

DOI: [10.12101/j.issn.1004-390X\(n\).202004062](https://doi.org/10.12101/j.issn.1004-390X(n).202004062)

基于高通量测序的日本医蛭转录组及 SSR 序列分析*

余 米, 封孝兰, 周 梦, 张承露, 曹 敏**
(重庆市药物种植研究所, 重庆 408435)

摘要:【目的】分析日本医蛭在不同温度处理下, 其转录组和简单重复序列标记 (simple sequence repeats, SSR) 信息。【方法】应用高通量测序技术, 对不同温度下日本医蛭不同组织分别进行转录组测序, 采用 MISA 软件对 unigene 进行 SSR 分析。【结果】共获得 125.08 G 有效数据, 平均错误率为 0.03%。经 Trinity 软件混合拼接后共得到 202 570 条 transcripts 和 52 206 条 unigene, 平均长度分别为 1 826 bp 和 1 390 bp, 最大长度均为 30 853 bp。将 unigene 与公共数据库进行比对, 得到注释的 unigene 为 29 352 条, 占总数的 56.22%。对 unigene 进行 SSR 检测, 共检测到 29 926 个 SSR 位点, 发生频率是 57.32%。SSR 的核苷酸基元重复数都在 5 次以上, 主要集中在 5~10 次 (86.27%) 之间, 以三核苷酸和单核苷酸重复 SSR 为主, 所占比例分别为 74.07% 和 15.10%, 核苷酸重复基元中, 以 A/T (4 255, 14.22%) 为主, 在三核苷酸重复基元中, ATC/ATG (9 496, 31.73%) 所占比例最高。不同温度下日本医蛭转录组基因表达量差异比较数据显示, 样本间表达差异基因共 1 603 个。【结论】不同温度处理后日本医蛭的 SSR 发生频率较高, 可为日本医蛭筛选分子遗传标记提供大量基础数据。

关键词: 日本医蛭; 转录组; 简单重复序列标记

中图分类号: S 917.4

文献标志码: A

文章编号: 1004-390X (2021) 02-0250-10

Transcriptome and SSR Analysis Based on RNA-Seq in *Hirudo nipponia*

YU Mi, FENG Xiaolan, ZHOU Meng, ZHANG Chenglu, CAO Min

(Chongqing Institute of Medicinal Plant Cultivation, Chongqing 408435, China)

Abstract: [Purpose] To analyze the transcriptome and SSR information of *Hirudo nipponia* under different temperatures. [Method] Illumina HiSeq high-throughput sequencing technology and MISA software were applied to obtain the transcriptome and SSRs from different tissues of *H. nipponia* at different temperatures. [Result] The 125.08 G effective data was obtained with an average error rate of 0.03%. After assembly by the software Trinity, a total of 202 570 transcripts and 52 206 unigenes were generated, corresponding to maximum length of 30 853 bp both, the average length of 1 826 bp and 1 390 bp respectively. Using BLASTX against the public databases, 29 352 unigenes were annotated at least one databases, accounting for 56.22% of the total. In addition, 29 926

收稿日期: 2020-04-28

修回日期: 2020-11-13

网络首发时间: 2021-03-26 10:45:50

*基金项目: 重庆市科学技术局项目 (cstc2018jxjl-jbky00003, cstc2019jscx-msxmX0357); 重庆市卫健委“药用动物学”重点学科建设。

作者简介: 余米 (1987—), 男, 四川巴中人, 硕士, 助理研究员, 主要从事药用水生动物养殖开发利用。

E-mail: 344565975@qq.com

**通信作者 Corresponding author: 曹敏 (1988—), 女, 四川宜宾人, 硕士, 助理研究员, 主要从事中药资源调查与保育。E-mail: 962103004@qq.com

网络首发地址: <http://kns.cnki.net/kcms/detail/53.1044.S.20210324.0902.001.html>



potential SSR loci were detected, with a frequency of 57.32%. Trinucleotide and mononucleotide were major types, accounting for 74.07% and 15.10%, respectively. ATC/ATG (9 496) and A/T (4 255) were respectively most frequent motifs in trinucleotide and mononucleotide repeats, accounting for 31.73% and 14.22%, respectively. There were 1 603 differentially expressed genes among the samples. [Conclusion] The SSR of *H. nipponia* has a high frequency, which provides basis for the screening of molecular genetic markers.

Keywords: *Hirudo nipponia*; transcriptome; simple sequence repeats (SSR)

日本医蛭 (*Hirudo nipponia*) 俗称水蛭、日本医水蛭或稻田医蛭, 属医蛭行亚目 (Hirudiniformes Caballero) 医蛭科 (Hirudinidae Whitman) 医蛭属 (*Hirudo*), 在中国南北多省均有分布^[1]。由于其含有高效抗凝血物质——水蛭素, 是心脑血管疾病药物的主要动物源材料。近年来, 随着中药材市场和药企需求的增加, 过度捕捞加之天然水域水质的改变, 使得天然水域日本医蛭的资源量减少和个体趋于小型化^[2], 各地对日本医蛭的养殖逐渐兴起。

高通量测序技术具有成本低、不需要预先得到物种基因信息和获得的数据丰富等优点, 可用于研究不同蛭类对不同环境的适应内在机制^[3]、宽体金线蛭在饲喂前后抗凝血因子的基因表达^[4]以及日本医蛭在饥饿胁迫后唾液腺基因表达水平的差异^[5]。以上研究都集中在蛭类的进化和抗凝血活性方面, 有关日本医蛭不同组织的基因转录和 SSR 分析未见报道。温度是影响日本医蛭生理的主要因素, 18 ℃ 是日本医蛭繁殖的适宜温度, 24 ℃ 是其生长的适宜温度^[2], 2 种温度下日本医蛭不同组织的转录组与 SSR 信息分析能为筛选日本医蛭性别特异分子标记、分子标记辅助育种以及提高抗凝血活性等提供基础数据。

本试验以日本医蛭肌肉 (M)、生殖环带 (C) 和头部 (H) 3 种组织为研究对象, 采用新一代高通量测序技术进行转录组测序, 通过数据组装、组织表达量比较、差异表达基因注释和 SSR 序列分析, 以期日本医蛭繁殖和抗凝血活性的分子机理以及分子标记辅助育种提供理论依据。

1 材料与方法

1.1 样品采集

采用来自重庆市药物种植研究所水蛭繁育基地的 40 尾日本医蛭, 体质量为 1.425~2.304 g, 体

长 9.80~10.52 cm。2019 年 11 月, 开始分别置于 24 和 18 ℃ 恒温培养箱内饥饿处理 28 d, 分别剪取其肌肉 (M)、环带 (C) 和头部 (H) 组织, 浸泡在 RNA-hold 保存液中备用, 24 ℃ 处理的日本医蛭不同组织简写为: H24、C24 和 M24; 18 ℃ 处理的日本医蛭不同组织简写为: H18、C18 和 M18。

1.2 总 RNA 提取

利用 Trizol 法提取不同组织 RNA。采用琼脂糖凝胶电泳分析样品 RNA 完整性及是否存在 DNA 污染, 利用 NanoPhotometer spectrophotometer 检测 RNA 纯度 (OD₂₆₀/OD₂₈₀ 及 OD₂₆₀/OD₂₃₀ 比值); 利用 Agilent 2100 bioanalyzer 精确检测 RNA 完整性。将得到的 RNA 溶液置于 -80 ℃ 保存或立刻用于后续试验。

1.3 建库测序和拼接组装

日本医蛭转录组测序和拼接组装委托北京诺禾致源科技股份有限公司完成, 建库中使用的建库试剂盒为 Illumina 的 NEBNext[®] Ultra[™] RNA Library Prep Kit, 建库完成后使用 Agilent 2100 bioanalyzer 对文库的 insert size 进行检测, insert size 符合预期后, 采用 qRT-PCR 对文库有效浓度进行准确定量 (文库有效浓度高于 2 nmol/L)。获得 clean reads 后, 采用 Trinity^[6]对 clean reads 进行无参拼接。

1.4 功能注释

为获得相应基因功能的功能注释信息, 使用 BLAST 软件 (2.2.28+版本)^[7]在线与非冗余蛋白库 (non-redundant protein sequence database, NR)、核酸序列数据库 (nucleotide sequence database, Nt)、Swiss-prot 蛋白序列数据 (manually annotated and reviewed protein sequence database, Swiss-prot) 和真核生物蛋白质同源簇数据库 (eukaryotic ortholog groups, KOG) 比对, *E*-value 值均为 1e-5; 使用 KAAS (r140224) 软件^[8]与日本京都基因和基

因组百科全书 (kyoto encyclopedia of genes and genomes, KEGG) 比对, *E*-value 值均为 1e-10; 使用 Hmmscan (HMMER 3) 软件与蛋白质家族的集合数据库 (protein family, Pfam)^[9] 比对, *E*-value 值均为 0.01; 使用 Blast2go (b2g4pipe_v2.5)^[10] 基因本体数据库 (gene ontology, GO) 比对, *E*-value 值为 1.0e-6。

1.5 SSR 标记筛选

采用 MISA 1.0 版 (默认参数) 对应的各个 unit size 的最少重复次数分别为: 1~10、2~6、3~5、4~5、5~5、6~5 (例如 1~10, 单核苷酸为重复单位时, 其重复数至少为 10 才可被检测到; 2~6, 以双核苷酸为重复单位时, 其最少重复数为 6) 对 unigene 进行 SSR 检测。采用 Primer3 2.3.5 版 (默认参数) 进行 SSR 引物设计。

1.6 差异基因表达分析

将 Trinity 拼接得到的转录组作为参考序列 (RefSeq), 将每个样品的 clean reads 往 RefSeq 上做 mapping, 分别过滤掉低质量的 reads、带接头污染的 reads 以及含 N 的 reads。比对过程中我们采用了 RSEM 软件^[11], RSEM 中使用到的 bowtie2 参数 mismatch 0 (bowtie2 默认参数)。

2 结果与分析

2.1 数据质量评估及序列组装分析

经过测序质量控制, 共得到 125.08 G 的 clean data, 过滤后的碱基分布平稳, 平均 GC 含量为 42.58%, 不同温度下日本医蛭环带组织样品 Q30 碱基百分比均不小于 92.01% (表 1)。

将过滤后的 reads 片段进行聚类及拼接组装 (表 2), 共得到 202 570 条 transcript 和 52 206 条 unigene, transcript 的 N50 和 N90 (拼接转录本按照长度从长到短排序, 并进行长度累加, 当累加值不小于 unigene 总长 50%/90% 的数值就是 N50/N90) 分别为 2 894 和 820, unigene 的 N50 和 N90 分别为 2 483 和 520, 组装完整性较高。transcript 和 unigene 的序列长度分布于 300~2 000 bp 以上, 其中 >2 000 bp 区间的 transcript 分布最多 (为 65 746, 占 32.5%), unigene 在 300~500 bp 分布最多 (为 17 231, 占 33%) (图 1)。transcript 和 unigene 序列的平均长度分别为 1 826 和 1 390 bp, 最大和最小长度均为 30 853 bp 和 301 bp。

2.2 功能基因注释与表达量分布

将 unigene 序列与 7 大数据库进行比对, 得出了 unigene 在每个数据库中的功能注释信息。

表 1 不同温度下日本医蛭各组织测序数据统计
Tab. 1 Statistical data of RNA-seq for *H. nipponia* at different temperatures

样本 samples	原始数据 raw reads	清洁数据 clean reads	碱基数量/G clean bases	错误率/% error	Q20/%	Q30/%	GC 百分比/% GC percent
H24_1	45 431 080	44 676 774	6.70	0.03	97.50	92.58	41.88
H24_2	45 719 004	44 872 164	6.73	0.03	97.30	92.14	41.71
H24_3	47 655 932	46 815 800	7.02	0.03	97.28	92.11	43.10
C24_1	45 996 730	45 117 404	6.77	0.03	97.23	92.01	42.27
C24_2	51 090 894	49 760 010	7.46	0.03	97.32	92.23	42.28
C24_3	47 917 682	46 952 310	7.04	0.03	97.47	92.52	42.43
M24_1	50 189 638	49 072 616	7.36	0.03	97.48	92.54	42.73
M24_2	47 912 728	47 118 008	7.07	0.03	97.33	92.20	42.49
M24_3	47 104 418	45 704 948	6.86	0.03	97.37	92.37	43.34
H18_1	48 066 694	47 299 224	7.09	0.03	97.24	92.03	41.95
H18_2	46 001 150	45 192 382	6.78	0.03	97.32	92.25	42.16
H18_3	46 693 840	45 834 804	6.88	0.03	97.26	92.10	42.20
C18_1	45 453 330	44 652 910	6.70	0.03	97.29	92.10	42.89
C18_2	48 072 830	47 297 166	7.09	0.03	97.36	92.27	42.95
C18_3	46 965 374	45 732 230	6.86	0.03	97.51	92.60	42.65
M18_1	47 919 804	46 592 206	6.99	0.03	97.41	92.42	41.69
M18_2	46 058 130	45 030 182	6.75	0.03	97.40	92.38	42.28
M18_3	47 057 392	46 211 360	6.93	0.03	97.39	92.34	42.28

表 2 拼接转录本长度分布

Tab. 2 Distribution of number and quality of unigenes and transcripts

项目 item	最小长度/bp Min. length	平均长度/bp mean length	最大长度/bp Max. length	N50总条数 N50 total number	N90总条数 N90 total number	总条数 total number	总核苷酸/nt total nucleotides
独立基因 unigenes	301	1 390	30 853	2 483	520	52 206	72 566 318
转录本 transcripts	301	1 826	30 853	2 894	820	202 570	369 978 282

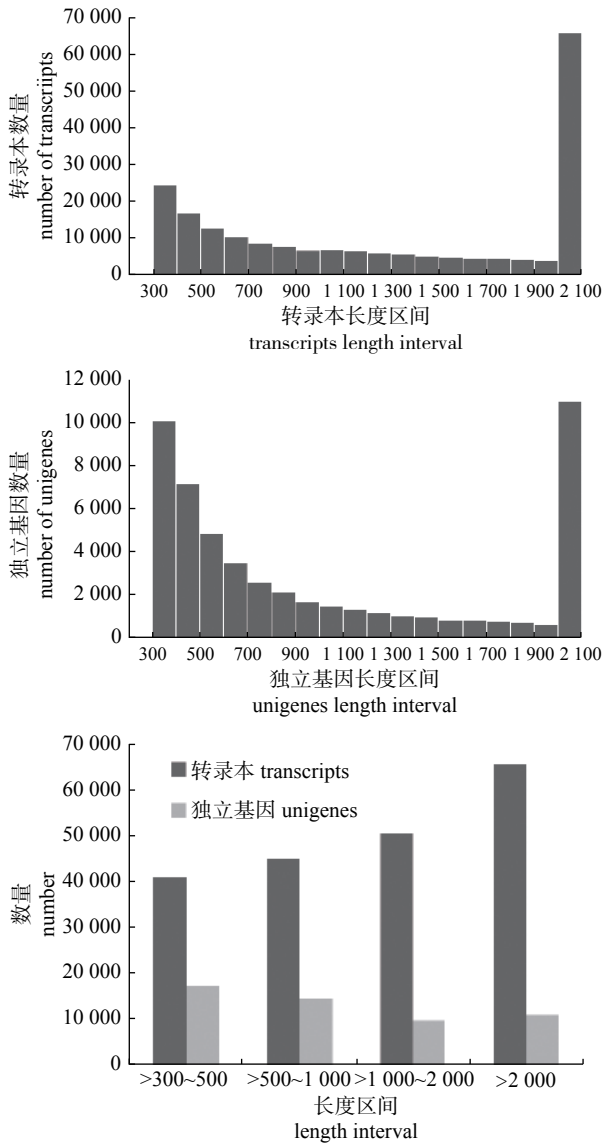


图 1 转录本与独立基因拼接长度和频数分布

Fig. 1 Distribution of splice length and frequency of transcripts and unigene

表 3 显示: 有 56.22% 的 unigene 至少被 1 个数据库注释。Pfam 和 GO 数据库均注释到 22 984 条, 占 unigene 总数的 44.02%, 其余依次为 NR 数据库 (21 967 条, 42.07%)、Swiss-Prot 数据库 (16 852, 32.27%)、KOG 数据库 (9 846, 18.86%)、KO 数据库 (6 408, 12.27%) 和 NT 数据库 (6 379, 12.21%)。

表 3 日本医蛭独立基因注释数量统计

Tab. 3 The statistical number of unigene that were functional annotated in *H. nipponia*

数据库 database	非重复序列基因数 number of unigene	比例/% percentage
NR	21 967	42.07
NT	6 379	12.21
KO	6 408	12.27
SwissProt	16 852	32.27
Pfam	22 984	44.02
GO	22 984	44.02
KOG	9 849	18.86
能被所有数据库注释数目 annotated in all databases	1 705	3.26
至少被 1 个数据库注释数目 annotated in at least one database	29 352	56.22
总独立基因数目 total unigene	52 206	100

2.2.1 注释基因的 NR 分类

通过与 NR 进行比对注释, 可以获取与日本医蛭转录组序列具有相似性的近缘物种信息, 结果 (图 2) 显示: 日本医蛭与淡水泽蛭、海蠕虫、鸭嘴舌形贝、舞行波豆虫和鞭虫具有较高的序列同源性, 其中与淡水泽蛭 (*H. robusta*) 具有最高的序列相似性 (42%)。

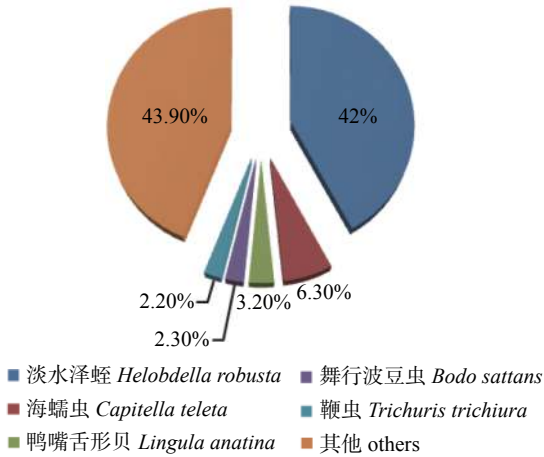


图 2 基于 NR 数据库比对的同源基因物种来源分布

Fig. 2 Species distribution of the top BLASTx hits against the NR database for unigene in *H. nipponia*

2.2.2 注释基因的 GO 分类

对 unigene 序列进行 GO 注释之后, 按照 GO 3 个基本大类 (生物学过程、细胞组分和分子功能) 将注释成功的全部基因按照进一步细分的功能进行聚类 (图 3)。基于序列同源性、被分成 3 大主要类别共有 56 个功能群。在 GO 分类的生物学过程中, 注释基因数量在二级分类排在前 3 位的分别是细胞过程 (13 933 条)、新陈代谢过程 (12 017 条) 和单有机体过程 (10 944 条)。在细胞组分分类中, 注释数量前三的依次为细胞 (8 385 条)、细胞成分 (8 385 条) 和细胞膜 (6 300 条)。在分子功能分类中, 注释最多和次多的分别为连接功能 (13 012 条) 与催化活性 (9 623 条)。

2.2.3 注释基因的 KOG 分类

KOG 数据库归为 26 个组, 将通过 KOG 成功注释的 unigene 按照分组进行分类 (图 4), 在 KOG 数据库中共注释到 9 849 条。在这 26 个分类中, 聚类数量最多的为信号转导机制 (1 553

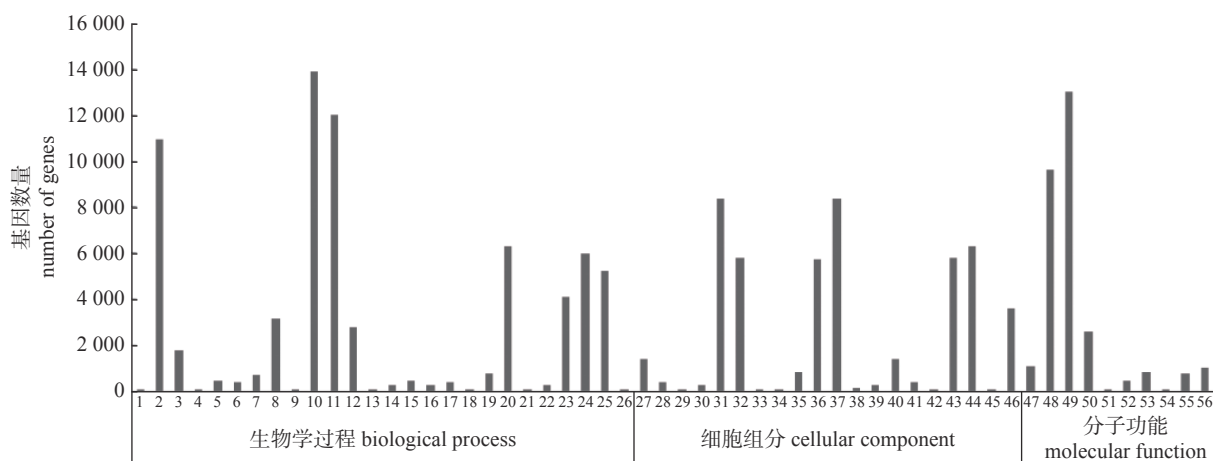
条, 15.77%), 其次是功能预测 (1 365 条, 13.86%), 第三是翻译后修饰、蛋白翻转和分子伴侣 (1 109 条, 11.26%), 聚类最少的是未修饰蛋白 (2 条)。

2.2.4 注释基因的 KEGG 分类

以 KEGG 代谢途径数据库为依据, 对 unigene 的注释结果进行分类 (图 5)。日本医蛭的 unigene 代谢途径可分为细胞过程 (A)、环境信息处理 (B)、遗传信息处理 (C)、代谢 (D) 和有机系统 (E) 32 个亚类; 其中, 涉及 unigene 数量最多的 5 个代谢途径分别为 B 分支中的信号转导 (969 条)、E 分支的内分泌系统 (575 条)、C 分支的翻译 (553 条)、A 分支中的运输与分解代谢 (493 条) 和细胞群落 (388 条)。

2.3 unigene 的 SSR 分析

SSR 被称为短串联重复序列或微卫星 DNA 标记。对日本医蛭转录组的 52 206 条 unigene 序列进行 SSR 分析, 共鉴定出 29 926 个 SSR, 总



注: 1. 生物相; 2. 单有机体过程; 3. 多有机体过程; 4. 生长; 5. 移动; 6. 生物正调控过程; 7. 发育过程; 8. 组织或生物源的细胞组分; 9. 节奏进程; 10. 细胞过程; 11. 新陈代谢过程; 12. 信号; 13. 细胞聚集; 14. 免疫反应过程; 15. 生物黏附; 16. 生殖过程; 17. 生物负调控过程; 18. 行为; 19. 多细胞生物体过程; 20. 生物调节; 21. 解毒; 22. 生殖; 23. 刺激反应; 24. 生物过程调控; 25. 定位; 26. 细胞杀伤; 27. 病毒部分; 28. 细胞外区域; 29. 突触成分; 30. 其他有机体; 31. 细胞; 32. 细胞器; 33. 突触; 34. 细胞结合; 35. 膜包围的内腔; 36. 细胞膜成分; 37. 细胞成分; 38. 细胞外基质; 39. 其他有机体成分; 40. 病毒; 41. 细胞外区域成分; 42. 细胞外基质成分; 43. 大分子聚合物; 44. 细胞膜; 45. 内核; 46. 细胞器成分; 47. 分子传导活力; 48. 催化活性; 49. 连接功能; 50. 载体活性; 51. 金属伴侣活性; 52. 转录因子活性, 蛋白质结合; 53. 分子功能调节; 54. 抗氧化活性; 55. 结构分子活力; 56. 核苷酸连接转录因子活性。

Note: 1. biological phase; 2. single-organism process; 3. multi-organism process; 4. growth; 5. locomotion; 6. positive regulation of biological process; 7. developmental process; 8. cellular component organization or biogenesis; 9. rhythmic process; 10. cellular process; 11. metabolic process; 12. signaling; 13. cell aggregation; 14. immune system process; 15. biological adhesion; 16. reproductive process; 17. negative regulation of biological process; 18. behavior; 19. multicellular organismal process; 20. biological regulation; 21. detoxification; 22. reproduction; 23. response to stimulus; 24. regulation of biological process; 25. localization; 26. cell killing; 27. virion part; 28. extracellular region; 29. synapse part; 30. other organism; 31. cell; 32. organelle; 33. synapse; 34. cell junction; 35. membrane-enclosed lumen; 36. membrane part; 37. cell part; 38. extracellular matrix; 39. other organism part; 40. virion; 41. extracellular region part; 42. extracellular matrix component; 43. macromolecular complex; 44. membrane; 45. nucleoid; 46. organelle part; 47. molecular transducer activity; 48. catalytic activity; 49. binding; 50. transporter activity; 51. metallochaperone activity; 52. transcription factor activity, protein binding; 53. molecular function regulator; 54. antioxidant activity; 55. structural molecule activity; 56. nucleic acid binding transcription factor activity.

图 3 日本医蛭 GO 分类

Fig. 3 GO classification

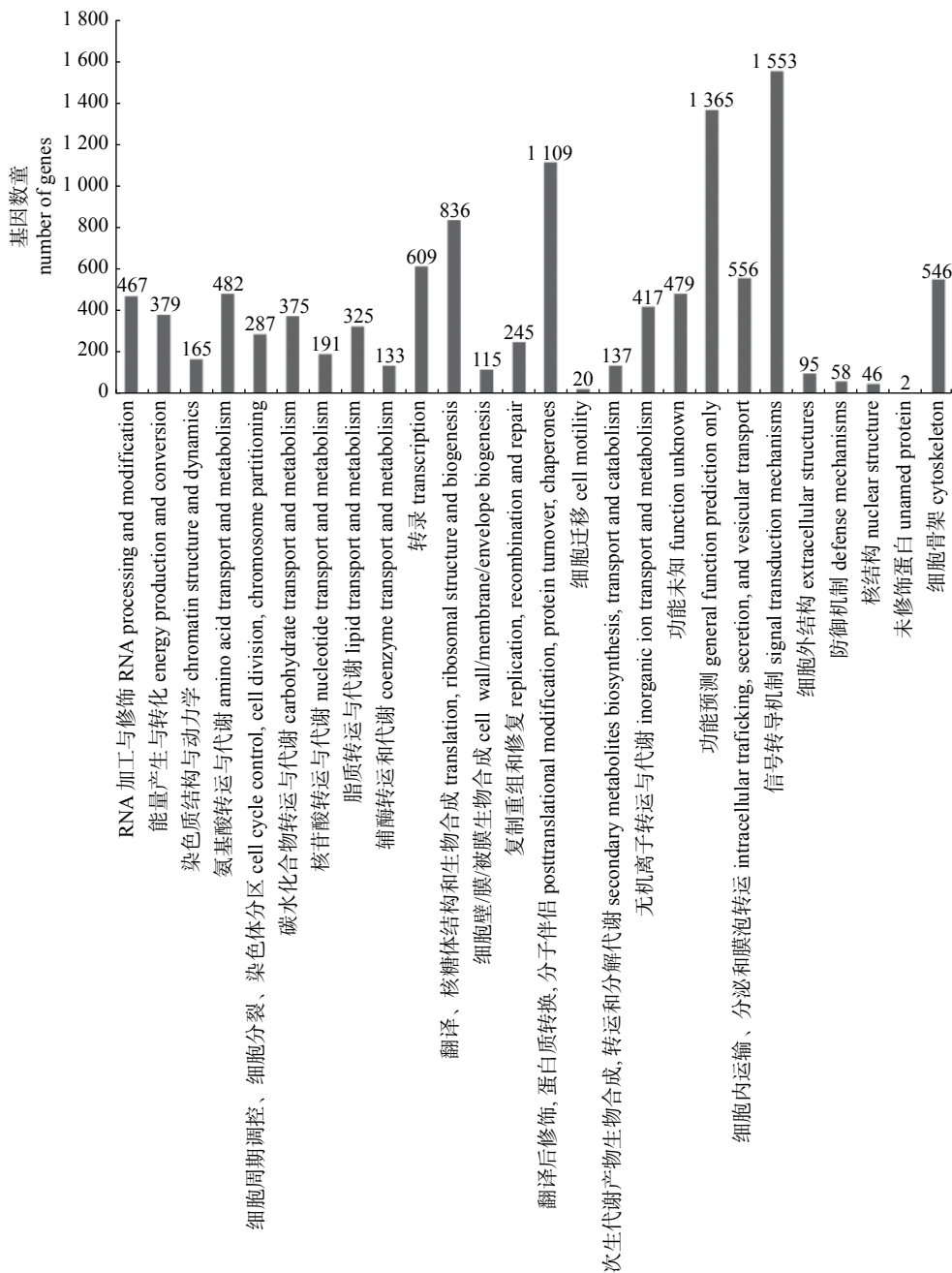


图 4 日本医蛭 KOG 分类
Fig. 4 KOG classification

长度 72 566 318 bp。这些 SSR 分布于 16 094 个序列中, 其中有 7 188 条含有 1 个以上 SSR, 其中在化学成分形成中的 SSR 数量有 5 664 个。图 6、7 表明: 三核苷酸 SSR 数量最多, 有 22 167 个, 占 6 种核苷酸重复类型的 74.07%, 其次是单核苷酸 (4 519 个, 15.10%)、四核苷酸 (1 860 个, 6.22%)、二核苷酸 (1 174 个, 3.92%)、六核苷酸 (126 个, 0.42%) 和五核苷酸 (80 个, 0.27%)。

2.4 日本医蛭 SSR 核苷酸基序重复类型及其频率特征分析

从日本医蛭转录组 SSR 核苷酸基序类型来看, 其中 29 926 个 SSR 位点包含了 92 种重复基元。由表 4 所示: 单核苷酸至六核苷酸重复分别有 2、4、10、23、24 和 29 种; 出现最多的基元是 ATC/ATG (9 496 个, 31.73%), 其次是 AAT/ATT (7 075 个, 23.64%) 和 AAC/GTT (3 083 个,

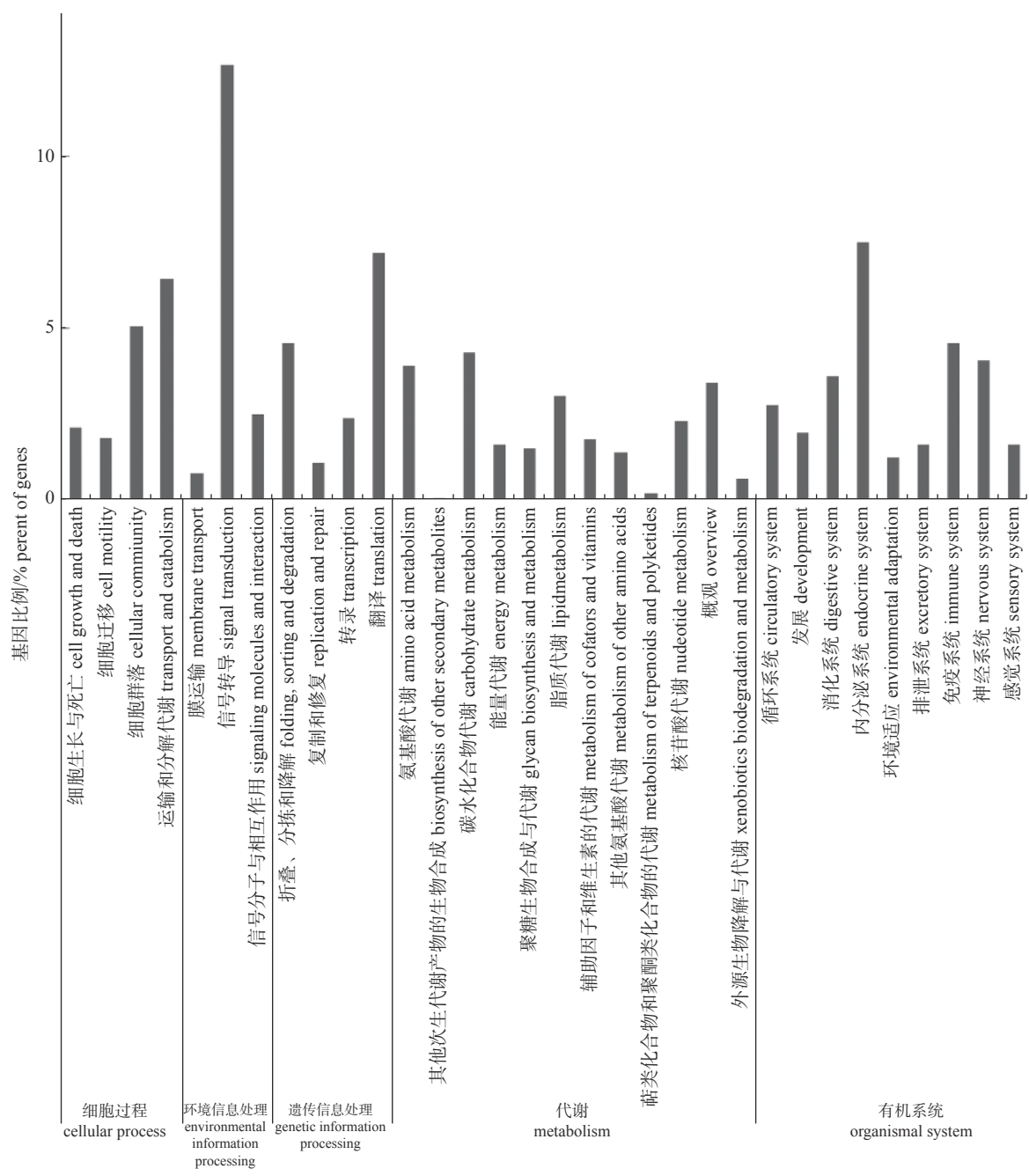


图 5 日本医蛭的 KEGG 分类
Fig. 5 KEGG classification

10.30%)。二核苷酸重复基元以 AC/GT、AG/CT 和 AT/AT 为主，分别占 56.81%、21.98% 和 20.70%。三核苷酸重复基元以 ATC/ATG、AAT/ATT 和 AAC/GTT 为主，分别占 42.84%、31.92% 和 13.90%。四核苷酸重复基元以 ATCC/ATGG、AAAT/ATTT 和 AATG/ATTC 为主，分别占 18.75%、12.5% 和 11.25%。五核苷酸重复基元以 AATAT/ATA-TT、AACAT/ATGTT 和 AAAAT/ATTTT 为主，

分别占 35.59%、18.17% 和 14.95%。六核苷酸重复基元以 AATGAT/ATCATT、ACCATC/ATG-GTG 和 AATGAT/ATCATT 为主，分别占 59.52%、4.76% 和 4.76%。二核苷酸基元重复数主要为 6 次，三~六核苷酸基元重复数均为 5 次。

2.5 日本医蛭不同部位差异基因表达分析
首先对原始的 read count 进行标准化 (normalization)，主要是对测序深度的校正；然后通过

统计学模型进行假设检验概率 (P -value) 的计算, 最后进行多重假设检验校正 (BH), 得到 FDR 值, 差异基因筛选标准为: $|\log_2(\text{FoldChange})| > 1$ 和 $\text{padj} < 0.05$ 。筛选结果如表 5 所示: 日本医蛭头部

在不同温度下差异表达 434 个基因, 上调 211 个, 下调 223 个; 环带部位在不同温度下差异表达 591 个基因, 上调 191 个, 下调 400 个; 肌肉组织差异表达 578 个, 上调 362 个, 下调 216 个。

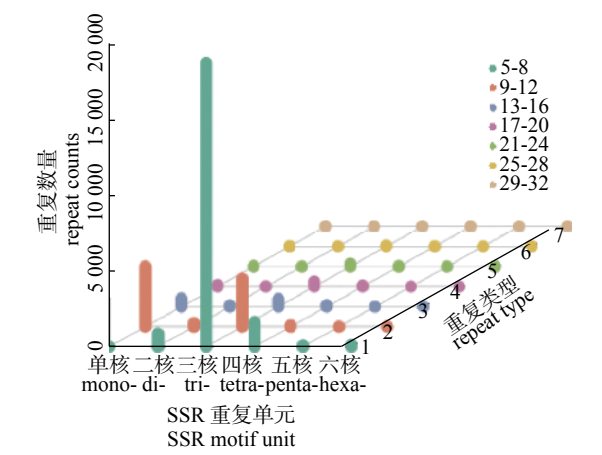


图 6 SSR 重复单元
Fig. 6 SSR motif unit

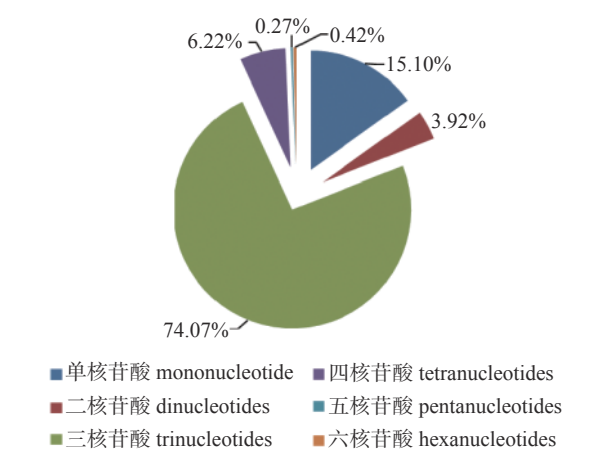


图 7 日本医蛭重复基元种类分布
Fig. 7 Types and distribution of SSRs

表 4 日本医蛭 SSR 不同核苷酸重复基元的类型及频率
Tab. 4 Types and frequencies of different nucleotide repeat motifs of SSR in *H. nipponia*

重复基元类型 repeat motif length	重复次数 repeat number							总计 total	频数/% frequency
	5	6	7	8	9	10	>10		
A/T						2 490	1 765	4 255	14.22
C/G						115	149	264	0.88
AC/GT	—	112	47	28	22	12	37	258	0.86
AG/CT	—	84	40	29	17	18	55	243	0.81
AT/AT	—	239	150	104	57	34	83	667	2.23
AAC/GTT	1 315	705	405	213	196	109	139	3 082	10.30
AAT/ATT	2 485	1 201	766	536	777	507	803	7 075	23.64
ATC/ATG	3 935	2 139	1 248	745	579	397	453	9 496	31.73
AAAT/ATTT	221	68	35	5	1	3	5	338	1.13
AATG/ATTC	110	54	67	10	2	1	34	278	0.93
ATCC/ATGG	199	126	159	34	9	20	115	662	2.21
AAAAT/ATTTT	6	3						9	0.03
AACAT/ATGTT	7	2					1	10	0.03
AATAT/ATATT	8	1	1			1	4	15	0.05
AATGAT/ATCATT	4				1	1		6	0.02
ACCATC/ATGGTG	3	1				1	1	6	0.02
ACGATG/ATCGTC	50	4	4	10	3	3	1	75	0.25
总计 total	8 343	4 739	2 922	1 714	1 664	3 712	3 645	26 739	89.35
频数/% frequency	82.88	87.08	88.81	91.61	95.52	98.86	96.97	89.35	

表 5 差异表达基因数

Tab. 5 Number of differentially expressed genes			
组织 tissue	差异表达基因 all DEGs	下调基因 down-regulated genes	上调基因 up-regulated genes
H18 vs H24	434	223	211
C18 vs C24	591	400	191
M18 vs M24	578	216	362

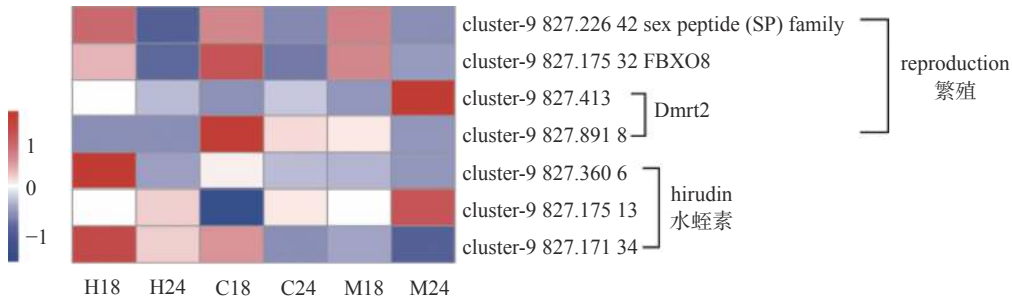


图 8 日本医蛭不同组织相关基因表达热图

Fig. 8 Heat map of gene expression in different tissues of *H. nipponia*

3 讨论

日本医蛭作为一种重要的心脑血管类药物的原材料，国内在日本医蛭的抗凝血方面做了大量的基础性研究，但在分子育种和代谢层面缺乏相关的基础数据资料。为了获取日本医蛭不同组织的转录组信息，本试验在不同温度下对其头部、环带和肌肉 3 种组织进行高通量转录组测序，对数据进行过滤与质控后，采用 Trinity 软件组装、拼接和转录本功能注释，得到 125.08 G 转录组 clean data 信息，这些数据的积累可为开展日本医蛭的繁殖和代谢提供参考资料。从转录组质量来看，共组装得到平均长度为 1 390 bp unigene 共计 52 206 条，未发现<200 bp 的序列。N50 和 N90 分别为 2 483 和 520，Q30 均在 92.01% 以上，其测序质量可靠。

本研究得到的 SSR 有 29 926 个，unigene 共 52 206 条，SSR 出现的频率是 57.32%，远高于蛭类中的棒纹牛蛭 (35.6%)、洞穴山蛭 (24.9%) 和宽体金线蛭 (22.3%)^[12]，也高于江鳕 (10.27%)^[13]和牙鲆 (7.95%)^[14]等鱼类。由于选取的是肌肉、环带和头部不同组织的混合样本，RNA 来源丰富，组装的转录组质量较好，所以筛选到的 SSR 数量相对较多，这些大量的序列为发现日本医蛭新基因提供了充足的转录组序列基源。同时也证实了转录组测序是开发 SSR 分子标记一种有效方法。

本研究所有鉴定到的微卫星中，SSR 的核苷

如图 8 所示：整体来说，水蛭素基因 cluster-9827.3606 与 cluster-9827.17134 在 H18 中表达水平最高；基因 cluster-9827.22642 和 cluster-9827.17532 在 H18、C18 和 M18 中表达均上调；cluster-9827.8918 在 C18 中的表达水平较高。

酸基元重复数都在 5 次以上，主要集中在 5~10 次之间 (86.27%)，以三核苷酸和单核苷酸重复 SSR 为主，所占比例分别为 74.07% 和 15.10%，六核苷酸和五核苷酸最少，占比分别为 0.42% 和 0.27%。刘洪涛等^[15]认为富含 A/T 的序列退火温度较低，有利于 DNA 解链以增加其出现的概率，同时由于蛋白质密码子为三碱基造成处于编码区的 SSR 三碱基重复占比最高。另外，本研究结果显示：在三核苷酸重复基元中，ATC 所占比例最高，为 9 496 次；单核苷酸重复基元中，以 A 为主，为 4 255 次；四核苷酸重复基元中以 ATCC 为主，达到了 662 次。这一结果与王斌等^[12]对蛭类的报道存在一定差异，可能是由于物种不同，其 SSR 核苷酸基序重复类型本身就存在差异，例如江鳕以一、二核苷酸基元为主^[13]，牙鲆和双须骨舌鱼以二、三核苷酸为主^[14, 16]。

日本医蛭其咽周腺体 (位于头部) 分泌的水蛭素，是高效的抗凝血活性物质，本次试验结果表明：在 18 ℃ 下，日本医蛭头部的水蛭素基因表达明显 (图 8)，故可在养殖时进行适当降温处理增加日本医蛭的水蛭素总体含量。我们在解剖取样时发现：经过 28 d 处理后日本医蛭体内依然储藏着大量的未凝固血液，18 ℃ 相较于 24 ℃ 处理组其储藏量更大，我们推测降温后，日本医蛭为了防止体内血液的凝固而大量分泌水蛭素，而对照组环境温度较高，其血液的代谢速度较快，水蛭素的分泌减少。

基因 cluster-9827.22642 [Sex peptide (SP) family] 是性肽家族一员,在 18 ℃ 3 种组织中表达均升高,性肽家族在日本医蛭体内参与雌性交配后的调节和信号转导等生物学过程,果蝇性肽 (SP) 被认为是一种精液成分,可诱导受精雌性的产后反应,如抑制再生和刺激产卵,SP 被认为在果蝇的性选择和性冲突中起着中心作用^[17-18]。基因 cluster-9827.17532 (FBXO 家族) 在泛素介导的蛋白质水解过程中具有底物识别的特性和功能,在本试验中 FBXO8 主要参与 ARF 蛋白信号转导、GTPase 酶活性调节和 DNA 双链断裂形成等生物学过程。此外,日本医蛭是雌雄同体、异体受精的生物,其在繁殖季节关于日本医蛭性别的扮演一直未见报道,Dmrt (double sex and mab-3 related transcription factor) 是重要的性别决定基因,本研究中发现日本医蛭在 18 ℃ 下性别决定相关基因 cluster-9827.8918 在 C18 中的表达水平较高,表明 18 ℃ 下日本医蛭性别可能发生转换。由于我们取样环带部位包括了日本医蛭所有生殖腺,故不能确定这些性别决定基因在雌雄生殖腺中的表达情况。但有文献证实性别决定的假定基因位于性别决定的遗传图谱上,具有丰富的小卫星标记^[19-21]。因此,本研究为进一步研究日本医蛭繁殖的分子机理和筛选性别特异分子标记积累了基础数据,同时有利于抗凝血相关的分子标记的开发。

4 结论

不同温度处理后日本医蛭的 SSR 发生频率较高,抗凝血与繁殖相关基因 (cluster-9827.3606、cluster-9827.22642 和 cluster-9827.413 等) 表达明显,其头部和环带组织可作为日本医蛭筛选抗凝血活性和分子遗传标记的供试组织。

[参考文献]

- [1] 杨潼. 中国动物志, 环节动物门, 蛭纲[M]. 北京: 科学出版社, 1996.
- [2] 余米, 刘嘉, 周梦, 等. 日本医蛭生境水质及生物学特性初探[J]. 中药材, 2019, 42(12): 2782. DOI: 10.13863/j.issn1001-4454.2019.12.009.
- [3] LIU Z C, ZHAO F, TONG X R, et al. Comparative transcriptomic analysis reveals the mechanism of leech environmental adaptation[J]. Gene, 2018, 664: 70. DOI: 10.1016/j.gene.2018.04.063.
- [4] KHAN M S, GUAN D, KVIST S, et al. Transcriptomics and differential gene expression in *Whitmania pigra* (Annelida: Clitellata: Hirudinida: Hirudinidae): contrasting feeding and fasting modes[J]. Ecology and Evolution, 2019, 9(8): 4706. DOI: 10.1002/ece3.5074.
- [5] 邢月婷, 管峰, 李诗语, 等. 日本医蛭唾液腺对饥饿胁迫响应的转录组比较[J]. 水产学报, 2020, 44(5): 754.
- [6] GRABHERR M G, HAAS B J, YASSOUR M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. Nature Biotechnology, 2011, 29(7): 644. DOI: 10.1038/nbt.1883.
- [7] ALTSCHUL S F, MADDEN T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs[J]. Nucleic Acids Research, 1997, 25(17): 3389. DOI: 10.1093/nar/25.17.3389.
- [8] MORIYA Y, ITOH M, OKUDA S, et al. KAAS: an automatic genome annotation and pathway reconstruction server[J]. Nucleic Acids Research, 2007, 35: 182. DOI: 10.1093/nar/gkm321.
- [9] FINN R D, TATE J, MISTRY J, et al. The Pfam protein families database[J]. Nucleic Acids Research, 2008, 36: 281. DOI: 10.1093/nar/gkh121.
- [10] GOTZ S, GARCIAGOMEZ J M, TEROL J, et al. High-throughput functional annotation and data mining with the Blast2GO suite[J]. Nucleic Acids Research, 2008, 36(10): 3420. DOI: 10.1093/nar/gkn176.
- [11] LI B, DEWEY C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome[J]. BMC Bioinformatics, 2011, 12(1): 323. DOI: 10.1186/1471-2105-12-323.
- [12] 王斌, 孙静, 刘凌云, 等. 蛭类转录组中EST-SSR分析及抗凝血相关分子标记的挖掘[J]. 中草药, 2017, 48(1): 172. DOI: 10.7501/j.issn.0253-2670.2017.01.024.
- [13] 孟玮, 蒋艳琳, 张林, 等. 基于RNA-seq技术的江鳕转录组SSR位点信息分析[J]. 淡水渔业, 2019, 49(6): 10. DOI: 10.13721/j.cnki.dsyy.2019.06.002.
- [14] 陈松波, 龚丽, 刘海金. 牙鲆EST资源的SSR信息分析[J]. 东北农业大学学报, 2010, 41(10): 82. DOI: 10.19720/j.cnki.issn.1005-9369.2010.10.016.
- [15] 刘洪涛, 刘金叶, 杨明秋, 等. 基于转录组测序的波纹唇鱼SSR和SNP多态特征分析[J]. 基因组学与应用生物学, 2020, 39(6): 2451. DOI: 10.13417/j.gab.039.002451.
- [16] 王且鲁, 刘奕, 宋红梅, 等. 双须骨舌鱼转录组EST-SSR标记开发与引物筛选[J]. 淡水渔业, 2016, 46(6): 8. DOI: 10.13721/j.cnki.dsyy.2016.06.002.
- [17] FRICKE C, WIGBY S, HOBBS R, et al. The benefits of male ejaculate sex peptide transfer in *Drosophila melanogaster*[J]. Journal of Evolutionary Biology, 2009, 22(2): 275. DOI: 10.1111/j.1420-9101.2008.01638.x.
- [18] WIGBY S, CHAPMAN T. Sex peptide causes mating costs in female *Drosophila melanogaster*[J]. Current Biology, 2005, 15(4): 316. DOI: 10.1016/j.cub.2005.01.051.
- [19] SANDRA G, NORMA M. Sexual determination and differentiation in teleost fish[J]. Reviews in Fish Biology and Fisheries, 2010, 20(1): 101. DOI: 10.1007/s11160-009-9123-4.
- [20] HATTORI R S, STRUSSMANN C A, FERNANDINO J I, et al. Genotypic sex determination in teleosts: insights from the testis-determining amhy gene[J]. General and Comparative Endocrinology, 2013, 192: 55. DOI: 10.1016/j.ygcen.2013.03.019.
- [21] MARTINEZ P, VINAS A, SANCHEZ L, et al. Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture[J]. Frontiers in Genetics, 2014, 5: 340. DOI: 10.3389/fgene.2014.00340.

责任编辑: 何承刚