

蛭类转录组中 EST-SSR 分析及抗凝血相关分子标记的挖掘

王 斌¹, 孙 静², 刘凌云¹, 乔永锋³, 程 霞¹, 仝向荣¹, 王德斌^{1*}

1. 昆明学院 云南省高校特色生物资源开发与利用重点实验室, 云南 昆明 650224

2. 北京协和医学院 中国医学科学院医学生物学研究所, 云南 昆明 650118

3. 昆明学院 化学科学与技术系, 云南 昆明 650224

摘要:目的 对蛭类转录组进行 SSR 分析, 是开发功能分子标记、开展对药用蛭类资源的开发利用及进行遗传背景监测的必要手段。方法 利用与医学密切相关的医蛭科棒纹牛蛭 *Poecilobdella javanica* 和日本医蛭 *Hirudo nipponia*、黄蛭科宽体金线蛭 *Whitmania pigra* 和山蛭科洞穴山蛭 *Heamadipsa cavatuses* sp.nov. 共 14 个样本, 通过 RNA 测序及转录组分析, 筛选吸血差异 unigene, 并比较 4 个种的 SSR 标记特征和分布, 用于挖掘抗凝血相关功能 EST-SSR。结果 通过 SSR 密度分析发现, SSR 在棒纹牛蛭、日本医蛭、宽体金线蛭和洞穴山蛭转录组中的间隔长度为 4 723、6 026、8 059、8 144 bp; SSR 组成分析发现三碱基重复的基序占蛭类 SSR 类型组成中绝大多数; 另外, 洞穴山蛭的 SSR 组成与其他 3 种存在明显区别。比较 4 种吸血蛭吸血前后转录组中表达丰度显著差异的 unigene 群, 发现 4 种蛭中包含 SSR 且抗凝血相关的重叠 unigene 类群在转录组中呈较高的片段多态性。结论 unigene 片段长度、SSR 的密度、组成和分布特点能反映物种之间的分类关系; 4 种蛭重叠的抗凝血相关分子功能对应的 EST-SSR, 可作为抗凝血相关的功能分子标记候选 EST-SSR。

关键词: EST-SSR; 抗凝血相关基因; EST-SSR 分布; 功能分子标记; 药用蛭类

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2017)01-0172-07

DOI: 10.7501/j.issn.0253-2670.2017.01.024

Analysis on EST-SSR in transcriptome of medicinal leeches and mining of anticoagulant related functional molecular markers

WANG Bin¹, SUN Jing², LIU Ling-yun¹, QIAO Yong-feng³, CHENG Xia¹, TONG Xiang-rong¹, WANG De-bin¹

1. Key Laboratory of Special Biological Resource Development and Utilization of Universities of Yunnan Province, Kunming University, Kunming 650224, China

2. Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming 650118, China

3. Department of Chemical Science and Technology, Kunming University, Kunming 650224, China

Abstract: Objective To expand the exploitation and genetic diversity conservation of medicinal leech resources, the effective functional molecular markers for identification are needed to be developed by transcriptome-SSR analysis. **Methods** Totally 14 samples including *Poecilobdella javanica* and *Hirudo nipponia* of Hirudinidae, *Whitmania pigra* *Whitman* of Haemopidae and *Heamadipsa cavatuses* sp. nov. of Haemadipsidae were used in transcriptomic analysis by RNA-sequencing, the simple sequence repeats (SSR) distribution characteristics of four species were compared, and transcriptomic comparison of hematophagia samples and hungry samples were carried out to investigate the anticoagulation involved unigene. Finally, the data in two groups were combined to mining functional EST-SSR involved in anticoagulation. **Results** The SSR density analysis showed the gap lengths of SSR in transcriptome of *Poecilobdella javanica*, *Hirudo nipponia*, *Whitmania pigra* *Whitman*, and *Heamadipsa cavatuses* were 4 723, 6 026, 8 059, and 8 144 bp. In the four species, trinucleotide repeats occupied overwhelming majority in composition of EST-SSR; On the whole, distribution of SSR types can distinct from each other, and *Heamadipsa cavatuses* was particularly different from others. Moreover, we found that the anticoagulation involved unigene group containing SSR had high fragment-polymorphism. **Conclusion** For the four leech species, the characteristics of density, composition, and distribution of EST-SSR can reflect its classification relationship. Meanwhile, the EST-SSRs in unigenes which involved in anticoagulation and overlapped in four leech species are ideal fragment group for screening anticoagulation functional marker.

Key words: EST-SSR; anticoagulation involved genes; EST-SSR distribution; functional marker; medicinal leech

收稿日期: 2016-07-19

作者简介: 王 斌 (1976—), 男, 博士, 副教授, 研究方向为生物化学与分子生物学。Tel: 13888824773 E-mail: wangbbiomed@126.com

*通信作者 王德斌 (1959—), 教授。Tel: 13108898255 E-mail: wangdebin1202@sina.com

蛭类 (leeches) 很早就被应用于医蛭疗法 (hirudotherapy) 和作为中药用于治疗心脏疾病、妇科疾病、泌尿疾病、外科手术、口腔疾病及眼科疾病^[1]。世界上有 600 余种蛭类, 中国有 100 余种。但是被 FDA 正式批准的治疗水蛭仅欧洲医蛭 *Hirudo medicinalis* 一种^[2]; 《中国药典》2015 年版中规定水蛭药材源动物仅包含水蛭科水蛭、蚂蟥、柳叶蚂蟥。李军德等^[3]对蛭类药材来源考证认为《中国药典》中规定的此 3 种水蛭正式中文名为日本医蛭 *Hirudo nipponia*, 归属于医蛭科 Hirudinidae 医蛭属 *Hirudo* Linnaeus; 蚂蟥正式中文名为宽体金线蛭 *Whitmania pigra* Whitman、柳叶蚂蟥正式中文名为尖细金线蛭 *Whitmania acranulata* Whitman, 二者归属于黄蛭科 (Haemopidae) 金线蛭属 *Whitmania* Whitman。并建议扩大水蛭药材来源包含: 菲牛蛭 *Poecilobdella amanillensis* Lesson、或山蛭科动物天目山蛭 *Haemadipsa tianmushana* Song 的。随着心血管等疾病对蛭类药材资源需求的日益增长, 扩大蛭类用药资源仍然迫切。分布在云南等中国西南的医蛭科牛蛭属棒纹牛蛭 *Poecilobdella javanica* Wahlberg^[4], 以及蝙蝠血为食物的陆生山蛭类新种洞穴山蛭 *Heamadipsa cavatuses* sp. nov.^[5], 是抗凝血活性极强的吸血蛭, 据中药中蛭类使用原则推测^[6]和《中国药典》对药用蛭质量鉴定的抗凝血酶活性标准, 棒纹牛蛭和洞穴山蛭是具有潜在药用价值的动物资源, 其吸血过程中唾液包含大量抗凝血功能物质, 可发展为治疗心血管疾病及脑部血栓等疾病药物。

由于水蛭具有活血化瘀等功效, 在中药中主要用于治疗心血管疾病, 《中国药典》2010 年版规定利用抗凝血酶活性来定义水蛭的效价。但是由于水蛭药用成分中发挥抗凝血功能的物质远比抗凝血酶复杂得多^[7], 吸血蛭类中吸血相关的大量功能物质都可能参与蛭类抗凝血过程。因而以凝血酶为靶点不能有效控制水蛭临床应用的有效性 & 安全性, 更不能作为药用蛭资源鉴定和扩大的标准。

本实验利用 RNA-Seq 技术对宽体金线蛭、棒纹牛蛭、洞穴山蛭、日本医蛭进行测序并从头组装分别得到各个物种的转录组数据, 通过 MISA 软件分析各转录组数据获得的各物种的 EST-SSR 信息。并比较各种物 EST-SSR 在转录组中的分布特点及规律; 另一方面, 对 4 种蛭类中相同的抗凝血分子基础进行重叠分析; 综合 2 部分数据挖掘 4 种蛭中抗凝血功能特有的通用功能 SSR 标记, 为蛭类药用资

源鉴定和遗传监测、质量控制的分子标记开发提供参考数据。

SSR 序列一般是指由 1、2、3、4、5 和 6 核苷酸重复单元串联排列的简单重复序列, 其广泛分布于各类生物基因组内^[8-9]。对于不同的生物而言, 其基因组内 SSR 的密度、类型及其分布区域都存在较大的差异^[10]。大量研究表明, 基因不同区域存在的 SSR 序列对物种的进化、基因表达的调节以及基因翻译产生的蛋白功能都有一定的影响^[11-13], 因此 SSR 可以作为特定功能的分子标记。

1 材料与方法

1.1 材料

所用蛭类样本包括棒纹牛蛭、日本医蛭、洞穴山蛭和金线蛭均采自云南省西南地区德宏州。其中棒纹牛蛭在处理时按其习性分别设置饥饿状态、吸血后 30 min、10 d、20 d、30 d 共 8 个样本; 洞穴山蛭包括饥饿状态、吸血后 30 min、10 d 共 3 个样本; 日本医蛭包含饥饿和吸血 30 min 2 个样本 (表 1)。所有样本采回后切开, 并用去离子水反复清洗体表及体腔, 然后用 75% 无水乙醇再次清洗后立即提取 RNA 用于后续分析。

表 1 用于转录组分析的各物种样品数量

Table 1 Number of samples prepared for transcriptomic analysis

不同状态	棒纹牛蛭	洞穴山蛭	日本医蛭	宽体金线蛭
饥饿	2	1	1	1
吸血后 30 min	1	1	1	
吸血后 10 d	2	1		
吸血后 20 d	2			
吸血后 30 d	1			

1.2 RNA 提取及转录组分析准备

在进行 Illumina 测序前, 先提取总 RNA, 用 Agilent 2100 Bioanalyzer RNA Nano chip 进行质量监控, 确保总 RNA 质量浓度 ≥ 400 ng/ μ L, A_{260}/A_{280} 在 1.8~2.2, RNA 28 S: 18 S ≥ 1 , 且 RNA 完整度指数 RNA Integrity Number (RIN) ≥ 7.0 。接着用 NEBNext Poly (A) mRNA Magnetic Isolation Module (NEB, E7490) 富集 mRNA, 并 NEBNext mRNA Library Prep Master Mix Set for Illumina (NEB, E6110) 和 NEBNext Multiplex Oligos for Illumina (NEB, E7500) 试剂盒构建测序文库, 用 1% 琼脂糖凝胶电泳检测合格后, 用 Library Quantification Kit-Illumina GA Universal (Kapa, KK4824) 试剂盒

进行 qPCR 定量检测, 检测合格的文库用 Illumina cbot 进行聚类, 之后用 Illumina HiSeq™ 2500 (Biomarker biotechnological Co. Beijing) 进行测序。

1.3 Illumine reads 加工及从头组装

经测序得到的原始 pair-end read 通过质量分数、碱基分布、数据评估和统计等校验后, 进一步滤掉接头、未识别的核酸和带模糊序列“N”的低质量 reads, 最终获得的高质量 reads 用 Trinity^[7] 进行各个物种分别从头组装 (3 个物种均无参考基因组信息)。所得到的 unigene 进行下一步的注释和分析。

1.4 基因注释和 SSR 分析

对所得 unigene 用 BLAST 软件 (E 值 $<1 \times 10^{-5}$) 以下公共蛋白数据库中进行同源查找, 按相似基因进行功能注释。包括: NCBI 非冗余数据库 (<http://www.ncbi.nlm.nih.gov/>)、Swiss-Prot 蛋白数据库 (<http://www.expasy.ch/sprot/>)、COG (Clusters of Orthologous Groups) 数据库 (<http://www.ncbi.nlm.nih.gov/cog/>) 和 KEGG 数据库 (<http://www.genome.jp/keg/>)。GO 功能分配和分类分别用 Gene Ontology (GO) (<http://www.geneontology.org/>) 和 WEGO (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>)。利用 MISA 软件筛选到的 1 kb 以上的 unigenes 做 SSR 分析。通过 SSR 简单重复序列的查找得到单核苷酸重复次数在 10 次或 10 次以上, 二核苷酸重复次数在 6 次或 6 次以上, 2~6 核苷酸重复次数在 5 次或 5 次以上的完全重复 SSR 以及中间被少数碱基 (间隔 ≤ 100) 打断的不完全重复的 SSR。所有得到的 SSR 序列通过 EXCEL 分析, 得到本实验结果。参考已有的研究分析^[9], 本实验将所有可循环序列及与其互补序列归为一类, 将单核苷酸重复序列归为 A、C 2 类; 二核苷酸重复序列归为 AC、AG、AT 和 CG 4 类; 3 核苷酸重复序列归为 10 类; 4 核苷酸重复序列归为 16 类; 5 核苷酸重复序列归为 50 类; 6 核苷酸重复序列可归为 160 类。如 3 核苷酸重复序列中, AAG 基序代表了所有 AAG、AGA、GAA、CTT、TCT 和 TTC 的 SSR 序列。随后利用 blast+ 算法将拼接得到的非重复序列基因与 CDD 库中的 COG 库进行比对, 对饥饿洞穴山蛭含 SSR 的不同非重复序列基因进行 COG 功能分类及分析。

2 结果与分析

2.1 4 种蛭类转录组测序及组装

对宽体金线蛭 (1 个样本)、棒纹牛蛭 (吸血前及吸血后不同时间段的 8 个样本)、洞穴山蛭 (吸

血前及吸血后不同状态下的 3 个样本) 及日本医蛭 (吸血前及吸血后 2 个样本) 各样本分别构建 cDNA 文库后, 用 Illumina HiSeq™ 2500 进行双末端 RNA 测序。原始双末端 reads 经数据评估后, 滤掉低质量的数据和 rRNA, 并用 Trinity 软件分别对不同物种的各个样品数据分别进行混合组装, 获得对应的转录本序列 (Transcripts), 在聚类单元中选取最主要的转录本作为 unigenes 序列, 用 cd-hit 聚类去冗余, 采用 TGICL 聚类组装策略最终得到 unigene 文库。金线蛭 cDNA 总共获得 5.29 Gb 数据, 去除低质量数据后使用 4.23 Gb 数据; 棒纹牛蛭 8 个样本测序总共数据量为 33.6 Gb, 使用 28.3 Gb; 洞穴山蛭总共获得 14.95 Gb 数据, 除去低质量数据后使用 12.55 Gb; 日本医蛭则获得和使用 14.58 Gb 数据。除日本医蛭一组的 3 个样本 Q30 (测序错误概率为 0.01) 值为 88% 外, 其余样本均大于 93%。最终棒纹牛蛭组装获得 57 410 个 unigene, 平均长度 1 413 bp; 洞穴山蛭组装获得 37 465 个 unigene, 平均长度 1 362 bp; 日本医蛭组装获得 39 812 个 unigene, 平均长度 1 012 bp; 宽体金线蛭组装获得 22 875 个 unigene, 平均长度 1 214 bp (表 2)。从 4 种蛭类片段分布长度来看, 日本医蛭与其他 3 类蛭有明显差异, 其 200~300 bp 短片段分布明显高于长片段分布, 而其他 3 种蛭类正好相反。

2.2 4 种蛭 SSR 分布特点及分析

利用 MISA 软件对筛选到的 1 kb 以上的 unigenes 进 SSR 分析。结果从棒纹牛蛭 unigene 文库选出 14 262 个 SSR, SSR 的频率为 35.6%, 平均每隔 4 723 bp; 洞穴山蛭文库中共筛到 5 150 个 SSR, SSR 的频率为 24.9%, 平均间隔 8 144 个 unigene; 相同的参数在日本医蛭和宽体金线蛭中分别是 SSR 总数 4 451 个和 2 698 个, SSR 频率分别为 25.2% 和 22.3%, 平均间隔长度分别为 6 026 bp 和 8 059 bp。从 SSR 分布密度特点来看, 棒纹牛蛭 unigene 中 SSR 出现的频率较高, 平均间隔最短; 其后依次是日本医蛭、宽体金线蛭和洞穴山蛭 (表 3)。日本医蛭 unigene 文库中由于低于 1 kb 片段较多 (表 2), 因此该结果可能低估了日本医蛭中 SSR 分布频率; 另外, 棒纹牛蛭由于测序所用的 8 个样本能涵盖更广泛的 mRNA, 因此获得更多的表达基因, 较大的数据本底可能是导致检出高频率的 SSR 原因之一。

表 2 4 种蛭中 RNA 测序后 unigene 的数量及分布

Table 2 Number and distribution of assembled unigenes in four leech species

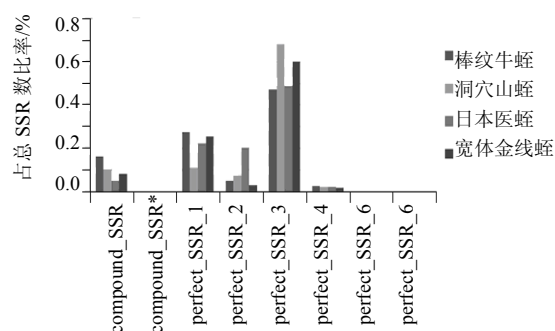
品种	长度/bp					总数	总长度/bp	N50 长度/bp	平均长度/bp
	200~300	300~500	500~1 000	1 000~2 000	>2 000				
棒纹牛蛭	14 185	10 160	9 069	10 095	13 895	57 410	81 132 096	2 749	1 413.2
洞穴山蛭	8 465	6 903	6 146	7 306	8 643	37 465	51 053 993	2 529	1 362.7
日本医蛭	13 569	8 504	6 146	5 564	6 029	39 812	40 288 201	2 160	1 012.0
宽体金线蛭	4 835	4 426	4 407	4 895	4 312	22 875	27 783 880	2 112	1 214.6

表 3 4 种蛭中 SSR 在 unigene 中密度分布比较

Table 3 Comparision of SSR density distribution in unigene of four leeches

品种	总 SSR	含 SSR unigene 个数							检测序列 总数	检测序列长度/bp	包 SSR 的 unigene 比率/%	SSR 平均间隔长度/bp
		2	3	4	5	6	7	总数				
棒纹牛蛭	14 262	1 956	618	241	59	19	17	8 548	23 990	67 361 719	35.60	4 723
洞穴山蛭	5 150	642	111	11	3	0	0	3 968	15 949	41 940 694	24.90	8 144
日本医蛭	4 451	568	121	27	3	0	0	3 535	14 015	26 822 526	25.20	6 026
宽体金线蛭	2 698	304	75	17	3	0	0	2 057	9 207	21 743 063	22.30	8 059

进一步对 4 种蛭中 SSR 基序组成类型分布分析发现, 4 种蛭中单碱基重复, 二碱基重复和三碱基重复占整个 SSR 类型的绝大多数, 其中三碱基重复占绝大多数, 体现了蛭类 SSR 碱基重复的共同规律。对 4 种蛭中 SSR 类型分布比率作图发现, 在 4 种蛭中同时占比最高的单碱基类型 (11%~27%) 和三碱基类型 (47%~68%) 中, 洞穴山蛭明显不同于其余 3 种, 其单碱基比率远低于其余 3 种, 而三碱基比率远高于其余三者。进一步证实了洞穴山蛭的 SSR 分布规律与其余 3 种蛭明显不同 (图 1)。



perfect_SSR_1、perfect_SSR_2、perfect_SSR_3、perfect_SSR_4、perfect_SSR_5、perfect_SSR_6 分别为 1~6 碱基重复的 SSR 类型; compound_SSR 为混合类型的 SSR; compound_SSR* 为混合导致的不确定的 SSR

perfect_SSR_1, perfect_SSR_2, perfect_SSR_3, perfect_SSR_4, perfect_SSR_5, perfect_SSR_6 represent mononucleotide to hexanucleotide repeat respectively; compound_SSR* represent uncertain SSR because of mixture

图 1 不同碱基类型的 SSR 类型所占比率比较

Fig. 1 Comparision of ratios of SSR type for four leeches

进一步对三碱基类型的 SSR 的不同基序深入分析发现, AAC/GTT、AAT/ATT、ATC/GAT 3 个基序所占比例在 4 个蛭中最大, 针对这 3 个基序类型作图发现, 洞穴山蛭所占比率明显不同于其余三者, 尤其 AAC/GTT 明显低于其余三者, 在 ATC/GAT 基序明显高于其余三者 (图 2)。

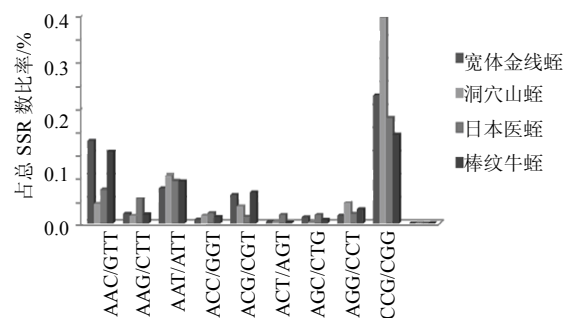


图 2 三碱基重复的 SSR 中各类基序在 4 种蛭中的比率分布

Fig. 2 Distribution of ratios of all trinucleotide repeat motif in four leeches

从 4 个蛭种的 SSR 分布密度及组成特点可以发现, 医蛭科牛蛭属的棒纹牛蛭与日本医蛭的 SSR 分布密度最为相近, 但在 SSR 组成上黄蛭科金线蛭属的宽体金线蛭和医蛭属的棒纹牛蛭最为相近; 无论从密度和组成来看, 山蛭科的洞穴山蛭与其余三者相对差异明显。由于日本医蛭可能由于测序时 cDNA 片段质量因素, 因此在 SSR 组成与棒纹牛蛭差异较大。因此综合 SSR 分布密度和组成来看, 棒纹牛蛭和日本医蛭最近, 其次是宽体金线蛭, 最远的是洞穴山蛭。SSR 分布规律与蛭类分类关系完全一致。

2.3 抗凝血相关分子标记的挖掘

药用蛭类主要用于治疗心血管疾病及脑部血栓等疾病,从分子机制上,蛭类药理作用主要与抗凝血相关分子相关,吸血蛭类都具有丰富的抗凝血相关分子,而非吸血蛭宽体金线蛭同样具有极强的抗凝血活性。因此可从 4 种蛭中挖掘抗凝血相关的分子标记。为了确定 4 种蛭中抗凝血的分子是否相似,将棒纹牛蛭、洞穴山蛭、日本医蛭吸血前后的 unigene 进行比对,获得吸血前后 RPKM 值 (Reads Per Kilobase per Million mapped reads, 反映对应 unigene 的表达丰度) 显著差异的 unigene 分别为:棒纹牛蛭 1 187 个,洞穴山蛭 897 个,日本医蛭 716 个。经 swissprot 蛋白数据库注释后,分别获得无冗余的功能 597、486、347 个。功能比对发现,在这 3 类吸血差异基因中,棒纹牛蛭与洞穴山蛭有 169 个相同、洞穴山蛭与日本医蛭有 124 个相同、日本医蛭与棒纹牛蛭有 143 个相同 (表 4)。虽然金线蛭为非吸血蛭,但其转录组中注释得到的 1 107 个功能中,仍有 178、145、111 个分别与棒纹牛蛭、洞穴山蛭和日本医蛭的抗凝血相关分子功能相同 (图 3)。该结果表明 4 种蛭中有共同的抗凝血相关分子基础。其中宽体金线蛭和日本医蛭抗凝血功能分子最为接近,随后是洞穴山蛭和棒纹牛蛭。有趣的是

表 4 4 种蛭之间抗凝血功能分子的重叠数据

品种	棒纹牛蛭	宽体金线蛭	洞穴山蛭	日本医蛭
棒纹牛蛭	597			
宽体金线蛭	178	1 107*		
洞穴山蛭	169	145	486	
日本医蛭	143	111	124	347

*表示宽体金线蛭转录组中注释的所有功能分子,区别于其余 3 个吸血蛭中吸血前后差异的吸血相关功能分子

*indicate number of total annotated unigenes in *Whitmania pigra* Whitman, distinct from the blood-sucking involved unigene in other three leeches.

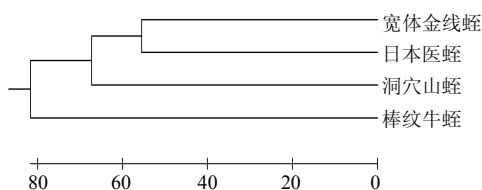


图 3 4 种蛭类中抗凝血相关功能分子重叠分析

Fig. 3 Overlap analysis for number of anticoagulating unigenes in four kinds of leeches

抗凝血相关功能分子并不因为洞穴山蛭亲缘关系较远而出现较大差异,相对于医蛭科中棒纹牛蛭,洞穴山蛭在吸血相关的抗凝功能分子上与作为药用蛭的宽体金线蛭和日本医蛭更相似。这与实际应用情况相似,即宽体金线蛭和日本医蛭都作为中药使用,而山蛭中很早就发现山蛭素^[14]而被广泛应用,相对而言牛蛭的应用较少。由此推测从 4 种蛭类相同抗凝相关差异分子能较好地反映水蛭应用于心血管等疾病治疗的情况,因此从中筛选 SSR 标记,可以较好地反映蛭类的药用价值。

基于以上推测,进一步进行重叠分析得到 4 种蛭中具有相同的抗凝血相关分子功能 48 个,每个功能均来自 unigene 在 Swissprot 数据库的序列相似性比对结果,因而每个注释功能在 4 种蛭中具有数量不等的包含 SSR 的 unigene 片段。每个功能在每个物种中对应的含 SSR 的 unigene 片段越多,意味着等位基因越丰富,即 SSR 滑动突变导致的基因越丰富,因此 SSR 分子标记多态性越高,按照该假设筛选到 22 个功能分子,包括 Uncharacterized histidine-rich protein DDB_G0274557 (Precursor), Putative leucine-rich repeat-containing protein DDB_G0290503, ADAM 17-like protease (Precursor), Ankyrin, Epidermal growth factor receptor substrate, Filamin-A, Serine/threonine-protein kinase par-1 Low-density lipoprotein receptor-related protein (Precursor), A disintegrin and metalloproteinase with thrombospondin motifs gon-1 (Precursor) 等抗凝血相关 Swissprot 注释功能,其中等位基因片段最丰富的 2 个,分别为 histidine-rich protein 和 leucine-rich repeat-containing protein,均为 SSR 三碱基基序重复导致的组氨酸重复序列和亮氨酸重复序列。以 SSR 出现频率最低的宽体金线蛭为本底,获得 22 个功能对应的 164 个主要以包含 ACC/GTT、AAT/ATT、ATC/GAT 基序的 SSR 及其对应的 137 个 unigene,作为开发药用蛭类质量鉴定和监测等分子标记的 EST-SSR 片段。

3 讨论

3.1 三碱基重复基序在蛭类中 EST-SSR 组成中占绝大多数

通过对蛭类 3 科 4 属 4 个物种 14 个样本的转录组测序分析,发现蛭类 EST-SSR 分布特点极为相似,三碱基基序出现频率接近或超过 50%,其次是单碱基 10%~30%和少量的二碱基。SSR 碱基基序的分布在不同物种中有不同的特点。

EST-SSR 分析发现鸡中 90%以上为单碱基和二碱基的 SSR 序列^[15]；山羊中 EST-SSR 分析结果发现二碱基、三碱基和六碱基出现的比例最高，其中二碱基比例最高^[16]，另武昌鱼 *Megalobrama amblycephala* Yih 和黄鱼 *Larimichthys polyactis* Bleeker 转录组分析表明二碱基重复达绝大多数，其次是占较小比例的三碱基^[17]；可见三碱基基序在 SSR 类型中占较大比例是蛭类 SSR 分布的重要特点。

3.2 EST-SSR 在蛭类中密度和分布反映物种间亲缘关系

从 4 种蛭类 SSR 分布特点可以发现洞穴山蛭在 EST-SSR 密度和组成分布特点上与医蛭科其他 3 种区别较大，而金线蛭又可与日本医蛭和棒纹牛蛭区别。SSR 的密度及组成分布特点甚至可以区分科间、属间关系。通过基因组中 SSR 分布分析发现，SSR 基序分布特点各物种有所不同，如按基序所占比例的先后顺序排序的话，人基因组中顺序是一、四、二，即中 SSR 分布主要是单碱基重复、其次是四碱基和二碱基重复；果蝇 *Drosophila melanogaster* Meigen 基因组中 SSR 基序比例分别是二、三、一、四碱基重复；秀丽隐杆线虫 *Caenorhabditis elegans* Bergerac 碱基基序比例顺序是一、二、三、四；拟南芥 *Arabidopsis thaliana* (L.) Heynh. 基因组中 SSR 基序比例也是一、二、三、四，但一、二、三占绝大部分；酿酒酵母 *Saccharomyces cerevisiae* Meyen ex E.C. Hansen 则为一、三、二、四^[18]。由此可见不同的物种可以按 SSR 基序的分布特点加以区分。SSR 作为 DNA 指纹图谱和多样性研究的理想的遗传标记，主要因为微卫星位点存在的片段长度多态性^[19-20]。同时，大量研究表明 SSR 密度、组成及分布特点在不同物种间都有所区别^[21-23]。基于抗凝血共同分子基础筛选的 SSR 标记可以作为医蛭药用资源鉴定的标记。

3.3 抗凝血关联的功能 SSR 候选 unigene

在基因组中，SSR 可以出现在编码区，也可以分布在非编码区。SSR 出现在编码区域时，就会导致编码蛋白的随机重复^[24]，许多研究认为编码区中包含 SSR 的基因群，可能是进化速度较快的蛋白^[25-26]，这些 SSR 具有与染色体稳定、RNA 折叠、氨基酸重复或人类疾病、重组热点、转座子相关的生物学功能^[18]。功能 SSR 通常出现在应对选择压力的基因群中，如包含 SSR 的 EST 序列在肺癌组织中具有不同的氨基酸组成和蛋白表达模式^[28]。以吸血为主的

蛭类尤其是医蛭科和山蛭科种属，抗凝血需求是其进化中的主要进化压力。

4 种蛭中包含 SSR 的 unigene 有很大部分与抗凝血相关基因重叠。以等位基因多态性进行排序获得的前 22 个功能分子几乎与抗凝血机制相关联。其中出现的多态性最高的 2 个分别为富含组氨酸重复和亮氨酸重复的蛋白，可能是直接由三碱基的 SSR 重复序列编码的结果。这些迹象显示抗凝血机制关联的包含 SSR 的 unigene 中，可能是筛选抗凝血功能 SSR 的 unigene 类群。

参考文献

- [1] Baskova I P, Kostjukova E S, Vlasova M A, et al. Proteins and peptides of the salivary gland secretion of medicinal leeches *Hirudo verbana*, *H. medicinalis*, and *H. orientalis* [J]. *Biochemistry (Mosc)*, 2008, 73(3): 315-320.
- [2] 鞠利雅, 法国蚂蟥获准 FDA 医用许可证书 [J]. 中国中药杂志, 2004, 29(10): 1022-1023.
- [3] 李军德, 黄璐琦, 唐仕欢, 等. 《中国药典》2010 年版一部部分动物药材来源探讨 [J]. 中国中药杂志, 2010, 35(16): 2051-2055.
- [4] 王德斌, 全向荣, 王 斌, 等. 云南省蛭类动物及其区系的重要性 [J]. 昆明学院学报, 2009, 31(6): 49-51.
- [5] 杨 潼, 莫 潇, 王德斌, 等. 洞穴吸血陆蛭一新种在中国云南省西陲的发现 [J]. 动物分类学报, 2009, 34(1): 125-129.
- [6] 吕 莉, 李欣燕, 韩国柱, 等. 重组水蛭素体内抗凝活性的实验研究 [J]. 中草药, 2002, 33(6): 520-522.
- [7] Corral R M, Macedo R S, Pereira P J, et al. Leech-derived thrombin inhibitors: from structures to mechanisms to clinical applications [J]. *J Med Chem*, 2010, 53(10): 3847-3861.
- [8] Emmanuel B, Neil J. The rise, fall and renaissance of microsatellites in eukaryotic genomes [J]. *BioEssays*, 2006, 28(10): 1040-1050.
- [9] Gur-Arie R, Cohen C J, Eitan Y, et al. Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism [J]. *Genome Res*, 2000, 10(1): 62-71.
- [10] Li Y C, Korol A B, Fahima T, et al. Microsatellites within genes: structure, function, and evolution [J]. *Mol Biol Evol*, 2004, 21(6): 991-1007.
- [11] Verstrepen K J, An J, Lewitter F, et al. Intragenic tandem repeats generate functional variability [J]. *Nature Genetics*, 2005, 37: 986-990.
- [12] Rocha E P, Blanchard A. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution [J].

- Nucleic Acids Res*, 2002, 30(9): 2031-2042.
- [13] Li Y, Korol A B, Fahima T, *et al*. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review [J]. *Mol Ecol*, 2002, 11(12): 2453-2465.
- [14] Strube K H, Kröger B, Bialojan S, *et al*. Isolation, sequence analysis, and cloning of haemadin. An anticoagulant peptide from the Indian leech [J]. *J Biol Chem*, 1993, 268(12): 8590-8595.
- [15] Bakhtiarizadeh M R, Arefnejad B, Ebrahimie E, *et al*. Application of functional genomic information to develop efficient EST-SSRs for the chicken (*Gallus gallus*) [J]. *Genet Mol Res*, 2012, 11(2): 1558-1574.
- [16] 赵传志, 李爱芹, 李长生, 等. 山羊 EST 资源的 SSR 分析 [J]. 山东农业科学, 2009(6): 6-9.
- [17] Gao Z, Luo W, Liu H, *et al*. Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*) [J]. *PLoS One*, 2012, 7(8): e42637.
- [18] Mukund V K, Prabhakar K R, Vidya S G. Differential distribution of simple sequence repeats in *Eukaryotic genome* sequences [J]. *Mol Biol Evol*, 2001, 18(7): 1161-1167.
- [19] Beckmann J S, SOLLER M. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites [J]. *Biotechnology*, 1990, 8: 930-932.
- [20] Morgante M, Olivieri A M. PCR-amplified microsatellites as markers in plant genetics [J]. *Plant J*, 1993, 3: 175-182.
- [21] Galindo C L, Mciver L J, McCormick J F, *et al*. Global microsatellite content distinguishes humans, primates, animals, and plants [J]. *Mol Biol Evol*, 2009, 26: 2809-2819.
- [22] Mayer C, Leese F, Tollrian R. Genome-wide analysis of tandem repeats in *Daphnia pulex*-a comparative approach [J]. *BMC Genomics*, 2010, 11(19): 277-285.
- [23] Pannebakker B, Niehuis O, Hedley A, *et al*. The distribution of microsatellites in the *Nasonia parasitoid* wasp genome [J]. *Insect Mol Biol*, 2010, 19(Suppl 1): 91-98.
- [24] Katti M V, Sami-Subbu R, Ranjekar P K. Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications [J]. *Protein Sci*, 2000, 9(5): 1203-1209.
- [25] Green H, Wang N. Codon reiterations and the evolution of proteins [J]. *Proc Natl Acad Sci USA*, 1994, 91: 4298-4302.
- [26] Huntley M, Golding G B. Evolution of simple sequence in proteins [J]. *J Mol Evol*, 2000, 51: 131-140.
- [27] Bakhtiarizadeh M R, Ebrahimi M, Ebrahimie E. Discovery of EST-SSRs in lung cancer: Tagged ESTs with SSRs lead to differential amino acid and protein expression patterns in cancerous tissues [J]. *PLoS One*, 2011, 6(11): e27118.