

文章编号: 1674-5566(2024)01-0124-11

DOI: 10.12024/jsou.20230404153

基于基因富集数据的中国金鱼品种谱系研究

周静彤^{1,2}, 何 为^{1,2}, 李晨虹^{1,2}, 张景航^{1,2}, 潘伊凌³

(1. 上海海洋大学 环境DNA技术与水生态健康评估工程中心, 上海 201306; 2. 上海海洋大学 海洋动物系统分类与进化上海高校重点实验室, 上海 201306; 3. 上海自然博物馆(上海科技馆分馆), 上海 200041)

摘要: 金鱼(*Carassius auratus*)起源于中国,是由野生鲫鱼驯化而来的一种观赏性鱼类。金鱼不同品种间存在巨大外形差异,长期的杂交与人工选择产生了众多的品种,但也导致其谱系关系混乱复杂,划分品种的难度加大。人们尝试了许多方法明确金鱼各品种间的谱系历史与杂交关系,但是至今未取得进展。本研究收集了11个中国金鱼品种,共计59个个体,利用跨物种靶基因富集和高通量测序获取了各品种目标基因序列。以斑马鱼4434个单拷贝基因位点为参考序列,将金鱼对应的基因按照同源四倍体亚基因组分组,富集的序列分别按照两个亚基因组进行比对分型,产生35298个单核苷酸多态性(SNPs)位点数据,用于金鱼品种谱系历史和杂交关系的研究。通过主成分分析,品种遗传结构以及系统发育网状图的构建,揭示了中国部分金鱼品种的遗传关系,解释了它们的遗传演化过程:(1)龙睛、琉金两种文种金鱼与和金(文种金鱼)关系较近;珍珠鳞、狮头不是来自于和金,而是独立演化而来。(2)以兰寿为代表的蛋种金鱼虽然早于水泡眼出现,但水泡眼是文种金鱼先出现水泡眼特征,然后文种水泡背鳍消失,演化出现代金鱼中的主流水泡眼。这也解释了蛋种金鱼中,水泡眼出现文种的概率比较大的现象。(3)龙睛鹤顶红与狮头关系密切,先产生了头瘤的性状,而不是从龙睛演化而来,龙睛的性状是在头瘤性状出现之后产生。(4)珍珠鳞、皇冠珍珠、龙睛皇冠珍珠在遗传上关系相近,皇冠珍珠是由珍珠鳞直接演化而来,先出现珍珠鳞而后出现皇冠特征。龙睛皇冠珍珠是最晚演化而来,从龙睛中获得基因,并且可认为是由龙睛与皇冠珍珠杂交而来。研究结果将帮助了解中国金鱼的演化历史,对于金鱼品种的分类和品种培育起到一定的借鉴作用。

关键词: 金鱼; 龙睛皇冠珍珠; 狮头; 琉金; 单拷贝保守核编码序列; 基因富集; 谱系研究

中图分类号: S 965.8 **文献标志码:** A

金鱼(*Carassius auratus*)属于脊索动物门(Chordata)硬骨鱼纲(Osteichthyes)鲤形目(Cypriniformes)鲤科(Cyprinidae)鲫属(*Carassius*)。因其丰富的颜色与外部形态,成为一种观赏性鱼类,受到人们的喜爱,至今人们已经培育出了上百种金鱼品种^[1-2]。WANG等^[3]分析了234条金鱼和大量本地标本的部分控制区(CR)和整个细胞色素b(Cytb)的核苷酸序列,确定了金鱼是由鲫鱼驯化而来,可能起源于中国南方长江中下游的鲫鱼。基于线粒体DNA(mtDNA)序列研究结果显示日本金鱼也是起源于中国鲫鱼,与日本鲫鱼无关^[4]。金鱼最早记载于中国晋代,一般认为“金鲫”是金鱼的始祖^[5-6]。

起初金鱼只有体色发生变化,明朝后半期形成了双尾、长鳍、短体等表型的金鱼,在进入清代之后,才出现了眼睛变异,头瘤,透明鳞,背鳍缺失等其他表型的金鱼品种^[5-6]。金鱼于1502年最早传入日本,1596年出现了琉金^[7-8]。18世纪左右出现蛋种金鱼,传入日本后培育出兰寿^[7-8]。在中国则独立演化成虎头、蛋凤等品种。在17世纪末,金鱼传入英国,18世纪传入欧洲,1874年传入美国,之后在世界广泛传播^[6]。

金鱼的分类目前仍没有形成公认的分类系统。主要的分类依据包括外部形态、颜色等特征。其中外部形态差异主要包括:背鳍的有无;体长、体高、体宽的比例;头瘤的有无和形态;眼

收稿日期: 2023-04-11 修回日期: 2023-05-22

基金项目: 上海科技馆2020-2022年MicroCT成像效果相关基础研究项目(SSTM/SOPZY-03-R2KZ-20200300001)

作者简介: 周静彤(1998—),女,硕士研究生,研究方向为海洋科学。E-mail: 1165409283@qq.com

通信作者: 何 为, E-mail: whe@shou.edu.cn

版权所有 ©《上海海洋大学学报》编辑部(CC BY-NC-ND 4.0)

Copyright © Editorial Office of Journal of Shanghai Ocean University (CC BY-NC-ND 4.0)

<http://www.shhydx.com>

的形态变异;鼻膜的形态变异;鳞片的形态变异以及尾鳍的形态变异等^[9]。分类和命名目前大多为民间常用的形式,缺乏统一的分类和命名规则^[10]。民间将有背鳍的金鱼称为文种金鱼,而失去背鳍的金鱼则称为蛋种金鱼。常见的文种金

鱼有和金、琉金、珍珠鳞、狮头、龙睛等,其中珍珠鳞金鱼的鳞片呈现坚硬的半球形凸起,狮头带有头瘤性状、龙睛带有眼睛凸出性状;蛋种金鱼包括兰寿、水泡眼等,水泡眼的眼部为水泡状,兰寿有头瘤变异(图1)。

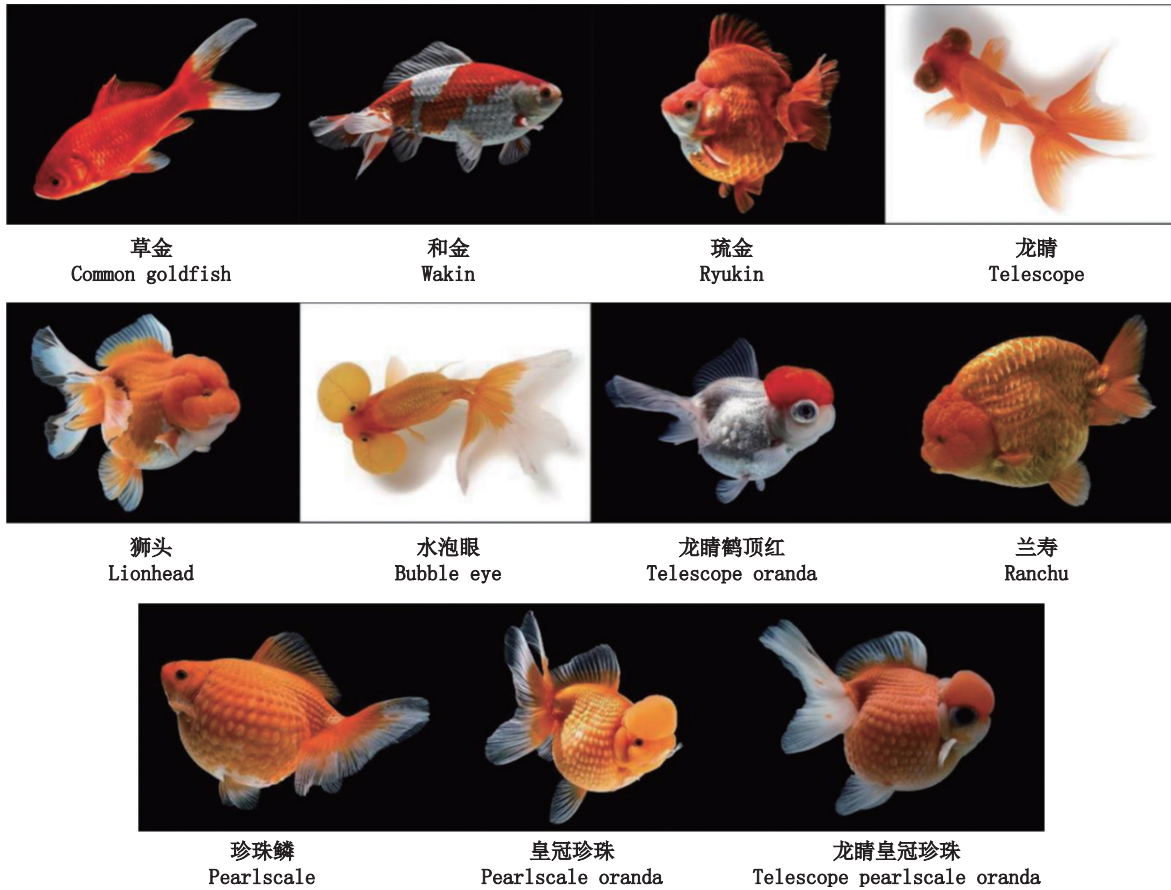


图1 用于本实验的中国金鱼品种

Fig. 1 Goldfish strains of China used in this study

长期的人工选择导致金鱼在外部形态上与鲫鱼相比出现了显著差异,并且在近几十年,人们通过大量的杂交方式培育金鱼,杂交子代在体型上与亲本相比也出现了差异,获得了大量重组性状的金鱼,这些金鱼成为了新的品种^[6, 11]。由于金鱼品种的增加,谱系关系混乱复杂,使得金鱼品种的划分难度加大,研究金鱼品种之间的谱系关系也成为了一个难题^[9, 12]。

人们通过众多的方法和遗传标记来尝试明确金鱼品种间的谱系关系与杂交来源^[6]。龙睛、琉金、珍珠鳞、狮头都属于文种金鱼,也有许多研究者对这些文种金鱼进行研究。细胞学研究表明,鲫鱼、红龙睛金鱼的染色体数目、形态都十分相似^[13]。罗莉中等^[14]对鲫鱼和红龙睛金鱼 8 种

组织器官分别进行了乳酸脱氢酶(LDH)的比较研究,并未发现鲫鱼和红龙睛金鱼之间的差异。通过 mtDNA 控制区与线粒体 CytB 进行研究无法厘清琉金、龙睛等金鱼品种之间的谱系关系^[15-17]。然而,这些研究未能确定各种文种金鱼是由草金这种鲫型金鱼分别独立培育而成,还是由草金培育成某一文种金鱼,再由这种金鱼培育成其它文种的品种。

王晓梅等^[15]利用金鱼基因组 DNA 进行 RAPD 研究,发现金鱼品种是先进化为草金,再由草金演化成文种金鱼,之后发展为蛋种金鱼,这与张利红等基于 SDS-PAGE 蛋白质标记和 AFLP 标记分析研究结果一致^[16-20]。王姝妍利用变异速率较快的线粒体控制区与受饱和效应影响较小

的CytB构建系统发育树只能明确蛋种金鱼出现在双尾鳍文种金鱼之后,但未能明确蛋种金鱼和文种金鱼品种之间的谱系关系^[9]。对此我们提出疑问:不同品种蛋种金鱼是分别如何培育而来?其中蛋种代表金鱼——水泡眼在生产培育过程中后代产生文种金鱼的概率相较于其他蛋种金鱼更高,所以针对这一现状,我们提出假设:水泡眼是先演化为蛋种金鱼再获得水泡眼性状还是先获得水泡眼性状再失去背鳍?

龙睛鹤顶红具有龙睛和头瘤两种形态特征,传统的分类认为龙睛鹤顶红的头瘤性状与鹤顶红一致,但由于其有额外的龙睛性状,并且目前研究并未详细讨论龙睛鹤顶红的演化历史,所以龙睛鹤顶红是先演化出头瘤性状还是先演化出龙睛性状有待进一步研究。

龙睛皇冠珍珠具有珍珠鳞、龙睛、头瘤3种特征;皇冠珍珠具有珍珠鳞、头瘤2种特征,龙睛鹤顶红具有龙睛、头瘤2种特征;狮头、龙睛、珍珠鳞各自只具有1种形态特征。由于这些金鱼出现的时间较晚,他们之间是否采用了杂交培育的手段,以及它们之间的谱系关系,也未见研究。

关于金鱼的研究基本上都基于线粒体基因数据。鱼类的线粒体基因是连锁的,导致了线粒体基因组实质上只是1个基因位点,又由于其具有母系遗传的特点,所以在研究金鱼谱系演化方面并不能很好的重建金鱼谱系演化历史和确定金鱼谱系间的杂交关系。因此,金鱼的谱系演化研究需要用到更多独立的核基因位点。金鱼是异源四倍体物种,比斑马鱼基因组多一轮全基因组复制(WGD)^[21-22],通常的系统发育和谱系杂交的分析方法是基于二倍体数据分析,无法按照常规的分析方法对金鱼的基因组数据进行分析,不然会出现SNP、等位基因频率等信息计算错误的现象。目前,金鱼基因组组装和分型工作取得了很好的进展,金鱼的基因组已经被完整地分别组装到两个亚基因组^[23]。因此,可以参考组装好的金鱼基因组,对金鱼的数据分别按照两个亚基因组分型,把每个亚基因组当作一个二倍体基因组分别计算SNP和等位基因频率,之后将两个亚基因组的SNP和等位基因频率数据合并,用以研究金鱼的谱系关系。

目前使用独立的核基因位点来研究金鱼问

题的有KON和CHEN。KON对27种家养金鱼和野生金鱼进行了全基因组测序,用于研究决定不同表型的分子机制,同时建立了部分金鱼品种的种群遗传结构,但使用的金鱼品种大部分为日本特有,中国金鱼品种数目较少^[23]。而CHEN等^[24]用基因组数据研究了可能导致金鱼出现和驯化有关的表型的分子机制。上述研究都未对我们提出的金鱼品种演化中的各类假设问题进行分析考证,特别是与中国的金鱼品种相关的问题。

至此,我们总结了提出的假设与疑问:1.文种金鱼是由哪种鲫型金鱼直接演化而来?是单尾的普通型草金还是双尾的和金?2.不同品种蛋种金鱼是如何分别培育而来,水泡眼是先演化为蛋种金鱼再获得水泡眼性状还是先获得水泡眼性状再失去背鳍?3.龙睛鹤顶红是先有龙睛,还是先有头瘤?4.龙睛皇冠珍珠、珍珠鳞、龙睛、皇冠珍珠之间的谱系关系如何,它们之间的杂交历史是什么样的?

由于金鱼品种数量众多,杂交关系复杂,为了回答上述的疑问,按照提出的猜想与假设,选择与假设有关、具有代表性、重组性状较少、与其他金鱼演化关系相比较为简单的金鱼来研究,由于无法收集到鹤顶红,所以选择了与鹤顶红(文种金鱼)一样具有头瘤性状的狮头代替。收集了11种中国金鱼品种,共计59个个体;通过跨物种靶基因富集,并结合异源四倍体亚基因组分型分析,获得金鱼A、B两个亚基因组内的单拷贝核苷酸序列;利用这些核苷酸序列开展金鱼品种间遗传结构研究、主成分分析,构建系统发育网状图,检验提出的关于金鱼品种间谱系演化的假设。期望研究结果能够帮助了解金鱼的演化历史,对于金鱼品种的分类和品种培育起到一定的参考作用。

1 材料与方法

1.1 样品收集

金鱼样品采自上海市浦东新区岚灵花鸟综合创意园市场和上海云鳞生物科技有限公司金鱼养殖场,共11个品种,计59个个体,其中包含草金4条、和金6条、琉金5条、龙睛鹤顶红5条、龙睛5条、狮头7条、兰寿7条、水泡眼5条、珍珠鳞5条、皇冠珍珠5条、龙睛皇冠珍珠5条。所有

个体剪右胸鳍固定于无水乙醇中,并在-4℃冰箱保存。

1.2 DNA 提取与建库

使用 Ezup 柱动物基因组 DNA 纯化试剂盒(生工生物工程(上海)股份有限公司,产品编号: B518251)提取 DNA,提取方法参照试剂盒使用说明书。用 NanoDrop3300 (Thermo Fisher Scientific) 荧光分光光度仪确定浓度。取 1 000 ng DNA 加蒸馏水至 270 μ L,用 Covaris M220 超声波破碎仪破碎至 250 bp。用琼脂糖凝胶电泳确定破碎后的 DNA 片段大小,取 130 μ L 用于后续建库,剩余 DNA 保存于-20℃冰箱。建库步骤参考 MEYER 等^[25]和 FISHER 等^[26]的方法,并对其部分流程操作进行改动。文库构建的主要步骤^[27]:(1)平末端修复:将破碎后的 DNA 产物用 T4 DNA 聚合酶修复 3'端、5'端,5'端磷酸基团用 T4 多核苷酸激酶连接。(2)接头连接:在第一步产物的两端使用 T4 DNA 连接酶加上 P5、P7 接头^[28]。(3)由于第二步产物接头没有磷酸化,只有模板链的一条单链存在缺口,使用 Bst 聚合酶填充产生的单链缺口并进行扩增。

1.3 DNA 富集

将不同个体的文库按等摩尔量混样,用于目标基因富集。10 个文库混为一管,每个文库为 200 ng。根据 JIANG 等^[29]开发的 4 434 个单拷贝保守核编码序列(CDS)设计带生物素的 RNA 探针。利用这些探针获取样本中的目标基因,基因富集的步骤大致如下^[27]:(1)使目标 DNA 文库与生物素化的 RNA 探针杂交。(2)通过被链霉亲和素包被的磁珠 Dynabeads M-270 对杂交文库进行清洗。(3)使用 KAPA HiFi Taq Ready Mix (2 \times)对产物进行 PCR 扩增,之后重复 1、2 步对文库进行二次富集。(4)将二次富集的产物两端增加 P7 index 与 IS4 接头并扩增。

基因富集扩增后的产物在 Azenta 公司(江苏,中国)的 Illumina HiSeq X10 平台上进行测序。用 Illumina bcl2fastq 2.20 根据每个样本的 P7 index 信息对数据按样本进行分组,获得了混合样品的富集序列。将获得的测序结果解压,将序列依照 inline index 来分组,之后去掉 inline index 接头,获得最终 trimmed 之后数据。

1.4 异源四倍体亚基因组 4 434 个单拷贝基因位点分组和 reads 序列分组

从 NCBI 数据库下载金鱼染色体水平的基因组(accession number: GCA_003368295.1)^[23],将基因组数据按照 KON 等研究结果分为两个亚基因组,命名为 G1、G2。用斑马鱼 4 434 个单拷贝保守核编码序列^[29]分别与金鱼的两个亚基因组进行比对(BLAST),blastn 运行的 e-value cut off 设为 0.000 1。比对结果有 3 种情况:(1)靶位点只在一个位置上与一条染色体比对上;(2)靶位点对比到多条染色体;(3)靶位点在一条染色体上比对到多个位置。只保留第一种情况的单拷贝保守核编码序列,按亚基因组分别合并为 fasta 格式文件,命名为 goldfish_4434_1.fas 和 goldfish_4434_2.fas。

将 trimmed 之后的富集序列转换为 fasta 格式,先用 merge 将双端测序结果连接成一条序列,再使用 seqkit rmdup(参数设置: -s)按照 fasta 的格式去掉碱基组成相同的简单重复序列^[30];然后使用 blastn(参数设置: -task blastn-query-db-outfmt 6-num_threads 30-evalue 0.001)将每个个体的富集序列与两个亚基因组的参考序列进行比对,比对结果有如下 3 种情况:(1)序列只能定位至一个亚基因组;(2)序列能定位至两个亚基因组;(3)序列不能定位与任意一个亚基因组上。将第一种情况的序列保留并归类至对应的亚基因组;将第二种情况的序列归类至相似度高的亚基因组,如果相似度一致,则同时归类于两个亚基因组;第三种情况的序列直接丢弃。将属于同一个亚基因组内的序列按照个体名称分类后再合并为 fasta 格式文件,作为后续分析备用。

1.5 单核苷酸多态性位点分型(SNP calling)

以 goldfish_4434_1.fas、goldfish_4434_2.fas 为参考序列,使用默认参数设置的 BWA-MEM (<http://cncbi.github.io/BWA-Manual-CN/>)将已经划分到两个亚基因组的序列分别映射到参考序列上。使用 GATK v4.1.8.1^[31]分别调用两个亚基因组单核苷酸多态性位点(SNPs),并对获得的单核苷酸多态性位点(SNP)进行质量过滤:使用 VCFtools(参数设置: -max-missing -remove-indels -recode -recode-INFO-all)处理输出的 vcf 格式的文件,删除缺失数据和 indels,最终获得两个亚基

因组的SNP结果。使用VCFtools附带的concat功能模块将两个亚基因组的单核苷酸多态性位点(SNP)信息进行合并,获得整个金鱼基因组的单核苷酸多态性位点(SNP)结果,命名为concat.vcf。

1.6 种群结构分析

Structure分析是基于贝叶斯模型,通过取对数的似然值计算DeltaK,以此确定品种的遗传结构^[32]。使用vcftosnps.pl脚本过滤两个亚基因组的单核苷酸多态性位点(SNP)结果,去除低覆盖率单核苷酸多态性位点(SNP),每个目标序列只选择一个单核苷酸多态性位点(SNP),以避免使用连锁不平衡的单核苷酸多态性位点(SNP)^[33]。分别将两个亚基因组的SNP数据作为STRUCTURE 2.3.4软件的输入文件^[32],设置参数MCMC重复为20万个、burnin为2万个,参数groups(K)设为2~4。每个groups(K)运行4次重复。将结果打包压缩上传至HARVEST进行最拟合groups(K)值的分析^[34]。下载HARVEST分析结果中获得的最拟合groups(K)值的indfile和popfile文件,将这两个文件复制到CLUMPP目录下。使用CLUMPP 1.1.2软件^[35],寻找最拟合groups(K)中的4次重复聚类分析的最佳结果,设置参数如下:DATATYPE即读取文件类型,设为0(表示读取indfile类型文件)和1(表示读取popfile类型文件);R,即Structure运算最拟合groups(K)时的重复运行次数,设为4;K,即Structure运算的最拟合groups(K),设为3;C按照读取文件类型,个体数indfile设置为60,群体数popfile设置为12。最终得到个体和群体的Q矩阵文件为:K3_i.outfile和K3_p.outfile。将CLUMPP得到的结果使用disstruct 1.1软件^[36]按照个体与群体对样本归属各个群体的概率进行绘图。将合并SNP数据concat.vcf重复上述分析,最终得到了G1、G2、合并亚基因组3组数据的个体与群体的品种结构分析结果。

1.7 主成分分析(PCA)

主成分分析(Principal Component Analysis, PCA)是一种线性降维算法,通过方差来衡量数据的差异性^[37]。这种方式可以减少降维后信息量的损失,保留原数据的特征与结构,更好地描述样本的特性。

本研究将两个亚基因组的SNP位点信息分别进行主成分分析。首先将含有两个亚基因组

SNP位点信息的vcf文件通过plink(<https://zzz.bwh.harvard.edu/plink/>)转化成PCA可识别的格式文件,生成eigenval、eigenvec、nosex 3种输出文件(参数:--vcf --pca --double-id --allow-extra-chr)。在eigenvec格式文件中输入中国金鱼的品种信息并将eigenvec格式文件更改为txt格式文件。将txt格式文件作为输入文件,使用R中的ggplot2包对中国金鱼的主成分分析结果进行绘图,并保存为TIF图片格式。将合并基因组的SNP信息的结果文件concat.vcf重复上述操作,最终得到了G1、G2、合并基因组3组数据的PCA分析结果,同时比较两个亚基因组的信号差异。

1.8 金鱼品种选育历史重建

Treemix是PICKRELL和PRITCHARD开发的软件^[32]。通过从多个种群中计算等位基因频率,构建种群之间的最大似然树,并推断种群间可能发生的杂交事件。Treemix使用OptM判断最优m值,即给定基因流次数范围内的相对最优值,OptM在同一m值的情况下,通过对bootstrap进行多次运算来计算参数 $\Delta m [\Delta m = L''(m)/L'(m)]$ ^[38],当 Δm 达到最大值时,该条件下的m为最优m值。

本研究使用Treemix重建所选中国金鱼品种的系统发育网状树,推断金鱼不同品种培育的历史和品种之间的谱系关系。将两个亚基因组G1、G2的SNPs结果文件进行压缩,使用VCFtools^[39]将压缩的vcf格式文件转化为tped格式,生成tped和tfam文件;修改tfam中的第一列数据,标记为中国金鱼品种信息。通过plink(<https://zzz.bwh.harvard.edu/plink/>)计算等位基因频率(参数设置:--threads 12 --tfile concat --freq --allow-extra-chr --within)并使用pigz命令将等位基因频率文件压缩。使用Treemix自带的plink2treemix.py将压缩的等位基因频率文件转换为Treemix输入文件格式。使用Treemix推断基因流(参数:-i -o -root -m -k -bootstrap -global -noss),m代表基因流数量,本实验取1~10,每个m值重复5次,root代表外群,本研究采用草金为外群。用R包OptM分析Treemix的结果,确认最佳基因流数量。载入Treemix的plotting_funcs.R脚本到R软件中,用最佳m值对应的结果文件作图并绘制网状树和残差拟合热图。将合并基因组的SNP信息的结果文件concat.vcf重复上述操作,最终得到了G1、

G2、合并基因组3组数据的系统发育网状树的结果。

2 结果

2.1 亚基因组分组和富集数据信息

2.1.1 亚基因组分组

亚基因组分组结果最终获得两个亚基因组的参考序列:goldfish_4434_1.fas 包含 2 635 个单拷贝保守核编码序列, goldfish_4434_2.fas 包含 3 014 个单拷贝保守核编码序列。后续分析将 goldfish_4434_1.fas、goldfish_4434_2.fas 作为参考基因。

2.1.2 富集数据信息

Illumina 测序共得到了 670 536 244 条原始序列,平均每管混合样品得到 111 756 041 条序列(共6管混合样品)。将混合样品中的序列按照接头序列分组后,去掉接头与低质量序列,得到 59 个样品的基因组富集序列,狮头(CL2947_4)的序列数最多,为 112 402 632;琉金(CL3122_1)的序列数最少,为 889 660。

2.2 金鱼品种遗传结构

数据包含 59 个个体,合并基因组有 2 467 个 SNP 位点用于 Structure 分析,其中亚基因组 G1 有 1 126 个,亚基因组 G2 有 1 341 个。分析结果显示,在合并基因组、亚基因组 G1、G2 中均显示 K 为 3 时 DeltaK 最大:合并基因组的 DeltaK 为 448.735 328、亚基因组 G1 的 DeltaK 为 62.810 301、亚基因组 G2 的 DeltaK 为 61.918 751,所以确定收集的中国金鱼品种的最拟合分组为 3 组。按照品种的 Q-矩阵绘图的分组结果如下:合并基因组、亚基因组 G1、亚基因组 G2 均显示龙睛鹤顶红、狮头、草金为第一组,兰寿、水泡眼、和金、琉金、龙睛为第二组,珍珠鳞、皇冠珍珠、龙睛皇冠珍珠为第三组(图 2a)。按照个体的 Q-矩阵绘图分组结果如下:合并基因组、亚基因组 G1、亚基因组 G2 均显示珍珠鳞、龙睛珍珠、皇冠珍珠、龙睛皇冠珍珠全部为第三组,龙睛鹤顶红全部属于第一组,但是狮头、草金、兰寿、水泡眼、和金、琉金、龙睛均显示部分个体与品种 Q-矩阵的绘图结果不一致(图 2b)。

2.3 主成分分析

亚基因组 G1 有 15 553 个 SNPs 用于 PCA 分析,亚基因组 G2 有 19 745 个 SNPs 用于 PCA 分

析,合并基因组有 35 298 个 SNPs 用于 PCA 分析。合并基因组的 PCA 的分析结果显示:皇冠珍珠、珍珠鳞、龙睛鹤顶红、龙睛皇冠珍珠 4 种品种的金鱼品种内的个体之间聚类紧密,和金、狮头、琉金 3 种品种的金鱼的个体聚类较散。珍珠鳞、龙睛、皇冠珍珠、龙睛皇冠珍珠、龙睛鹤顶红分布在 PC1 数值 0 的左侧,琉金、和金、草金、兰寿、狮头分布在 PC1 数值 0 右侧。珍珠鳞、皇冠珍珠、龙睛、水泡眼分布在 PC2 数值 0 的下侧,龙睛皇冠珍珠、龙睛鹤顶红、和金分布在 PC2 数值 0 的上侧。珍珠鳞、皇冠珍珠聚类较近(图 3)。

G1 亚基因组和 G2 亚基因组的 PCA 结果显示珍珠鳞、皇冠珍珠、龙睛皇冠珍珠 3 种金鱼聚在一起,与合并数据的结果一致。

2.4 系统发育网状树

根据 OptM 分析显示:合并基因组的最佳基因流数量为 2 次,亚基因组 G1 的最佳基因流数量为 4 次,亚基因组 G2 的最佳基因流数量为 6 次。合并基因组的系统发育网状树结果显示:狮头、龙睛鹤顶红为一支,和金、龙睛、琉金、兰寿、水泡眼为一支,皇冠珍珠、珍珠鳞、龙睛皇冠珍珠为一支。龙睛祖先有基因渗入到龙睛皇冠珍珠中,狮头与龙睛鹤顶红的祖先有基因流入和金中(图 4)。

3 讨论

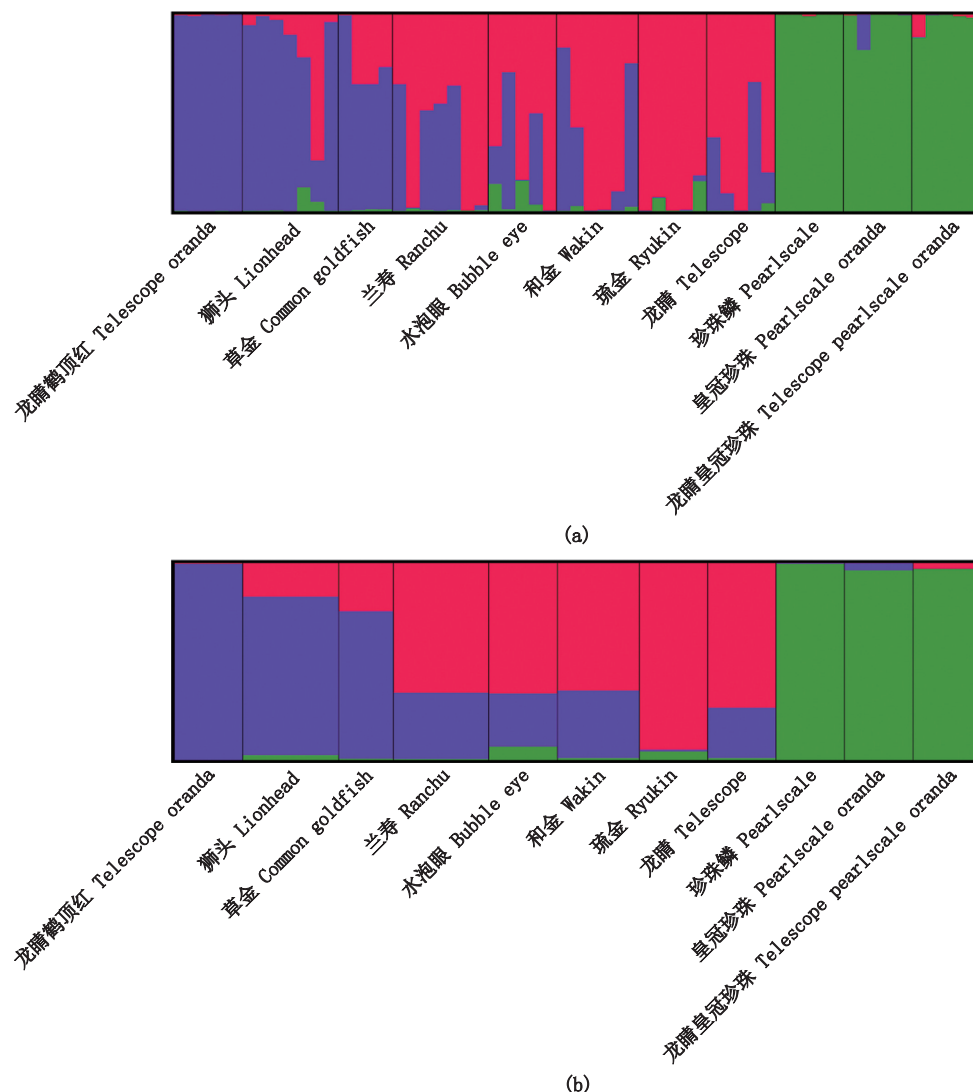
3 组数据 Structure 结果均显示文种金鱼之间的关系复杂,通过 Structure 无法推测出文种金鱼之间的演化关系。而在所有的 PCA 结果中草金、和金、琉金、狮头之间的关系混乱,但是草金、和金、琉金的分布较散且与大部分的金鱼品种重合,所以认为草金、和金、琉金属于早期的金鱼品种。合并亚基因组的 Treemix 结果表明,和金、龙睛、琉金这 3 种文种金鱼为一支,狮头与龙睛鹤顶红为一支,珍珠鳞与和金分开。结果说明龙睛和琉金与和金这种文种金鱼关系较近,这与传统的观点保持一致^[9-10]。珍珠鳞和狮头这两种文种金鱼在传统分类中与和金同属于文种金鱼^[9-10],但是研究结果显示它并不是由和金直接演化而来,推测它们应该是独立演化而来的。

由于金鱼杂交关系复杂,Structure 结果与 PCA 结果显示水泡眼、兰寿与其他文种金鱼并未

明显区分,无法推测兰寿、水泡眼的演化历史。合并亚基因组的 Treemix 结果显示兰寿与水泡眼并未聚为一支,兰寿出现的时间早于水泡眼且独立进化而来,水泡眼与和金、龙睛、琉金 3 种文种金鱼为一支且距离龙睛较近,推测水泡眼是先发生眼部变异,后失去背鳍,解释了生产实践中相较于其他蛋种金鱼,水泡眼后代出现文种的概率更大的现象^[11]。

通常从形态学来看,普遍认为龙睛、龙睛鹤顶红、狮头之间表现出部分性状的相似性,并且龙睛鹤顶红是由鹤顶红与龙睛杂交而来^[9-10],在合并基因组的 Structure 分析结果中显示龙睛鹤顶红并未与龙睛、龙睛皇冠珍珠分为一组,且基因

组成差距较大,这与两个亚基因组的 Structure 分析结果相同,所以我们推测龙睛鹤顶红与龙睛关系相较于狮头更远。合并两个亚基因组的 PCA 的分析结果显示狮头与龙睛鹤顶红之间有部分重合,未与龙睛之间重合。亚基因组 G1 与 G2 的 PCA 结果也显示相同信号,也确定龙睛鹤顶红与狮头关系比与龙睛关系近,与 Structure 分析结果一致。合并亚基因组的 Treemix 结果显示狮头与龙睛鹤顶红为一支,因为狮头与鹤顶红直接性状相似,所以进一步推测龙睛鹤顶红先产生了头瘤的性状,而不是从龙睛演化而来,龙睛的性状是在头瘤性状出现之后产生,且独立于龙睛,这与传统认为杂交获得的结果不同^[9-10]。



(a)为中国金鱼个体间遗传结构分析结果;(b)为中国金鱼品种间遗传结构分析结果。

(a) The results of genetic structure analysis results among individual goldfish in China; (b) The results of genetic structure analysis among Chinese goldfish strains.

图2 基于合并基因组的SNPs的Structure分析结果
Fig. 2 Structure analysis of SNPs based on merging genomes

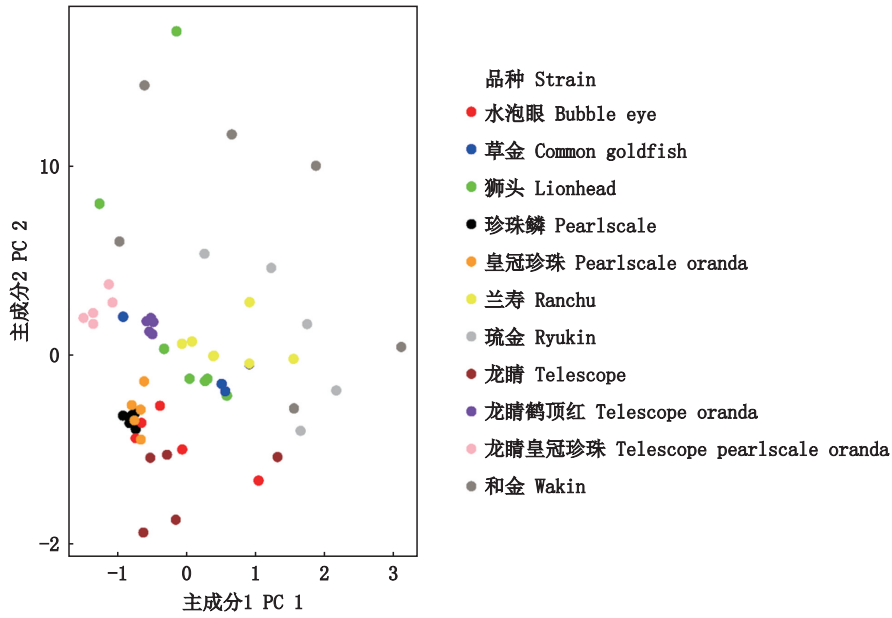


图3 合并基因组的PCA分析结果
Fig. 3 PCA analysis results of the combined genome

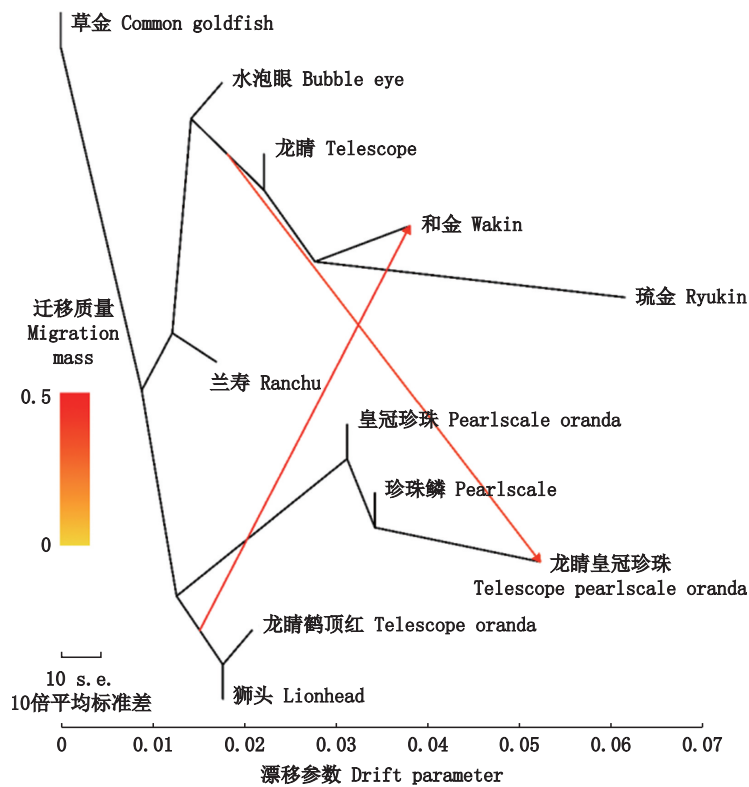


图4 基于合并基因组的SNPs的谱系关系构建结果
Fig. 4 Results of phylogenetic relationship construction of SNPs based on the combined genome

珍珠鳞、皇冠珍珠、龙睛皇冠珍珠在3种Structure分析结果中均显示为一组,说明了珍珠鳞、皇冠珍珠、龙睛皇冠珍珠在遗传上关系相近。3种PCA的分析结果也显示珍珠鳞、皇冠珍珠关系较近,与structure结果显示一致。在合并亚基因组的

Treemix里珍珠鳞、皇冠珍珠、龙睛皇冠珍珠也聚为一致,这为珍珠鳞、皇冠珍珠、龙睛皇冠珍珠在遗传上关系相近提供了进一步证据。3种PCA结果显示珍珠鳞在皇冠珍珠、龙睛皇冠珍珠这两种金鱼当中,距离皇冠珍珠最近,其次为龙睛皇冠珍珠。在

合并亚基因组的Treemix中也显示皇冠珍珠与珍珠鳞关系最近,然后是龙睛皇冠珍珠,这与PCA和Structure结果一致。龙睛皇冠珍珠从龙睛中获得基因,推测是由龙睛与皇冠珍珠杂交而来。传统形态学分类认为皇冠珍珠是由珍珠鳞演化而来,龙睛皇冠珍珠与龙睛有关与本研究的結果一致。

亚基因组G2的OptM结果显示的最佳基因交流次数有与合并亚基因组、亚基因组G1存在差异,我们认为出现这种情况的原因有两个:(1)主要发生基因交流的是G1亚基因组的基因,G2亚基因组的基因更为保守,应该同两个亚基因组的不对称进化有关^[33]。(2)Treemix分析的算法依旧存在局限性。

本研究揭示了部分中国金鱼品种的谱系关系,解释了实验金鱼品种的遗传演化过程:(1)龙睛、琉金两种文种金鱼与和金这种文种金鱼关系较近;珍珠鳞、狮头并不是由和金直接演化而来,推测应该是独立演化而来的。(2)水泡眼是文种金鱼先出现水泡眼特征,然后文种水泡背鳍消失,演化出现代金鱼中的主流水泡眼,以兰寿为代表的蛋种金鱼虽然早于水泡眼出现。(3)龙睛鹤顶红与狮头关系密切,先产生了头瘤的性状,而不是从龙睛演化而来,龙睛的性状是在头瘤性状出现之后产生。(4)珍珠鳞、皇冠珍珠、龙睛皇冠珍珠在遗传上关系相近,皇冠珍珠是由珍珠鳞直接演化而来,龙睛皇冠珍珠是最晚演化而来,龙睛皇冠珍珠从龙睛中获得基因,认为龙睛皇冠珍珠是由龙睛与皇冠珍珠杂交而来。本研究因为材料的所限,无法回答蛋种龙睛是由蛋种金鱼获得龙睛性状还是由龙睛丢失背鳍等其他问题,尚需要更多的金鱼品种来明确它们之间的谱系关系。

参考文献:

- [1] 金晓峰,张涛. 金鱼养殖现状及发展的几点建议[J]. 科学养鱼, 2019(8): 77-78.
JING X F, ZHANG T. Several suggestions for present status and development of golden fish [J]. Scientific Fish Farming, 2019(8): 77-78.
- [2] 林学明,游美珍,程惠泉. 我国金鱼养殖业现状与发展的几点建议[J]. 福建农业科技, 2004(1): 45-46.
LIN X M, YOU M Z, CHENG H Q. Several suggestions for present status and development of Chinese golden fish [J]. Fujian Agricultural Science and Technology, 2004(1): 45-46.
- [3] WANG S Y, LUO J, MURPHY R W, et al. Origin of Chinese goldfish and sequential loss of genetic diversity accompanies new breeds [J]. PLoS One, 2013, 8(3): e59571.
- [4] KOMIYAMA T, KOBAYASHI H, TATENO Y, et al. An evolutionary origin and selection process of goldfish [J]. Gene, 2009, 430(1/2): 5-11.
- [5] 殷嘉远,唐娜,彭尽晖. 金鱼品系演化与中国古典园林的关系探究[J]. 山东林业科技, 2020, 50(6): 72-77.
YIN J Y, TANG N, PENG J H. The relationship between the evolution of goldfish strains and Chinese classical gardens [J]. Journal of Shandong Forestry Science and Technology, 2020, 50(6): 72-77.
- [6] 姚红伟,吴明,石建香. 金鱼起源及遗传多样性研究进展[J]. 河北渔业, 2016(9): 53-55.
YAO H W, WU M, SHI J X. Advances in the origin and genetic diversity of goldfish [J]. Hebei Fisheries, 2016(9): 53-55.
- [7] 李振德,张正农,WHITE T. 浅析日本金鱼[J]. 水族世界, 2012(4): 76-81.
LI Z D, ZHANG Z N, WHITE T. A brief analysis of Japanese goldfish [J]. Aquarium, 2012(4): 76-81.
- [8] 天山雪. 兰寿起源之寻脉追踪[J]. 水族世界, 2010(6): 84-91.
TIAN S X. Ranchu origin of pulse tracking [J]. Aquarium, 2010(6): 84-91.
- [9] 王姝妍. 古环境变化和人工选择对鲫鱼鱼类遗传结构的影响[D]. 北京: 中国科学技术大学, 2013.
WANG S Y. The impacts of paleoenvironmental changes and anthropogenic selection on the genetic structure of *Carassius* [D]. Beijing: University of Science and Technology of China, 2013.
- [10] 王昭明. 日本金鱼良种概述[J]. 水产学杂志, 1998, 11(2): 50-60.
WANG Z M. About Japanese strains of Chinese ornamental goldfish [J]. Chinese Journal of Fisheries, 1998, 11(2): 50-60.
- [11] 毕东旭,张饮江,王芳,等. 中国金鱼起源与发展研究进展[C]//2018年中国水产学会学术年会论文摘要集. 西安: 中国水产学会, 2018: 88.
BI D X, ZHANG Y J, WANG F, et al. Research progress on the origin and development of Chinese goldfish [C]// 2018 China Society of Fisheries. Xi'an, 2018: 88.
- [12] 周潘. 中国金鱼及其养殖[J]. 北京水产, 2000(5): 33-43.
ZHOU X. Goldfish and cultivation in China [J]. Journal of Beijing Fisheries, 2000(5): 33-43.
- [13] 王春元,李延龄. 金鱼(*Carassius auratus*)染色体组型的研究 I. 鲫鱼和红龙睛金鱼染色体组型的比较[J]. 遗传学报, 1982, 9(3): 238-242.
WANG C Y, LI Y L. Studies on the karyotype of goldfish (*Carassius auratus*) I. A comparative study of the

- chromosomes in crucian and red dragon-eye goldfish [J]. *Journal of Genetics and Genomics*, 1982, 9(3): 238-242.
- [14] 罗莉中, 毕世华, 王春元. 金鱼乳酸脱氢酶的同工酶的发生遗传学研究——I. 鲫鱼和红龙睛金鱼各组织器官乳酸脱氢酶的同工酶的比较[J]. *遗传学报*, 1982, 9(5): 375-380.
- LUO L Z, BI S H, WANG C Y. Studies on developmental genetics of lactate dehydrogenase isozymes in goldfish (*Carassius auratus* L.) I. Comparison of lactate dehydrogenase isozymes in various organs of crucian and red dragon-eye goldfish [J]. *Journal of Genetics and Genomics*, 1982, 9(5): 375-380.
- [15] 王晓梅, 宋文芹, 李秀兰, 等. 用 RAPD 技术检测野生鲫鱼和四个金鱼代表品种的基因组 DNA 多态性[J]. *遗传*, 1998, 20(5): 7-11.
- WANG X M, SONG W Q, LI X L, et al. Detection of the genomic DNA polymorphisms in the wild crucian and four representative varieties of goldfish using RAPD technique [J]. *Hereditas (Beijing)*, 1998, 20(5): 7-11.
- [16] 张利红. 中国金鱼品种间遗传多样性的 SDS-PAGE 蛋白质标记和 AFLP 标记分析研究[D]. 杭州: 浙江大学, 2007.
- ZHANG L H. Research on genetic diversities of varieties of goldfish (*Carassius auratus* var.) by SDS-PAGE and AFLP markertechniques[D]. Hangzhou: Zhejiang University, 2007.
- [17] 张利红, 陈坚, 吴信忠. 金鱼两个品种遗传多样性的 AFLP 研究[J]. *水产科学*, 2008, 27(3): 139-141.
- ZHANG L H, CHEN J, WU X Z. Analysis of genetic diversity of two varieties of goldfish by AFLP [J]. *Fisheries Science*, 2008, 27(3): 139-141.
- [18] 朱雪莲, 王志勇, 韩志强. 借助 mtDNA 控制区序列分析金鱼与不同地域鲫的亲缘关系[J]. *大连水产学院学报*, 2010, 25(2): 153-157.
- ZHU X L, WANG Z Y, HAN Z Q. The phylogenetic relationship between goldfish and crucian carp *Carassius auratus* in different regions based on mtDNA D-loop region sequence [J]. *Journal of Dalian Fisheries University*, 2010, 25(2): 153-157.
- [19] 庄远红. 不同品系金鱼和鲫鱼的线粒体分子标记及其系统发育分析[D]. 长沙: 湖南师范大学, 2008.
- ZHUANG Y H. Mitochondrial DNA molecular markers and phylogenetic relationship of goldfishes and crucian carps [D]. Changsha: Hunan Normal University, 2008.
- [20] 庄远红, 周毅, 牛艳东, 等. 不同品种金鱼和鲫鱼的分子系统发育关系研究[J]. *激光生物学报*, 2009, 18(1): 46-49.
- ZHUANG Y H, ZHOU Y, NIU Y D, et al. Molecular phylogenetic relationships of different varieties of goldfishes (*Carassius auratus* Var) and crucian carps (*Carassius auratus*) [J]. *Acta Laser Biology Sinica*, 2009, 18(1): 46-49.
- [21] CHEN Z L, OMORI Y, KOREN S, et al. De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication [J]. *Science Advances*, 2019, 5(6): eaav0547.
- [22] KUANG Y Y, ZHENG X H, LI C Y, et al. The genetic map of goldfish (*Carassius auratus*) provided insights to the divergent genome evolutions in the Cyprinidae family [J]. *Scientific Reports*, 2016, 6(1): 34849.
- [23] KON T, OMORI Y, FUKUTA K, et al. The Genetic basis of morphological diversity in domesticated Goldfish [J]. *Current Biology*, 2020, 30(12): 2260-2274.
- [24] CHEN D, ZHANG Q, TANG W Q, et al. The evolutionary origin and domestication history of goldfish (*Carassius auratus*) [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(47): 29775-29785.
- [25] MEYER M, KIRCHER M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing [J]. *Cold Spring Harbor Protocols*, 2010, 2010(6): pdb.prot5448.
- [26] FISHER S, BARRY A, ABREU J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries [J]. *Genome Biology*, 2011, 12(1): R1.
- [27] LI C H, HOFREITER M, STRAUBE N, et al. Capturing protein-coding genes across highly divergent species [J]. *BioTechniques*, 2013, 54(6): 321-326.
- [28] WANG Y, YUAN H, HUANG J M, et al. Inline index helped in cleaning up data contamination generated during library preparation and the subsequent steps [J]. *Molecular Biology Reports*, 2022, 49(1): 385-392.
- [29] JIANG J M, YUAN H, ZHENG X, et al. Gene markers for exon capture and phylogenomics in ray-finned fishes [J]. *Ecology and Evolution*, 2019, 9(7): 3973-3983.
- [30] SHEN W, LE S, LI Y, et al. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation [J]. *PLoS One*, 2016, 11(10): e0163962.
- [31] MCKENNA A, HANNA M, BANKS E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome Research*, 2010, 20(9): 1297-1303.
- [32] PRITCHARD J K, STEPHENS M, DONNELLY P. Inference of population structure using multilocus genotype data [J]. *Genetics*, 2000, 155(2): 945-959.
- [33] BRAASCH I. Genome evolution: domestication of the allopolyploid goldfish [J]. *Current Biology*, 2020, 30(14): R812-R815.
- [34] EARL D A, VONHOLDT B M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method [J]. *Conservation Genetics Resources*, 2012, 4(2): 359-361.
- [35] JAKOBSSON M, ROSENBERG N A. CLUMPP: a cluster

- matching and permutation program for dealing with label switching and multimodality in analysis of population structure[J]. *Bioinformatics*, 2007, 23(14): 1801-1806.
- [36] ROSENBERG N A. DISTRUCT: a program for the graphical display of population structure [J]. *Molecular Ecology Notes*, 2004, 4(1): 137-138.
- [37] SHLENS J. A tutorial on principal component analysis[J]. arXiv:1404.1100, 2014.
- [38] FITAK R R. *OptM*: estimating the optimal number of migration edges on population trees using *Treemix* [J]. *Biology Methods and Protocols*, 2021, 6(1): bpab017.
- [39] DANECEK P, AUTON A, ABECASIS G, et al. The variant call format and VCFtools [J]. *Bioinformatics*, 2011, 27(15): 2156-2158.

Phylogenetic history of Chinese goldfish based on gene enrichment data

ZHOU Jingtong^{1,2}, HE Wei^{1,2}, LI Chenhong^{1,2}, ZHANG Jinghang^{1,2}, PAN Yiling³

(1. *Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, Shanghai 201306, China*; 2. *Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China*; 3. *Shanghai Natural History Museum, The Branch of the Shanghai Science & Technology Museum, Shanghai 200041, China*)

Abstract: Goldfish, is an ornamental fish originated in China, which was domesticated from wild crucian carp. There are great differences between different types of goldfish, such as presence or absence of dorsal fin, telescope or bubble eyes, oranda and other characters. Long-term hybridization and artificial selection have produced a large number of goldfish strains, but it also leads to the confusion and complexity in determining pedigree relationship of goldfish, which makes it more difficult to distinguish goldfish strains. People tried many ways to clarify the pedigree history and hybridization relationship between various goldfish strains, but so far no progress has been made. That is why the selection history of most goldfish strains has not been clarified and the hybridization relationships of many goldfish strain have not been confirmed. In this study, tissue samples of 59 individuals from 11 Chinese goldfish strains were collected. Cross-species target gene enrichment and high-throughput sequencing were used to obtain the target gene sequencing data of each goldfish individual. Using 4 434 single-copy gene loci of zebrafish as reference sequences, the corresponding goldfish genes were grouped according to its heterotetraploid subgenome, and the enriched sequences were compared and categorized according to two subgenomes respectively. A total of 35 298 single nucleotide polymorphism (SNPs) loci were called, which were used to study the pedigree history and hybridization relationship of the goldfish strains. Principal component analyses, genetic structure analyses and phylogenetic network map of the two subgenomes were reconstructed to reveal genetic relationship of these Chinese goldfish strains and explain their selection and hybridization history: (1) The telescope and Ryukin were directly evolved from the wakin, while pearlscale and lionhead were not directly evolved from the wakin. (2) Ranchu evolved earlier than egg bubble eye, while the bubble eye had secondly lost its dorsal fin. (3) The telescope-oranda is closely related to the lionhead, which first generated the character of oranda, and later obtained the telescope eyes. (4) Pearlscale, pearlscale-oranda and telescope-pearlscale-oranda are genetically similar. Pearlscale-oranda directly evolved from pearlscale, telescope-pearlscale-oranda is the latest bred form, probably a hybrid of telescope and pearlscale-oranda. The results of this study may help to understand the evolution history of Chinese goldfish and provide some reference for the classification and breeding of goldfish varieties.

Key words: goldfish; telescope-pearlscale-oranda; lionhead; Ryukin; single-copy conserved nuclear-coding sequence; gene enrichment; lineage study