

DOI: 10.12264/JFSC2023-0174

虾肝肠胞虫全基因组 SSR 位点分布及比较分析

全德润^{1,2}, 王李宝¹, 黎慧¹, 史文军¹, 朱健强¹, 顾舒文¹, 顾晨¹, 任乾², 万夕和¹

1. 江苏省海洋水产研究所, 江苏 南通 226007;

2. 南京师范大学海洋科学与工程学院, 江苏 南京 210034

摘要: 为开发虾肝肠胞虫(*Enterocytozoon hepatopenaei*, EHP)及其近缘种的分子标记, 本研究采用生物信息学方法, 对虾肝肠胞虫全基因组中 SSR (simple sequence repeat, 简单重复序列)位点的数量、分布及频率等信息进行统计分析, 并与其他 5 种微孢子虫真菌物种基因组进行了比较。在虾肝肠胞虫基因组中共搜索到 794 个 SSR, 总长度为 21773 bp, 占基因组的 0.72%, SSR 的平均长度为 27.43 bp, 相对丰度为 262 loci/Mb, 相对密度为 7184.43 bp/Mb。6 种 SSR 类型中, 四核苷酸 SSR 数量最多, 有 570 个, 总长度为 12152 bp, 占 SSR 总数的 71.79%, 平均长度为 21.32 bp, 相对丰度为 188.08 loci/Mb, 相对密度为 4009.79 bp/Mb; 二核苷酸 SSR 数量最少, 仅有 20 个, 总长度为 280 bp, 占 SSR 总数的 2.52%, 平均长度为 14 bp, 相对丰度为 6.6 loci/Mb, 相对密度为 92.39 bp/Mb。基因组高频率(≥ 10)出现的 SSR 重复类型为: AACT, TTAG, TAGT, ACTA, AGTT, CTAA, AACTA, GTTA, TTAGT, TAAC, AT。选择序列长度不小于 20 bp 的多核苷酸 SSR 共计 30 个, 进行引物设计用于物种鉴定。利用 GO 数据库对含 SSR 基因进行功能注释, 共有 28 个基因(3.53%)注释到分子功能、生物进程及细胞组分 3 类中。将虾肝肠胞虫与其他微孢子虫真菌相比, 虾肝肠胞虫中 SSR 的数量及相对丰度适中, 由此可知基因组大小与 SSR 数量无关。本研究为虾肝肠胞虫的种群遗传及进化相关研究奠定了基础。

关键词: 虾肝肠胞虫; 基因组; 微卫星; 分子标记

中图分类号: S941

文献标志码: A

文章编号: 1005–8737–(2023)11–1398–11

虾肝肠胞虫(*Enterocytozoon hepatopenaei*, EHP)是一种微孢子虫^[1-2], 隶属于真菌界(Fungi)微孢子虫门(Microsporidia)单倍期纲(Haplophasea)壶目(Chytridiopsida)肠胞虫科(Enterocytozoonidae)肠胞虫属(*Enterocytozoon*)。虾肝肠胞虫主要感染十足类肝胰腺上皮细胞^[3], 是引起虾类肝胰腺微孢子虫病的病原^[4], 于 2009 年在泰国养殖的斑节对虾(*Penaeus monodon*)中首次发现并被表征和命名^[3]。虾肝肠胞虫可导致养殖虾出现一定程度的生长迟缓^[5], 降低其免疫力, 使其易受到细菌和病毒的继发感染^[6]。虾肝肠胞虫可通过养殖水体、粪便或同类相食进行水平传播^[7], 也可在亲子代

间进行垂直传播^[8]。据报道, 虾肝肠胞虫不仅可以感染斑节对虾^[3]、南美白对虾(*Litopenaeus Vannamei*)^[9]、细角滨对虾(*Litopenaeus stylirostris*)^[10]及罗氏沼虾(*Macrobrachium rosenbergii*)^[11], 也可感染多毛类、软体类以及卤虫等^[4,10]。目前, 虾肝肠胞虫已在越南、印度、文莱、中国、印度尼西亚、马来西亚、委内瑞拉、澳大利亚、韩国等多国养虾业中发现并流行^[7,9-10,12-15]。虽然虾肝肠胞虫感染不会引起虾类大面积死亡, 但会显著减缓虾的生长并使虾更易受到其他病原感染, 造成重大经济损失, 正在成为对虾养殖业的重大威胁^[16]。目前, 国内外对于虾肝肠胞虫的研究主要集中在生理结

收稿日期: 2023-10-07; **修订日期:** 2023-11-17.

基金项目: 江苏现代农业(特色水产)产业技术体系创新中心项目(JATS[2022]419, JATS[2023]380); 江苏省六大人才高峰项目(2019-NY-106); 江苏省农业科技自主创新资金项目(CX(21)3160).

作者简介: 全德润(1999–), 男, 硕士研究生, 研究方向为水产养殖. E-mail: 852250713@qq.com

通信作者: 万夕和, 研究员, 研究方向为水产健康养殖与病害防控. E-mail: wxh1708@163.com

构、组织病理、传播机制以及检测防控等方面,而在分子遗传学方面较为缺乏。

微卫星(microsatellite)即简单重复序列(simple sequence repeat, SSR),或称短串联重复序列(short tandem repeats, STRs),由 1~6 个碱基串联重复组成^[17-19]。SSR 在真核生物、原核生物^[20]及病毒的基因组中^[21]均有发现。SSR 的产生有多种原因,包括 DNA 复制、修复期间的滑链错配^[22],或者姐妹染色单体出现了不均等交换^[23]。由于其广泛的分布范围、众多的数量以及共显性的遗传特性,SSR 在遗传多样性研究、遗传图谱构建以及物种和亲属关系的确定等方面发挥着关键作用^[24-25]。目前有多种 SSR 分子标记开发技术,经典的构建基因组文库方法操作繁琐、实验成功率低^[26],SSR 引物富集法和杂交富集法工作量大、耗时长^[27]。随着高通量测序技术的不断发展,大量真菌的全基因组序列被公开^[28],使用生物信息学技术对其基因组数据进行研究,为 SSR 分子标记提供了便捷的途径。目前,已有多种大型真菌基因组 SSR 数据得到研究,如双色蜡蘑(*Laccaria bicolor*)^[29]、黑松露(*Tuber melanosporum*)^[30]和香菇(*Lentinula edodes*)^[31]。然而关于微孢子虫门真菌的基因组 SSR 的研究却尚未出现。2021 年虾肝肠胞虫全基因组序列的报道^[32],使得虾肝肠胞虫的遗传密码被成功破译,为寻找并挖掘虾肝肠胞虫的关键功能基因提供了珍贵的数据资源。

本研究运用生物信息学的技术手段,对虾肝肠胞虫的全基因组序列进行了 SSR 位点的分析,对比了其他 5 种微孢子虫真菌物种的 SSR 分布情况,该结果为探讨 SSR 在基因组结构中的作用提供了依据,为虾肝肠胞虫及其近缘种在分子标记开发、种群进化和遗传多样性研究等方向提供理论支持。

1 材料与方法

1.1 基因组数据来源

虾肝肠胞虫基因组(NCBI 登录号 GCA_02307-9535.1)由本实验室进行测序和组装,基因组大小为 3.03 Mb,获得 15 个重叠群,最大重叠群长度为 399434 bp, N50 长度为 230105 bp (0.23 Mb),GC 含量为 25.22%。中华绒螯蟹微孢子虫[*Hepatospora*

eriocheir (NCBI 登录号 ASM208788v1)]、肠孢子虫[*Enterospora canceri* (NCBI 登录号 ASM208791v1)]、线虫孢子虫[*Nematocida homosporus* (NCBI 登录号 nhom_JUm1504)]、水蚤微孢子虫[*Mitosporidium daphnia* (NCBI 登录号 UGP1.1)]和东方蜜蜂微孢子虫[*Nosema ceranae* (NCBI 登录号 ASM98816v1)]的基因组信息来自美国国家生物技术信息中心 NCBI。

1.2 SSR 序列搜索及引物设计

利用软件 Krait^[33](<https://github.com/lmdu/krait>)搜索基因组中满足定义的 SSR 位点。SSR 的筛选标准如下:将单核苷酸单元的重复数最小值设置为 10,二核苷酸单元的重复数最小值设置为 6,三核苷酸单元的重复数最小值设置为 5,四核苷酸、五核苷酸和六核苷酸单元的重复数最小值设置为 4。使用 Excel 软件对基因组中不同类型 SSR 数量、长度等信息进行统计分析,使用 GraphPad Prism 9.5 绘图。将筛选得到的 SSR 利用软件内集成的 Primer 3 采用默认参数设计引物。

1.3 含 SSR 基因的 GO 功能注释

利用 Blast 工具将虾肝肠胞虫基因组中含 SSR 的基因与蛋白质数据库 GO (Gene Ontology)进行比对(E 值 $<10^{-5}$)。利用 R 语言从 GO 数据库中对含 SSR 的基因进行功能注释。

2 结果与分析

2.1 SSR 在虾肝肠胞虫基因组中的分布特征

在虾肝肠胞虫基因组中共搜索到 794 个 SSR,其总长度为 21773 bp,占基因组的 0.72%,SSR 的平均长度为 27.43 bp,相对丰度为 262 loci/Mb,相对密度为 7184.43 bp/Mb。6 种 SSR 类型中,四核苷酸 SSR 数量最多,有 570 个,总长度为 12152 bp,占 SSR 总数的 71.79%,平均长度为 21.32 bp,相对丰度为 188.08 loci/Mb,相对密度为 4009.79 bp/Mb;二核苷酸 SSR 数量最少,仅有 20 个,总长度为 280 bp,占 SSR 总数的 2.52%,平均长度为 14 bp,相对丰度为 6.6 loci/Mb,相对密度为 92.39 bp/Mb;其余类型 SSR 特征见表 1。较长 SSR 能够呈现出更高的多态性,且多核苷酸 SSR 的突变率较高,满足虾肝肠胞虫鉴定等实际需求。在虾肝肠胞虫基因组中选择 30 个序列长度不小于 20 bp 的多核苷酸 SSR 进行引物设计(表 2)。

表 1 虾肝肠胞虫基因组中不同类型 SSR 统计

Tab. 1 Summary of different types of SSR in the genome of *Enterocytozoon hepatopenaei*

类型 type	数量 counts	长度/bp length	占比/% percent	平均长度/bp average	相对丰度/(loci/Mb) relative abundance	相对密度/(bp/Mb) relative density
单核苷酸 mononucleotide	23	250	2.9	10.87	7.59	82.49
二核苷酸 dinucleotide	20	280	2.52	14	6.6	92.39
三核苷酸 trinucleotide	26	432	3.27	16.62	8.58	142.55
四核苷酸 tetranucleotide	570	12152	71.79	21.32	188.08	4009.79
五核苷酸 pentanucleotide	127	6385	15.99	20.28	41.91	2106.86
六核苷酸 hexanucleotide	28	2274	3.53	81.21	9.24	750.35

表 2 在虾肝肠胞虫基因组中 30 个较长 SSR 的引物

Tab. 2 Set of primers corresponding to the 30 long SSR loci from the *Enterocytozoon hepatopenaei* genome

基序 motif	重复数 repeat	正向引物序列(5'-3') forward primer sequences (5'-3')	反向引物序列(5'-3') reverse primer sequences (5'-3')
AATACT	25	GCTGATCCATATGCACATAAGGC	CAATATGTTTGACTATGGTGATGGC
TATAT	65	ACTGCTAATGGGAGGGATGC	ACACAACCTGTCCCTGAACC
AATT	5	AGCACTAAAGCTAAACAGAAATTAACG	TGAAGAAATACCGCAATTTTTGGG
AATTA	9	GGCAAGTCCAAAAACCAAGTGG	CCGGATAAACTTATGCTACTCCC
ATATT	73	CTGGGGCAGAAGAAGAGGG	TTTAACCAAAAAGGAAACAGAGGC
AATAA	4	TCACTATCCATTGAATCTGCTGC	AGGGAGGAAGCAAAAAGATTGG
ATATAA	38	TTTCAACCTCCTTGAGTATGCC	TGTGCAACCATAGCAAAGGC
TTATA	4	CATCCACATAAAGGCGCACG	GCTTCCATGTGAAATAACTGGCC
TTAT	5	GTTGCCCCACTCTATGTTTG	TGCCTTTTCAATCTCTTTATTGCC
TAGT	5	AGGTTAGGTTAGGTAGTTTAGGTTAGG	AAACCACTAACTACTACCCTAAAACC
AACT	6	ACTTAACTAACTAACCTAACCTAACCG	CGGTTTCGTTAGTTTAGTTTAGTTTAGG
CTT	7	ACTTCATCTTTGGTGTGTCACG	TGTGTTGATGGCATGATTGATGG
TATAT	40	GTAAGGAGACTGCCACCAGG	GAGCAAAAAGAGGCAGGTGC
TTGTA	46	TGCAGTTTCTACGTCACCCC	ACATTGGCAGTGATCATTTGAGG
ATATA	51	CCAGTGCATCAGAAGGTGGG	AGACAGACACTCTCCTTGTATTCC
AAT	7	ACAGCAGAATAGTAGGGGAGG	TAGACCAGCACATGCACAGG
AAAT	6	ACATGCACCTTCATTTGAACTACC	AGATCAGAACCATGTGATAGCG
TATAT	4	AGTATTGGTGTAGTGTATTTTGTGC	ACATGTCAGCAGAAGGAGGG
TATTTG	28	AGTGGTTTCAGTAATTTCTTCGCC	TGTGCTAGGGGTGTGTTTGG
TCATTA	4	CCTGTATGCTTTGTTGTATTGTGC	GCAAGTGTTTATTACACAATATGAGC
TATAA	4	GCCTCCACTTGCTATTTACC	CAATGAAAGGGGAAAGCACC
ATTAT	149	AGGAAGGAAGTAGAGAAGTATGGG	GGTGATGGTATGTTGCAAAGAGC
ACTA	6	CCTAACCTAACTAACCTAACCTACC	GCTATTGACTAGATATCTTACCTACGG
ACTA	5	AACTAACTAACTAACTAACCTACC	GCTATTGACTAGATATCTTACCTACGG
TCTTTT	4	TGTTTCTTAGGCACGCTTG	TGACGATGATGGTGGTGACG
TCA	7	AGGGAAACTACTTCAAGTGTGC	ACTTCAGGTGATAAGCCAGGC
ATACT	14	TCAAAAATGTCAGGGTCAAAAAGG	TGGATTGATGGGCAGGAGG
AATGGT	4	AAACCTGGCAATGGCAATGG	ATTACCACCACTGCCGTTGC
TGAAAA	5	CCAGTGAAGATGATAGGTGGGG	TTCGTTACAAGCAATGTATGTGC
TATAT	18	TGTAAATTCATCCTCATCTCTCTCC	GGAGGAAGTGGTGAGACACG

2.2 虾肝肠胞虫 SSR 核心序列拷贝数变化分析

本研究中发现在虾肝肠胞虫的基因组中，随

着 SSR 拷贝数的增长，其数量反而呈现出下降的

趋势。不同类型的 SSR 核心序列在拷贝数上的优

势分布存在显著差异。例如, 在 6 种主要的碱基类型中, 单碱基 SSR 的拷贝次数最高可达 10 次, 而二碱基 SSR 则以 6 次为最常见, 三碱基 SSR 的拷贝次数达到 5 次, 四碱基、五碱基以及六碱基 SSR 的拷贝次数均为 4 次。在 SSR 的核心序列拷贝数分布方面, 单碱基 SSR 的核心序列拷贝数通常为 10 次, 而双碱基核心序列拷贝数则主要落在 6 至 8 次之间。对于三碱基核心序列, 其拷贝数一般在 5 至 7 次之间, 而四碱基核心序列的拷贝数则倾向于在 4 至 8 次之间。同样的, 五碱基核心序列的拷贝数也大多集中在 4 至 6 次之间, 而六碱基核心序列的拷贝数则只有 4 次。具体来说, 1 至 6 个碱基类型的 SSR 在其主要集中范围内的数量占据了相应种类 SSR 总数的 82.61%、90%、96.15%、93.33%、79.53% 以及 57.14% (如图 1 所示)。

2.3 虾肝肠胞虫 SSR 各重复类别特征分析

如表 3 所示, 在虾肝肠胞虫的基因组中, 统计分析了一系列具有高频次(至少 10 次)的 SSR 重复序列, 这些序列共包含了 11 种不同的类型, 总计 600 个, 占据了所有 SSR 序列的 75.57%。其中, 四核苷酸 SSR 重复序列是最常见的, 共有 8 种, 数量达到 549 个, 占据了高频次 SSR 序列的 91.5%, 且该类 SSR 重复序列在基因组中的分布频率存在显著差异, AACT 类型的重复序列频率最高, 达到了 23.55%, 而像 TAAC 这样的低频率重复序列仅为 1.51%。五核苷酸 SSR 重复序列中, AACTA 重复序列最为常见, 共有 22 个, 其次为 TTAGT, 共有 18 个。在二核苷酸 SSR 重复序列中, 只有 AT 重复序列出现了高频次的重复, 至于单核苷酸、三核苷酸和六核苷酸 SSR 重复序列, 并没有出现高频次的重复序列。

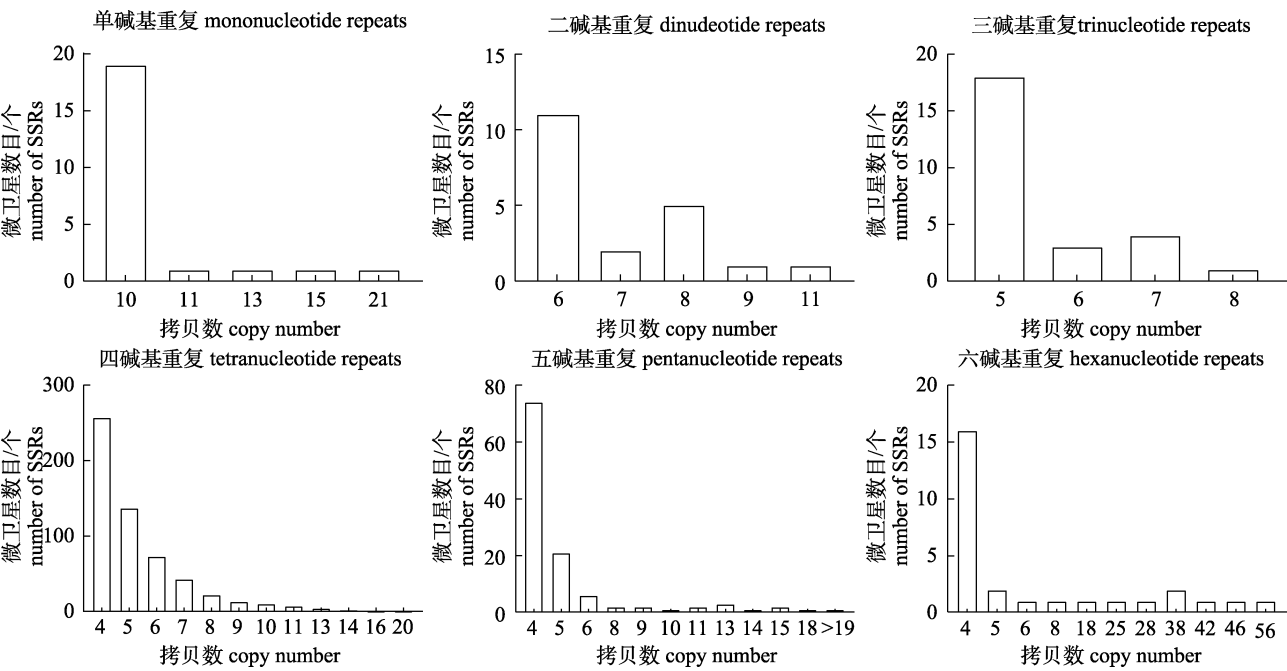


图 1 虾肝肠胞虫各类型 SSR 核心序列数分布
Fig. 1 Distribution of different copy numbers of various types of microsatellites in *Enterocytozoon hepatopenaei*

2.4 虾肝肠胞虫与其他 5 种微孢子虫门真菌基因组中 SSR 分布的比较

对虾肝肠胞虫以及 5 种微孢子虫的基因组进行了 SSR 位点分布的统计研究, 以对比它们的数量、相对丰富程度及相对密度的差异(详见表 4)。这些微孢子虫的基因组规模范围从 3030581 bp

(虾肝肠胞虫)到 5690748 bp (东方蜜蜂微孢子虫)不等, 它们包含的 SSR 总数范围从 261 个(中华绒螯蟹微孢子虫)至 794 个(虾肝肠胞虫), 分别占基因组的 0.24%和 0.72%。虾肝肠胞虫基因组的 SSR 数量、相对丰度和相对密度均高于其他 5 种微孢子虫基因组。研究结果显示, SSR 的数量和相对丰

表 3 虾肝肠胞虫基因组中高频率(≥10)出现的
SSR 重复类型

Tab. 3 The most frequently occurring SSR motifs in
the *Enterocytozoon hepatopenaei* genome

SSR 类别 SSR type	SSR 出现次数 SSR occurrence	占各碱基类型 比例/% proportion of each base type/%	占高频率 SSR 比例/% proportion of the most frequently occurring SSR
AACT	187	32.81	31.17
TTAG	175	30.70	29.17
TAGT	59	10.35	9.83
ACTA	47	8.25	7.83
AGTT	25	4.39	4.17
CTAA	24	4.21	4.00
AACTA	22	17.32	3.67
GTTA	20	3.51	3.33
TTAGT	18	14.17	3.00
TAAC	12	2.11	2.00
AT	11	55.00	1.83

度与其基因组大小无关。东方蜜蜂微孢子虫拥有最大的基因组, 其中包含 608 个 SSR (相对丰度为 160.84 loci/Mb), 相比之下, 虾肝肠胞虫的基因组较小, 但它却包含了 794 个 SSR (相对丰度为 262.0 loci/Mb), 中华绒螯蟹微孢子虫基因组(长度为 4570278 bp)介于虾肝肠胞虫和东方蜜蜂微孢子虫之间, 其包含的 SSR 数量为 261 个(相对丰度为 57.51 loci/Mb), 是 SSR 数量最少的。

对 6 种微孢子虫的基因组进行深入分析, 探

讨了各种重复序列在数量、相对丰度以及相对密度上的差异。结果显示, 在东方蜜蜂微孢子虫、肠孢子虫和水蚤微孢子虫的 SSR 总数中, 短重复单元(从单核苷酸到三核苷酸)占绝大多数, 比例分别为 95.71%、81.14%和 72.43%。在东方蜜蜂微孢子虫中, 单核苷酸 SSR 所占比例最大, 为 89.97 %, 而其他 5 种 SSR 仅占 10.03%。其他 3 种微孢子虫(虾肝肠胞虫、线虫孢子虫和中华绒螯蟹微孢子虫)的长重复单元(从四核苷酸到六核苷酸)在 SSR 总数中占绝大多数, 比例分别为 91.31%、84.36%和 65.52%。在虾肝肠胞虫和中华绒螯蟹微孢子虫的 SSR 中, 四核苷酸 SSR 所占比例最大, 分别为 71.79%和 63.6%, 而其他 5 种 SSR 分别占 28.21%和 36.4%。在线虫孢子虫的 SSR 中, 五核苷酸 SSR 所占比例最大, 为 71.68%, 而其他 5 种 SSR 仅占 28.32% (表 5, 图 2)。

对以上微孢子虫进一步统计, 以确定各种 SSR 重复类型的出现频率。在单核苷酸重复单元的 SSR 中, A 重复单元在虾肝肠胞虫、肠孢子虫、中华绒螯蟹微孢子虫、水蚤微孢子虫和东方蜜蜂微孢子虫中占主要比例, 而 T 重复单元则在线虫微孢子虫中占主要比例。在二核苷酸重复单元的 SSR 中, AT 重复单元在虾肝肠胞虫和东方蜜蜂微孢子虫中占主要比例, 而其他 4 种微孢子虫中占主要比例的二核苷酸重复单元 SSR 则各不相同。在三、四、五和六核苷酸重复单元中, 这 6 种微孢子

表 4 虾肝肠胞虫与其他 5 种微孢子虫门真菌基因组及 SSR 信息

Tab. 4 Overview of genomes and SSR in *Enterocytozoon hepatopenaei* and other five Microsporidia genomes and SSR

指标 item	虾肝肠胞虫 <i>Enterocytozoon hepatopenaei</i>	肠孢子虫 <i>Enterospora canceri</i>	中华绒螯蟹 微孢子虫 <i>Hepatospora eriocheir</i>	线虫孢子虫 <i>Nematocida homosporus</i>	水蚤微孢子虫 <i>Mitosporidium daphniae</i>	东方蜜蜂微 孢子虫 <i>Nosema ceranae</i>
所分析基因组长度/bp sequence analyzed	3030581	3095385	4570278	4282417	5635072	5690748
GC 含量/% GC content	25.23	40.15	22.6	41.04	43.0	25.36
SSRs 数量 number of SSRs	794	350	261	338	341	608
相对丰度/(loci/Mb) relative abundance	262.0	113.08	57.51	79.17	60.52	106.84
SSRs 总长度/bp total length of SSRs	21773	10028	10520	7145	6035	6972
相对密度/(bp/Mb) relative density	7184.43	3239.87	2318.15	1668.58	1071.03	1225.15
占基因组的含量/% genome content	0.72	0.33	0.24	0.17	0.11	0.13

表 5 6 种微孢子虫基因组中各种 SSR 的数量、总长、相对丰度及相对密度

Tab. 5 Number, total length, relative abundance, and relative density of SSRs in six Microsporidia genomes

类型 type		虾肝肠胞虫 <i>Enterocytozoon hepatopenaei</i>	肠孢子虫 <i>Enterospora canceri</i>	中华绒螯蟹 微孢子虫 <i>Hepatospora eriocheir</i>	线虫孢子虫 <i>Nematocida homosporus</i>	水蚤微孢子虫 <i>Mitosporidium daphniae</i>	东方蜜蜂微 孢子虫 <i>Nosema ceranae</i>
单核苷酸 mononucleotide	数量 counts	23	62	56	1	107	547
	总长度/bp total length	250	843	579	10	1140	5718
	平均长度/bp average length	10.87	13.6	10.34	10	10.65	10.45
	相对丰度/(loci/Mb) relative abundance	7.59	20.03	12.34	0.23	18.99	96.12
	相对密度/(bp/Mb) relative density	82.49	272.36	127.59	2.34	202.32	1004.79
二核苷酸 dinucleotide	数量 counts	20	119	20	4	32	22
	总长度/bp total length	280	4284	248	50	560	322
	平均长度/bp average length	14	36	12.4	12.5	17.5	14.64
	相对丰度/(loci/Mb) relative abundance	6.6	38.45	4.41	0.93	5.68	3.87
	相对密度/(bp/Mb) relative density	92.39	1384.08	54.65	11.68	99.38	56.58
三核苷酸 trinucleotide	数量 counts	26	103	14	48	108	13
	总长度/bp total length	432	2892	213	831	2034	216
	平均长度/bp average length	16.62	28.08	15.21	17.31	18.83	16.62
	相对丰度/(loci/Mb) relative abundance	8.58	33.28	3.08	11.21	19.17	2.28
	相对密度/(loci/Mb) relative density	142.55	934.35	46.94	194.06	360.97	37.96
四核苷酸 tetranucleotide	数量 counts	570	48	166	10	37	20
	总长度/bp total length	12152	1396	9372	168	688	332
	平均长度/bp average length	21.32	29.08	56.46	16.8	18.59	16.6
	相对丰度/(loci/Mb) relative abundance	188.08	15.51	36.58	2.34	6.57	3.51
	相对密度/(bp/Mb) relative density	4009.79	451.02	2065.18	39.23	122.1	58.34
五核苷酸 pentanucleotide	数量 counts	127	9	3	243	32	4
	总长度/bp total length	6385	205	60	5240	755	150
	平均长度/bp average length	50.28	22.78	20	21.56	23.59	37.5
	相对丰度/(loci/Mb) relative abundance	41.91	2.91	0.66	56.75	5.68	0.7
	相对密度(bp/Mb) relative density	2106.86	66.23	13.22	1223.7	133.99	26.36
六核苷酸 hexanucleotide	数量 counts	28	9	2	33	25	2
	总长度/bp total length	2274	408	48	846	858	234
	平均长度/bp average length	81.21	45.33	24	25.64	34.32	117
	相对丰度/(loci/Mb) relative abundance	9.24	2.91	0.44	7.71	4.44	0.35
	相对密度/(bp/Mb) relative density	750.35	131.82	10.58	197.57	152.27	41.12
所有 SSR all SSRs	数量 counts	794	350	261	339	341	608
	总长度/bp total length	21773	10028	10520	7145	6035	6972
	平均长度/bp average length	27.43	28.66	40.31	21.08	17.7	11.47
	相对丰度/(loci/Mb) relative abundance	262	113.08	57.51	79.17	60.52	106.84
	相对密度/(bp/Mb) relative density	7184.43	3239.87	2318.15	1668.58	1071.03	1225.15

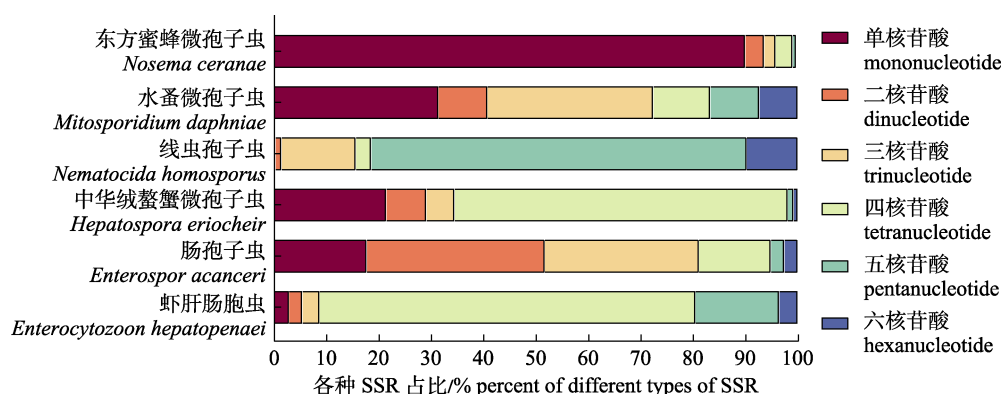


图 2 6 种微孢子虫门真菌基因组中各种 SSR 占其 SSR 总数的比例

Fig. 2 Distribution of SSRs in the genomes of six Microsporidia fungi

虫中高频率出现的 SSR 类型各不相同。与其他 5 种微孢子虫相比, 虾肝肠胞虫基因组中长重复单元的相对密度最高, 其中四核苷酸重复单元的相对密度为 4009.79 bp/Mb, 在该物种高频率出现的 SSR 重复类型中共有 8 种, 其中 AACT 重复单元占主导地位, 数量达到 187 个, 占总 SSR 数的 23.55%。

2.5 虾肝肠胞虫基因组中含 SSR 基因的 GO 功能注释

利用 GO 数据库对含 SSR 基因进行功能注释, 共有 28 个基因(3.53%)得到 GO 分类号。其中注释到分子功能(molecular function)的基因最多, 共 25 个; 其次是生物学过程(biological process), 共 12 个基因; 细胞组分(cell component)的注释基因最少, 共 10 个。在分子功能这一类别中, 以注释到结合活性(binding)亚组的基因最多, 共有 19 个, 占 GO 注释基因的 67.86%, 其次是催化活性(catalytic activity)亚组, 共有 10 个, 占 GO 注释基因的 35.71%。在生物学过程类别中, 以细胞进程(cellular process)和代谢进程(metabolic process)为最大的 2 个亚组, 分别包含 10 个和 9 个基因, 占 GO 注释基因的 35.71%和 32.14%。在细胞组分类别中, 细胞(cell)和细胞组分(cell part)这两个子类是功能注释的主要部分, 它们各自包含了 10 个被注释的基因, 占有所有具有 GO 注释基因总数的 17.39% (详见图 3)。

3 讨论

本研究利用生物信息学方法, 基于虾肝肠胞虫基因组数据, 对 SSR 进行搜索统计分析。在虾

肝肠胞虫基因组中, 共搜索到 794 个 SSR, 总长度为 21773 bp, 占基因组的 0.72%, 平均长度为 27.43 bp, 相对丰度为 262 loci/Mb, 相对密度为 7184.43 bp/Mb。相较于其他一些常见的哺乳动物如人类 (*Homo sapiens*, 3.00%)^[34]、小鼠 (*Mus musculus*, 2.85%)^[35]、牛 (*Bos taurus*, 4.70%) 以及绵羊 (*Ovis aries*, 4.80%)^[36], 虾肝肠胞虫的基因组中的 SSR 数量相对较少。与美丽硬仆骨舌鱼 (*Scleropages Formosus*, 0.79%)^[37]、红鳍东方鲀 (*Takifugu rubripes*, 0.73%)、菊黄东方鲀 (*Takifugu flavidus*, 0.73%)^[38] 等物种 SSR 含量相似。然而与云芝 (*Trametes versicolor*, 0.05%)、双孢蘑菇 (*Agaricus bisporus*, 0.15%)、黑管孔菌 (*Bjerkandera adusta*, 0.04%)、红缘拟层孔菌 (*Fomitopsis pinicola*, 0.06%)、金针菇 (*Flammulina velutipes*, 0.06%) 和鲜红密孔菌 (*Pycnoporus cinnabarinus*, 0.09%)^[39] 等真菌相比, 虾肝肠胞虫的 SSR 数量明显更高。分析导致其差异的原因可能在于不同生物种类之间的基因组规模、基因组内的碱基构成以及排列模式不同^[40]。

在绝大多数真菌及其他真核生物中, 单核苷酸 SSR 是数量最多的 SSR 重复类型, 而在绝大多数原核生物的基因组中分布最多的是三核苷酸 SSR^[21,41]。本研究表明, 不同微孢子虫基因组中优势 SSR 类型不尽相同, 单核苷酸 SSR 是东方蜜蜂微孢子虫基因组中占比最高的 SSR 类型, 二核苷酸 SSR 是肠孢子虫基因组中占比最高的 SSR 类型, 三核苷酸 SSR 是水蚤微孢子虫基因组中占比最高的 SSR 类型, 四核苷酸 SSR 是虾肝肠胞虫和中华

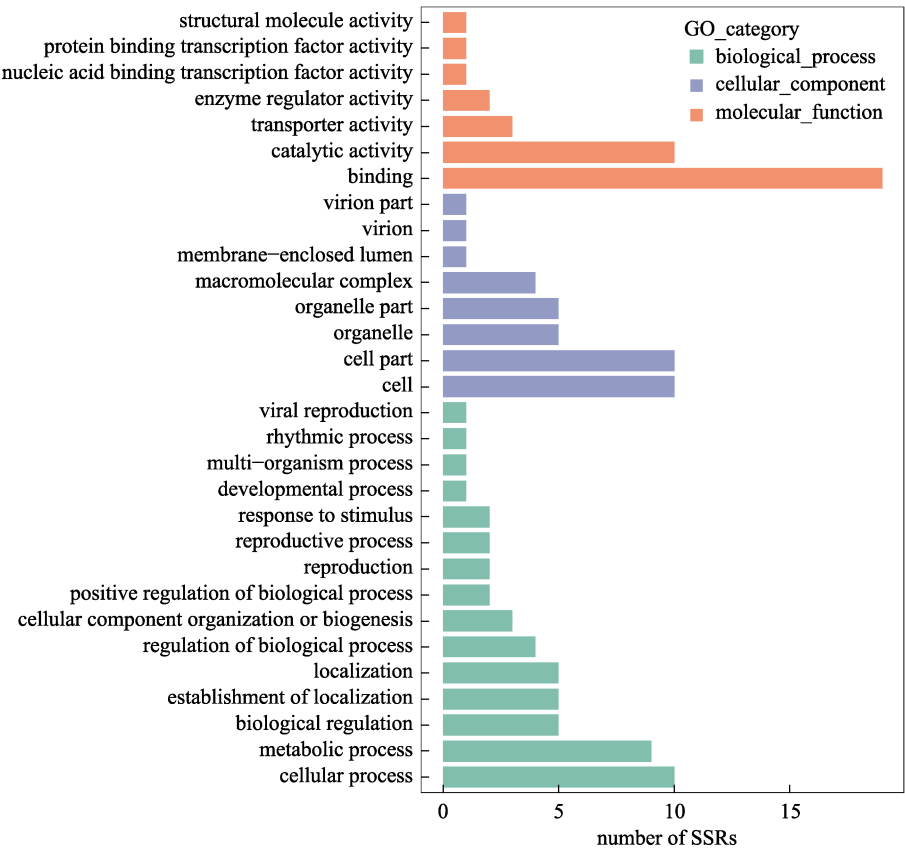


图 3 虾肝肠胞虫 SSR 基因 GO 功能注释

Fig. 3 GO function annotation of genes with SSRs in *Enterocytozoon hepatopenaei*

绒螯蟹微孢子虫基因组中占比最高的 SSR 类型, 五核苷酸 SSR 是线虫孢子虫基因组中占比最高的 SSR 类型。这些研究结果证实, 不同微孢子虫基因组 SSR 分布特征具有较大差异。

通过对包含 SSR 序列的基因进行功能性注释, 笔者发现在虾肝肠胞虫基因组中共有 28 个基因被赋予了特定的功能, 占有所有基因的 3.53%。通过分析这些基因的功能特性, 发现它们主要与基础代谢有关。从分子功能层面分析, 这些基因主要参与到结合活性和催化活动之中; 从生物学过程角度来看, 它们主要参与细胞进程和代谢进程; 而从细胞组成的角度看, 它们主要定位在细胞和细胞组分上。值得注意的是, 该研究结果与其他真菌如云芝(*Trametes versicolor*)^[39]、灵芝(*Ganoderma lucidum*)^[42]、草菇(*Volvariella volvacea*)^[43]和糙皮侧耳(*Pleurotus ostreatus*)^[44]等已知 GO 注释结果相吻合。本研究对虾肝肠胞虫中含有 SSR 的基因进行了 GO 功能注释, 初步探索了 SSR 在虾肝肠

胞虫的基础代谢过程中所扮演的角色, 这一发现对于深入研究虾肝肠胞虫感染过程以及其与环境因素之间的互动关系具有重要意义, 有助于制定有效的预防策略。同时该研究挖掘了较多的全基因组 SSR 的数据, 为进一步开发 SSR 标记提供了有力支持, 有助于深入分析虾肝肠胞虫群体的多样性, 精确识别其种群, 以及帮助构建遗传图谱并进行性状相关性的分析。

参考文献:

[1] James T Y, Pelin A, Bonen L, et al. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia[J]. Current Biology, 2013, 23(16): 1548-1553.

[2] Corradi N. Microsporidia: Eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers[J]. Annual Review of Microbiology, 2015, 69: 167-183.

[3] Tourtip S, Wongtripop S, Stentiford G D, et al. *Enterocytozoon hepatopenaei* sp. nov. (Microsporida: Enterocytozoonidae), a parasite of the black tiger shrimp *Penaeus monodon* (Decapoda: Penaeidae): Fine structure and phylogenetic relationships[J].

- Journal of Invertebrate Pathology, 2009, 102(1): 21-29.
- [4] Thitamadee S, Prachumwat A, Srisala J, et al. Review of current disease threats for cultivated penaeid shrimp in Asia[J]. Aquaculture, 2016, 452: 69-87.
- [5] Han J E, Tang K F J, Kim J H. The use of beta-tubulin gene for phylogenetic analysis of the microsporidian parasite *Enterocytozoon hepatopenaei* (EHP) and in the development of a nested PCR as its diagnostic tool[J]. Aquaculture, 2018, 495: 899-902.
- [6] Patil P K, Geetha R, Ravisankar T, et al. Economic loss due to diseases in Indian shrimp farming with special reference to *Enterocytozoon hepatopenaei* (EHP) and white spot syndrome virus (WSSV)[J]. Aquaculture, 2021, 533: 736231.
- [7] Tang K F J, Han J E, Aranguren L F, et al. Dense populations of the microsporidian *Enterocytozoon hepatopenaei* (EHP) in feces of *Penaeus vannamei* exhibiting white feces syndrome and pathways of their transmission to healthy shrimp[J]. Journal of Invertebrate Pathology, 2016, 140: 1-7.
- [8] Tangprasittipap A, Srisala J, Chouwdee S, et al. The microsporidian *Enterocytozoon hepatopenaei* is not the cause of white feces syndrome in whiteleg shrimp *Penaeus (Litopenaeus) vannamei*[J]. BMC Veterinary Research, 2013, 9(1): Article No.139.
- [9] Rajendran K V, Shivam S, Ezhil Praveena P, et al. Emergence of *Enterocytozoon hepatopenaei* (EHP) in farmed *Penaeus (Litopenaeus) vannamei* in India[J]. Aquaculture, 2016, 454: 272-280.
- [10] Tang K F J, Pantoja C R, Redman R M, et al. Development of in situ hybridization and PCR assays for the detection of *Enterocytozoon hepatopenaei* (EHP), a microsporidian parasite infecting penaeid shrimp[J]. Journal of Invertebrate Pathology, 2015, 130: 37-41.
- [11] Jang G I, Kim S M, Oh Y K, et al. First report of *Enterocytozoon hepatopenaei* infection in giant freshwater prawn (*Macrobrachium rosenbergii* de man) cultured in the republic of Korea[J]. Animals (Basel), 2022, 12(22): 3149.
- [12] Tang K F J, Aranguren L F, Piamsomboon P, et al. Detection of the microsporidian *Enterocytozoon hepatopenaei* (EHP) and Taura syndrome virus in *Penaeus vannamei* cultured in Venezuela[J]. Aquaculture, 2017, 480: 17-21.
- [13] Aldama-Cano D J, Sanguanrut P, Munkongwongsiri N, et al. Bioassay for spore polar tube extrusion of shrimp *Enterocytozoon hepatopenaei* (EHP)[J]. Aquaculture, 2018, 490: 156-161.
- [14] Kim B S, Jang G I, Kim S M, et al. First report of *Enterocytozoon hepatopenaei* infection in Pacific whiteleg shrimp (*Litopenaeus vannamei*) cultured in Korea[J]. Animals (Basel), 2021, 11(11): 3150.
- [15] Kim J H, Lee C, Jeon H J, et al. First report on *Enterocytozoon hepatopenaei* (EHP) infection in Pacific white shrimp (*Penaeus vannamei*) cultured in Korea[J]. Aquaculture, 2022, 547: 737525.
- [16] Karthikeyan K, Sudhakaran R. Experimental horizontal transmission of *Enterocytozoon hepatopenaei* in post-larvae of whiteleg shrimp, *Litopenaeus vannamei*[J]. Journal of Fish Diseases, 2019, 42(3): 397-404.
- [17] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes[J]. Nucleic Acids Research, 1984, 12(10): 4127-4138.
- [18] Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers[J]. Nucleic Acids Research, 1989, 17(16): 6463-6471.
- [19] Levinson G, Gutman G A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution[J]. Molecular Biology and Evolution, 1987, 4(3): 203-221.
- [20] Sinden R R. Biological implications of the DNA structures associated with disease-causing triplet repeats[J]. American Journal of Human Genetics, 1999, 64(2): 346-353.
- [21] Sharma P C, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes[J]. Trends in Biotechnology, 2007, 25(11): 490-498.
- [22] Zhao X Y, Tan Z Y, Feng H P, et al. Microsatellites in different Potyvirus genomes: Survey and analysis[J]. Gene, 2011, 488(1-2): 52-56.
- [23] Zhang Z C, Hou X L. Strategies for development of SSR molecular markers[J]. Hereditas (Beijing), 2004, 26(5): 763-768. [张增翠, 侯喜林. SSR 分子标记开发策略及评价[J]. 遗传, 2004, 26(5): 763-768.]
- [24] Wu X P, Feng Y W, Jiang H L, et al. A preliminary genetic linkage map of *Sinonovacula constricta* (Lamarck, 1818) based on microsatellites derived from RAD sequencing[J]. Journal of Ocean University of China, 2018, 17(4): 947-956.
- [25] Jighly A, Al-Abdallat A M, Alnemer L M. The role of microsatellites in streptophyta gene evolution[J]. Journal of Molecular Evolution, 2017, 84(2-3): 144-148.
- [26] Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: A review[J]. Molecular Ecology, 2002, 11(1): 1-16.
- [27] Wang Z, Liu H B, Liu J, et al. Mining new microsatellite markers for Siberian apricot (*Prunus sibirica* L.) from SSR-enriched genomic library[J]. Scientia Horticulturae, 2014, 166: 65-69.
- [28] Grigoriev I V, Nikitin R, Haridas S, et al. MycoCosm portal: Gearing up for 1000 fungal genomes[J]. Nucleic Acids Research, 2014, 42(D1): D699-D704.
- [29] Labbé J, Murat C, Morin E, et al. Survey and analysis of

- simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers[J]. *Current Genetics*, 2011, 57(2): 75-88.
- [30] Murat C, Riccioni C, Belfiori B, et al. Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers[J]. *Fungal Genetics and Biology*, 2011, 48(6): 592-601.
- [31] Lee H Y, Moon S, Shim D, et al. Development of 44 novel polymorphic SSR markers for determination of shiitake mushroom (*Lentinula edodes*) cultivars[J]. *Genes*, 2017, 8(4): 109.
- [32] Wang L B, Li H, Shi W J, et al. Whole-genome sequencing and comparative genomic analysis of a pathogenic *Enterocytozoon hepatopenaei* strain isolated from *Litopenaeus vannamei*[J]. *Aquaculture International*, 2023, 31(1): 523-546.
- [33] Du L M, Zhang C, Liu Q, et al. Krait: An ultrafast tool for genome-wide survey of microsatellites and primer design[J]. *Bioinformatics*, 2018, 34(4): 681-683.
- [34] Subramanian S, Mishra R K, Singh L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions[J]. *Genome Biology*, 2003, 4(2): R13.
- [35] Tong X L, Dai F Y, Li B, et al. Microsatellite repeats in mouse: Abundance, distribution, and density[J]. *Acta Zoologica Sinica*, 2006, 52(1): 138-152.
- [36] Qi W H, Jiang X M, Xiao G S, et al. Seeking and bioinformatics analysis of microsatellite sequence in the genomes of cow and sheep[J]. *Acta Veterinaria et Zootechnica Sinica*, 2013, 44(11): 1724-1733. [戚文华, 蒋雪梅, 肖国生, 等. 牛和绵羊全基因组微卫星序列的搜索及其生物信息学分析[J]. *畜牧兽医学报*, 2013, 44(11): 1724-1733.]
- [37] Duan Y N, Liu Y, Hu Y C, et al. Distribution regularity of microsatellites in *Scleropages formosus* genome[J]. *Chinese Agricultural Science Bulletin*, 2019, 35(23): 152-158. [段永楠, 刘奕, 胡隐昌, 等. 美丽硬仆骨舌鱼全基因组微卫星分布规律特征[J]. *中国农学通报*, 2019, 35(23): 152-158.]
- [38] Xu J J, Zheng X, Zhang X Y, et al. Analysis of distribution characteristics of microsatellites in four genomes of puffer fish[J]. *Genomics and Applied Biology*, 2021, 40(4): 1441-1451. [徐杰杰, 郑翔, 张鑫宇, 等. 4 种河鲀全基因组微卫星分布特征分析[J]. *基因组学与应用生物学*, 2021, 40(4): 1441-1451.]
- [39] Cao Y, Zhang Y Z, Cheng S J, et al. Genome-wide distributional and comparative analysis of SSR loci in *Trametes versicolor*[J]. *Mycosystema*, 2017, 36(11): 1524-1542. [曹云, 张延召, 程抒劼, 等. 云芝全基因组 SSR 位点分布及比较分析[J]. *菌物学报*, 2017, 36(11): 1524-1542.]
- [40] Peng Y, Li J, Wang T, et al. Preliminary study on distribution characteristics and positioning of microsatellites in whole genome of *Pelteobagrus vachelli*[J]. *South China Fisheries Science*, 2022, 18(1): 90-98. [彭冶, 李杰, 王涛, 等. 瓦氏黄颡鱼全基因组微卫星的分布特征及其定位的初步研究[J]. *南方水产科学*, 2022, 18(1): 90-98.]
- [41] Murat C, Riccioni C, Belfiori B, et al. Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers[J]. *Fungal Genetics and Biology*, 2011, 48(6): 592-601.
- [42] Qian J, Xu H B, Song J Y, et al. Genome-wide analysis of simple sequence repeats in the model medicinal mushroom *Ganoderma lucidum*[J]. *Gene*, 2013, 512(2): 331-336.
- [43] Wang Y, Chen M J, Wang H, et al. Microsatellites in the genome of the edible mushroom, *Volvariella volvacea*[J]. *BioMed Research International*, 2014, 2014: 281912.
- [44] Qu J B, Huang C Y, Zhang J X. Genome-wide functional analysis of SSR for an edible mushroom *Pleurotus ostreatus*[J]. *Gene*, 2016, 575(2): 524-530.

Distribution and comparative analysis of SSR sites in the whole genome of *Enterocytozoon hepatopenaei* (EHP)

QUAN Derun^{1,2}, WANG Libao¹, LI Hui¹, SHI Wenjun¹, ZHU Jianqiang¹, GU Shuwen¹, GU Chen¹, REN Qian², WAN Xihe¹

1. Jiangsu Institute of Oceanology and Marine Fisheries, Nantong 226007, China;

2. College of Marine Science and Engineering, Nanjing Normal University, Nanjing 210034, China

Abstract: *Enterocytozoon hepatopenaei* (EHP) is a microsporidian parasite that infects shrimp. Although EHP infection does not cause widespread mortality in shrimp, it significantly slows down their growth and makes them more susceptible to other pathogens, resulting in significant economic losses. Thus, it has emerged as a major threat to the shrimp farming industry. The successful decoding of its genome sequence and genetic code has provided a robust foundation for various fundamental research endeavors. These endeavors provide valuable data essential for the identification and exploration of important functional genes in EHP. Here, we aimed to study the basic information regarding microsatellites in the EHP genome and provide theoretical support for molecular marker development, population evolution, and genetic diversity studies of EHP and its related species. We used bioinformatics to analyze the quantity, distribution, and frequency of SSR (simple sequence repeat) sites in the whole EHP genome and compared them with the corresponding properties of five other microsporidian fungal species. A total of 794 SSRs were identified in the EHP genome, with a total length of 21773 bp, accounting for 0.72% of the genome. The average length of SSRs was 27.43 bp, with a relative abundance of 262 loci/Mb and a relative density of 7184.43 bp/Mb. Of the six identified types of SSRs, tetranucleotide SSRs were the most prevalent, with 570 SSRs and a total length of 12152 bp; they accounted for 71.79% of the total SSRs, the average length of SSRs was 21.32 bp, with a relative abundance of 188.08 loci/Mb and a relative density of 4009.79 bp/Mb. Dinucleotide SSRs were the least prevalent, with only 20 SSRs and a total length of 280 bp; they accounted for 2.52% of the total SSRs, the average length of SSRs was 14 bp, with a relative abundance of 6.6 loci/Mb and a relative density of 92.39 bp/Mb. The most frequently occurring SSR in the genome (≥ 10) included AACT, TTAG, TAGT, ACTA, AGTT, CTAA, AACTA, GTTA, TTAGT, TAAC, and AT. Thirty multi-nucleotide SSRs with sequence lengths ≥ 20 bp were selected for primer design for species identification. The genes containing SSRs were functionally annotated using the Gene Ontology database, and a total of 28 genes (3.53%) were annotated into three categories: molecular function, biological process, and cellular component. Compared to other microsporidian fungi, EHP has a moderate abundance of SSRs, and the quantity of SSRs is independent of genome size. This study provides a theoretical foundation for population genetics and evolutionary research of EHP.

Key words: *Enterocytozoon hepatopenaei*; genome; microsatellite; molecular marker

Corresponding author: WAN Xihe. E-mail: wxh1708@163.com