 ML 进行位点的分型。为了保证 SNP 位点分型的准确性和严谨性, 进行以下条件的过滤: 在个体内标签深度在 500 以上的不进行分型。只留取标签内最多有 3 个 SNP 的标签位点; 标记在 80%的个体中可以

**Table 1.** Information of sequencing library  
**表 1.** 文库构建测序表

样品编号	原始 reads 数	高质量 reads 数	百分比	unique 标签数	标签深度
海阳-1	6,396,349	5,054,386	79.02%	143,337	20
海阳-2	6,396,349	5,587,054	87.35%	145,862	21
海阳-3	6,396,349	5,798,571	90.65%	145,740	22
海阳-4	6,396,349	5,545,037	86.69%	146,059	21
海阳-5	6,396,349	4,045,274	63.24%	144,101	16
海阳-6	7,065,887	5,445,752	77.07%	144,571	21
海阳-7	7,065,887	5,972,173	84.52%	144,777	23
海阳-8	7,065,887	6,462,181	91.46%	145,127	25
海阳-9	7,065,887	6,444,063	91.20%	143,720	25
海阳-10	7,065,887	4,748,605	67.20%	143,837	19
天津-1	6,646,800	4,688,066	70.53%	145,915	18
天津-2	6,646,800	5,804,242	87.32%	146,395	21
天津-3	6,646,800	5,728,174	86.18%	145,486	21
天津-4	6,646,800	5,573,184	83.85%	146,024	21
天津-5	6,646,800	4,244,575	63.86%	145,211	16
天津-6	7,027,055	5,482,653	78.02%	145,998	21
天津-7	7,027,055	6,021,305	85.69%	146,142	22
天津-8	7,027,055	6,082,922	86.56%	145,434	23
天津-9	7,027,055	5,766,252	82.06%	146,083	22
天津-10	7,027,055	4,655,306	66.25%	146,354	18
天津野生-1	5,866,332	4,209,487	71.76%	143,403	16
天津野生-2	5,866,332	5,149,297	87.78%	143,053	20
天津野生-3	5,866,332	5,138,678	87.60%	142,469	20
天津野生-4	5,866,332	4,993,461	85.12%	143,425	19
天津野生-5	5,866,332	3,645,928	62.15%	142,983	14
天津野生-6	7,404,738	4,473,240	60.41%	145,430	16
天津野生-7	7,404,738	5,977,990	80.73%	143,673	22
天津野生-8	7,404,738	5,375,319	72.59%	143,216	20
天津野生-9	7,404,738	6,021,680	81.32%	143,343	22
天津野生-10	7,404,738	4,184,500	56.51%	143,167	15

分型。最小等位基因频率  $MAF \geq 0.05$ ；共完成 32,746 个分型结果。

## 2.5. 数据处理方法

利用 Genepop 软件(version 4.2.2)、PowerMarker 软件(version 3.25)、GCTA 软件(version 1.25.0)、treebest

(Version: 1.9.2)以及 SPSS 对测序结果进行统计分析。

### 3. 结果及分析

#### 3.1. 三个群体遗传多样性分析

根据筛选得到的 32,746 个 SNP 位点, 利用 Genepop 软件(version 4.2.2) (Rousset, 2008)统计分析每个 SNP 位点在 3 个舌鳎群体中的 SNP 位点在所有群体内遗传变异(Fis)、群体间的遗传变异(Fst)、总变异(Fit); 使用 GD (version: 0.8)软件根据分型结果计算核苷酸多样性( $\theta\pi$ )及 Tajima'D 中性检验及核苷酸不配对分析。

##### 3.1.1. 群体间的遗传结构分析

表 2 中数据显示, 3 个群体两两群体间的 Fst 值从 0.0731~0.1635 不等。天津群体和天津野生群体的遗传分化最小, 天津群体和海阳群体以及天津野生群体和海阳群体的遗传分化大。遗传分化系数(Fst)值小于 0.05, 表明变异主要来自于群体内, 且群体间无分化。其大于 0.05, 表明遗传变异大多发生在群体间, 但群体间呈低度分化。

##### 3.1.2. 选择消除分析

Tajima' D 是表征 DNA 序列在进化过程中是否遵循中性进化模型的指标。在标准中性进化模型下, Tajima's D 的理论值为零, 如果实际的 Tajima's D 值明显偏离零, 表明实际的等位基因频率相对于中性进化模型的期望存在偏倚。表 3 中, Tajima's D 中性检验结果海阳群体较高, 表明这个群体偏离了中性模式, 可能受到群体扩张和自然选择的作用。天津及天津野生有不同程度的偏离零, 表明相对于中性进化模式的期望存在偏倚。

而  $\theta\pi$  是衡量种群多态程度的一个重要指标, 表 3 中,  $\theta\pi$  结果显示: 天津野生群体种群多态程度最高, 天津养殖群体次之, 海阳养殖群体最低。

#### 3.2. SNP 多态性分析

根据筛选得到的 32,746 个 SNP 位点, 利用 PowerMarker 软件(version 3.25)分别统计海阳-、天津、天津野生三个群体中每个 SNP 位点的多态信息含量(PIC), 观测杂合度(Ho)和期望杂合度(He)如表 4。

分别统计海阳、天津、天津野生三个群体中每个 SNP 位点的多态信息含量(PIC), 天津野生最高为 0.220, 天津养殖次之, 海阳养殖最低, 为 0.175, 天津野生和天津养殖差异并不显著。

**Table 2.** The value of Fst of three groups in *Cynoglossus semilaevis*

**表 2.** 半滑舌鳎 3 个群体间的 Fst 值

群体	海阳	天津
天津	0.1635	
天津野生	0.1778	0.0731

**Table 3.** The value of  $\theta\pi$  and Tajima' D of three groups in *Cynoglossus semilaevis*

**表 3.** 半滑舌鳎 3 个群体的  $\theta\pi$  以及 Tajima' D 值

群体名称	海阳	天津	天津野生
$\theta\pi$	0.118343	0.185802	0.206763
Tajima' D	1.457026	0.448594	0.59124



**Table 4.** PIC of SNP of three groups in *Cynoglossus semilaevis*  
**表 4.** SNP 多态性分析统计表

群体		期望杂合度(He)	观测杂合度(Ho)	多态信息含量(PIC)
海阳-	范围	0~0.5	0~1	0~0.375
	平均值	0.2212	0.3115	0.1752
天津	范围	0~0.5	0~1	0~0.375
	平均值	0.2531	0.2877	0.2066
天津野生	范围	0~0.5	0~1	0~0.375
	平均值	0.2473	0.2805	0.220

### 3.3. 主成分分析(Principle Component Analysis, PCA)

利用软件 GCTA (version 1.25.0)进行主成分分析。根据前两个主成分将各个个体描绘成散点图,一般的,通过前两个主成分能够将个体进行聚类。由于图中散点密集,加上样品名会有部分样品名重叠,结果详见图 1。

#### 分析结果:

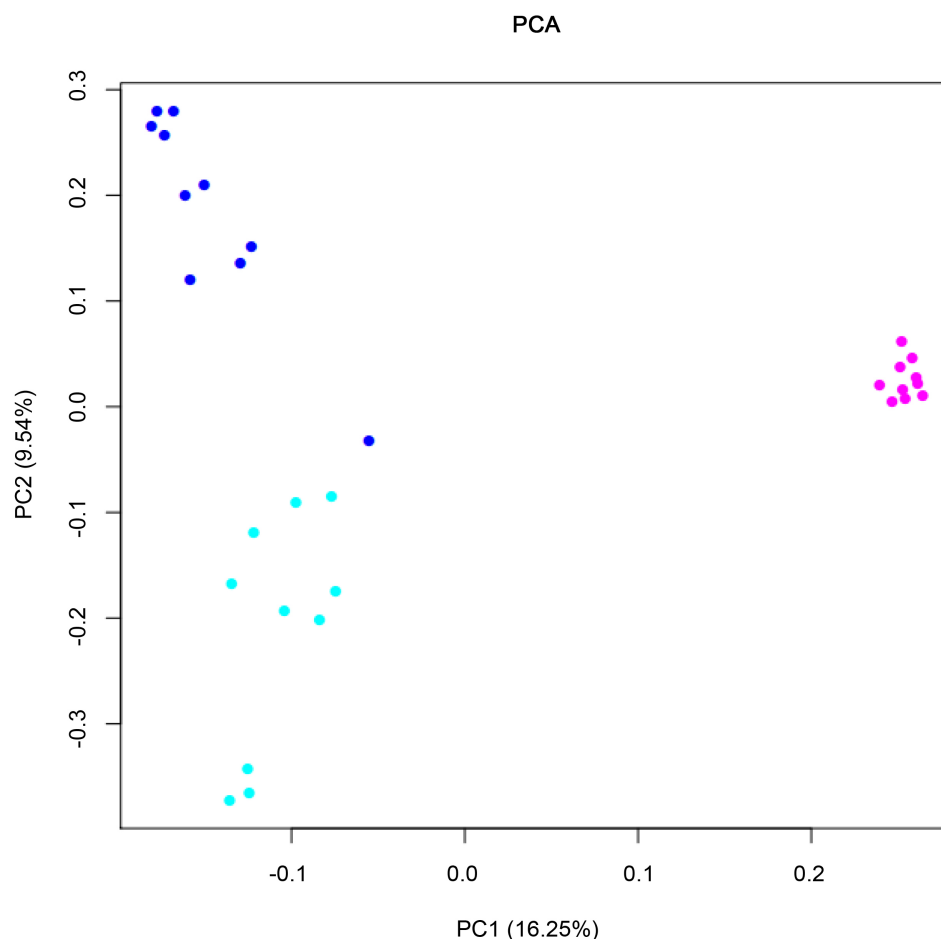
利用软件 GCTA 进行主成分分析,PCA 结果显示,前两个主成分 PC1 和 PC2 贡献率分别为 16.25% 和 9.54%,累积贡献 25.79%。二维聚类结果显示,天津野生少部分区域与天津养殖间隔较小,但是两个主成分大体上还是可以明显区分为两大类,且二者与海阳养殖群体分类界限明显。

### 3.4. 基于个体共享标签序列的系统发生树构建

选取在 30 个个体中共有的 SNP,如果缺失相应的位点,则用“-”代替。将获得的序列利用 treebest (Version: 1.9.2),最大似然法构建 NJ (Neighbor-joining)树,树的可靠性通过 bootstrap 法进行检验(重复 1000 次);其中,图 2 中内部节点处的数值代表 bootstrap 值的百分比,叶子节点处为分析对象名称。海阳-1——海阳-10 对应于 Q1——Q10;天津-1——天津-10 对应于 T1——T10;天津野生-1——天津野生-10 对应于 T\_W1——T\_W10

## 4. 讨论

2b-RAD 技术利用 IIB 型限制性内切酶,通过基因组酶切产生等长的 33~36 bp 的酶切标签,经过富集后用于下游的高通量测序反应,通过生物信息学分析实现全基因组范围高通量 SNP 筛查和分型分析,批量的 SNP 位点可以为群体遗传多样性分析奠定良好的基础[4] [5]。本研究中利用 SNP 分型来进行三个半滑舌鳎群体的遗传多样性分析,通过群体间遗传多样性分析、选择消除分析、SNP 位点遗传多样性分析、主成分分析等发现天津群体和天津野生群体的遗传分化最小,天津群体和海阳群体以及天津野生群体和海阳群体的遗传分化大。地理隔离造成的遗传多样性具有一定的作用。SNP 位点多样性、多态性信息含量、群体遗传多样性比较结果,所取的半滑舌鳎天津群体较海阳群体具有较高的遗传多样性和种群多态程度,其中天津野生群体和天津养殖群体虽较海阳养殖群体有差异,但前两者之间差异并不显著,揭示天津野生和养殖群体可能有一定程度的混杂。武菲[6]等通过对 3 个野生群体(北屯、乌伦古湖、博斯腾湖)和 2 个养殖群体(五家渠、北湖)进行线粒体 DNA 控制区基因序列测定,比较了野生群体和人工养殖群体分子遗传多样性差异,发现河鲈养殖群体的核苷酸多样性( $0.0015 \pm 0.00075$ )低于野生群体( $0.0023 \pm 0.0011$ )。分析其可能的原因,是由于长时间的单一亲代繁殖、封闭群体的亲本数量的限制、遗传漂变及养殖鱼类对养殖环境的适应等因素,加速养殖群体种质同质化引起的。毛寿康等[7]利用 14 对微卫星引物,对采捕于辽宁葫芦岛、山东青岛、江苏连云港以及浙江舟山附近海域的梭鱼(*Liza haematocheila*) 4 个野生

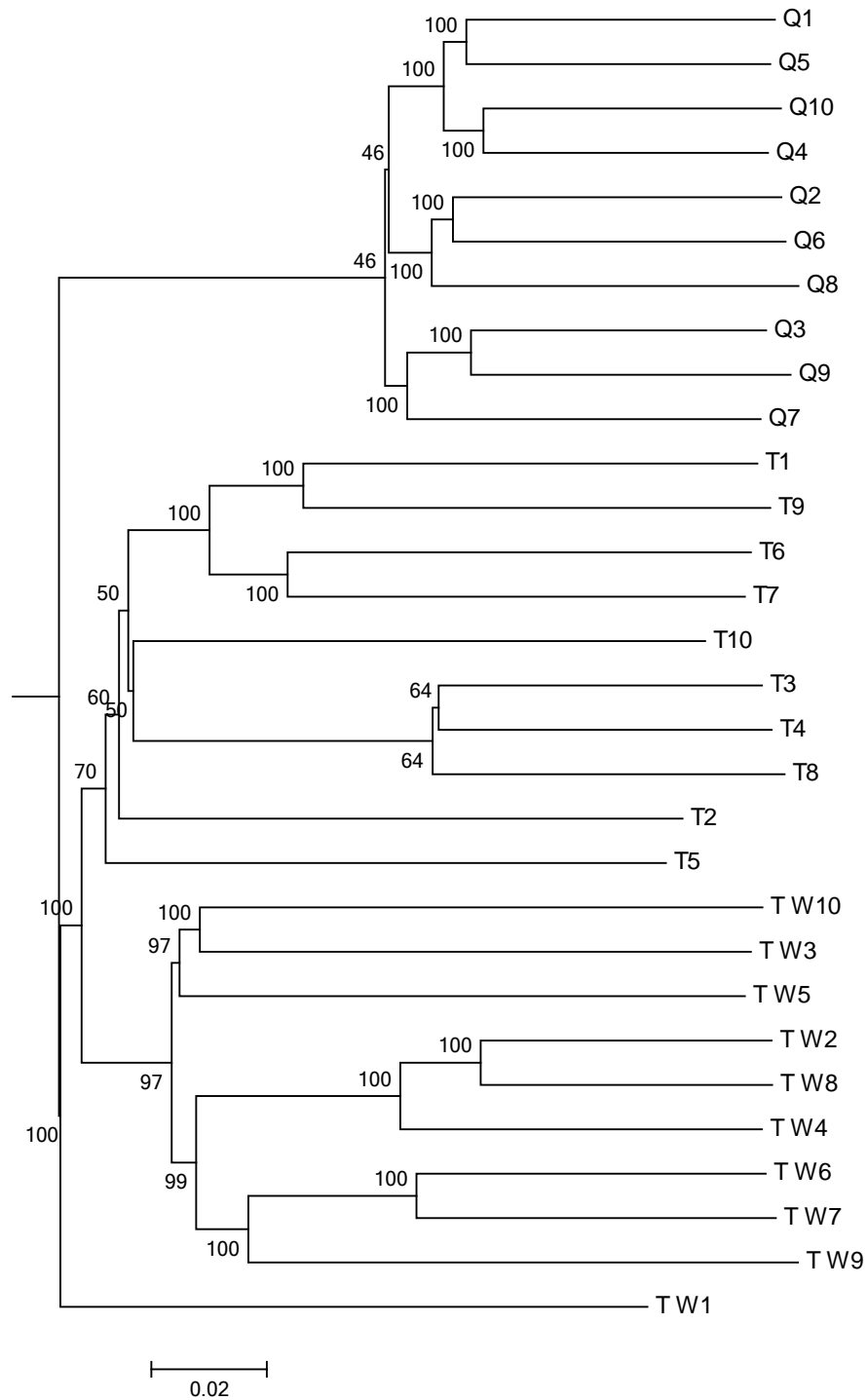


**Figure 1.** PCA of three groups in *Cynoglossus semilaevis*. Dark blue points present wild population in Tianjin, sky blue points present cultured population in Tianjin and pink points present Haiyang population in Shandong

**图 1.** 主成分分析散点图。图中深蓝色圆点表示天津野生群体；浅蓝色圆点表示天津群体；粉红色为海阳-群体

地理群体进行群体遗传多样性分析，发现 4 个群体遗传多样性处于中等多态水平，遗传多样性差异无统计学意义( $P > 0.05$ )，各群体大部分微卫星位点偏离 Hardy-Weinberg 平衡( $P < 0.05$ )。群体间出现一定程度基因流。青岛和葫芦岛的遗传距离最近(0.202)，而连云港和舟山的最远(0.473)，这可能与梭鱼幼体的扩散能力及近海沿岸生态环境及群落结构有关。

本研究三个群体的选择，天津养殖和青岛养殖群体的选择旨在考察不同养殖群体的遗传多样性差异和遗传分化和遗传距离，而天津养殖和天津野生群体的选择旨在考察同一地理条件下，野生群体和养殖群体遗传多样性的差异和分化情况，事实证明，这种差异并不显著，因为各地增殖放流活动的实施，使得养殖群体可能部分放流后混入到野生群体中，造成二者在统计学上无显著差别。对半滑舌鳎种质资源保护及开发利用的几点建议：1) 野生群体越来越少，遗传多样性降低，应该进行产卵场及保育场的保护，维持天然的种群结构的同时，应加大放流力度；2) 由于半滑舌鳎野生群体存在偏离 Hardy-Weinberg 平衡，故在对沿岸海域进行人工增殖放流时，要注意其遗传结构，避免人为因素加快半滑舌鳎天然种质资源的衰竭；3) 对放流群体亲本要进行遗传多样性的分析，避免近亲繁殖，避免放流群体遗传结构过于简单，对自然群体结构造成干扰。



**Figure 2.** The consensus phylogenetic relationship of *C. semilaevis* of three groups from Neighbor-Joining (NJ) analyses. The numbers on the branches are bootstrap values for NJ  
**图 2.** NJ 法构建半滑舌鲷三个群体进化关系。节点数字代表 NJ 构建法的 bootstrap 值

### 基金项目

现代农业产业技术体系专项经费资助(CARS-47-Z1); 天津现代产业技术体系-水产-海水鱼养殖岗位 ITTFRS2017011; 天津市农业科技成果转化与推广项目(201604090)。



## 参考文献 (References)

- [1] 张文静, 余育, 沈韞芬. 微卫星 DNA 遗传分析在原生动植物学中的研究进展[J]. 水生生物学报, 2003, 27(2): 185-190.
- [2] 陆丽君, 马爱军, 王新安, 等. 5 个红鳍东方鲀养殖群体微卫星 DNA 遗传多态性分析[J]. 渔业科学进展, 2013, 34(4): 27-34.
- [3] Delphine, V.I., Albrecht, E.M., Claude, L., *et al.* (2010) Population Structure and Genetic Diversity in a Commercial Maize Breeding Program Assessed with SSR and SNP Markers. *Theoretical and Applied Genetics*, **120**, 1289-1299. <https://doi.org/10.1007/s00122-009-1256-2>
- [4] Fu, X., Dou, J., Mao, J., *et al.* (2013) RADtyping: An Integrated Package for Accurate De Novo Codominant and Dominant RAD Genotyping in Mapping Populations. *PLoS ONE*, **8**, e79960. <https://doi.org/10.1371/journal.pone.0079960>
- [5] Gerdes, J.T., Behr, C.F., Coors, J.G., *et al.* (1994) Compilation of North American Maize Breeding Germplasm. CSSA, Madison.
- [6] 武菲, 胡文革, 王翠华, 等. 河鲈养殖与野生群体遗传多样性比较分析[J]. 水生生物学报, 2016, 40(1): 182-188
- [7] 毛守康, 马爱军, 丁福红, 等. 梭鱼(*Liza haematocheila*) 4 个野生地理群体遗传多样性的微卫星分析[J]. 渔业科学进展, 2016, 37(2): 67-75.

### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2373-1443, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [ojfr@hanspub.org](mailto:ojfr@hanspub.org)