

DOI: 10.19663/j.issn2095-9869.20190810002

http://www.yykxjz.cn/

宁梓健, 姜宏波, 刘琦, 包杰. 虾肝肠胞虫全基因组分泌蛋白的预测分析. 渔业科学进展, 2020, 41(6): 165–173

Ning ZJ, Jiang HB, Liu Q, Bao J. Genome-wide prediction and analysis of secreted proteins of *Enterocytozoon hepatopenaei*. Progress in Fishery Sciences, 2020, 41(6): 165–173

虾肝肠胞虫全基因组分泌蛋白的预测分析*

宁梓健 姜宏波 刘琦 包杰^①

(沈阳农业大学 畜牧兽医学院 沈阳 110866)

摘要 分泌蛋白是在细胞内合成后分泌到细胞外起作用的蛋白质, 其在许多原生动植物寄生虫操纵宿主细胞和影响虫体的毒力中起着重要作用。虾肝肠胞虫(*Enterocytozoon hepatopenaei*, EHP)是一种可感染多种经济虾类的专性细胞内寄生虫, 是近年来影响全球对虾养殖生产较严重的病害之一。本研究利用真核生物分泌蛋白预测流程 EuSecPred2.0 对 EHP 全基因组的分泌蛋白进行预测, 并对分泌蛋白长度、信号肽长度、切割位点处氨基酸分布等信息进行分析, 对分泌蛋白功能进行分析。结果显示, 分泌蛋白氨基酸长度主要集中在 30~400 aa 之间; 信号肽长度集中在 9~32 aa 之间; 信号肽切割位点处以疏水性氨基酸为主。对信号肽进行基序分析, 发现存在基序 NV[VT][IK]CA[ED][SA]。对所获蛋白质进行功能注释, 发现了多种与微孢子虫黏附侵染、调节细胞周期和免疫反应等相关的关键蛋白。研究结果有助于了解 EHP 对宿主的侵染机制, 同时为进一步明确 EHP 的致病相关因子提供理论依据。

关键词 虾肝肠胞虫; 基因组; 分泌蛋白; 信号肽; 基序

中图分类号 Q18 **文献标识码** A **文章编号** 2095-9869(2020)06-0165-09

虾肝肠胞虫(*Enterocytozoon hepatopenaei*, EHP)是一种严格细胞内寄生的微孢子虫, 成熟的孢子呈椭圆形, 大小为 $(1.4\pm 0.3)\times(0.8\pm 0.1)\ \mu\text{m}$, 后端有 1 个空泡, 胞内含 1 个细胞核, 5~6 圈极丝, 1 个与极丝相连的锚定盘和 1 层电子致密的孢子壁(乔毅等, 2018)。EHP 于 2009 年首次从斑节对虾(*Penaeus monodon*)中发现并命名(Tourtip *et al.*, 2009), 感染 EHP 的对虾往往正常进食, 但生长缓慢甚至停滞, 直接影响到养殖产量, 给养殖户带来严重的经济损失(刘珍等, 2016; 宋增磊等, 2019), 如何预防和治疗 EHP 已经成为对虾养殖产业急需解决的科学问题。明确 EHP 的入侵机理是切断 EHP 的传播途径和针对性用药的基础。EHP 的一些生物合成途径和

三羧酸循环的基因缺失表明其对宿主有很强的依赖性(Katinka *et al.*, 2001), 需要从宿主吸收营养完成生活史发育, 因此, 在进化过程中微孢子虫进化出一种极其复杂和独特的感染机制以抵抗宿主免疫系统的攻击。

分泌蛋白(Secreted protein)是在细胞内合成后分泌到细胞外起作用的蛋白质, 已有研究表明, 其是介导真核病原微生物与宿主之间相互作用的重要因子(McKerrow *et al.*, 1993; Gupta *et al.*, 2012)。在其他微孢子虫中已有研究表明, 一些分泌蛋白在虫体入侵宿主过程中扮演着重要角色(李田等, 2013; Wang *et al.*, 2015)。因此, 查找和分析 EHP 的分泌蛋白对于揭示入侵机理和虫体的免疫逃避将具有重要意义。目前,

* 现代农业产业技术体系(CARS-48)资助 [This work was supported by China Agriculture Research System (CARS-48)].

宁梓健, E-mail: 897290003@qq.com

① 通讯作者: 包杰, E-mail: baojie@syau.edu.cn

收稿日期: 2019-08-10, 收修改稿日期: 2019-10-09

对 EHP 在虾体内的感染机制了解甚少,尤其是在宿主-寄生虫相互作用水平上。本研究在基因组测序数据的基础上,利用生物信息学方法在全基因组范围内预测 EHP 的分泌蛋白,同时,对所得分泌蛋白的功能进行注释,并对其序列特征和基序进行分析,研究结果可为筛选 EHP 致病相关因子和探究其侵染机理提供参考。

1 材料与方法

1.1 基因组数据来源和主要分析软件

虾肝肠胞虫基因组数据来源于 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov>),其 NCBI 登录号为 MNPJ00000001.1。

用到的主要分析软件和网站如下:蛋白质跨膜结构域预测软件 Tmhmm-2.0c(Krogh *et al*, 2001),线粒体蛋白预测软件 MitoProt II (Claros *et al*, 1996),细胞核蛋白预测软件 Nucpred-1.1(Brameier *et al*, 2007)、NLStradamus.1.7(Nguyen Ba *et al*, 2009)和 PredictNLS (<https://roslab.org/owiki/index.php/PredictNLS>),蛋白质信号肽预测软件 SignalP-4.0(Petersen *et al*, 2011),蛋白质 GPI 锚定位点预测软件 Kohgpi-1.5(Fankhauser *et al*, 2005),蛋白质亚细胞定位预测软件 TargetP-1.1 (Emanuelsson *et al*, 2000)和 WoLF PSORT(Horton *et al*, 2007),真核生物分泌蛋白预测程序 EuSecPred2.0 (<https://silkipathdb.swu.edu.cn/eusecpred>),氨基酸组成分析软件 WebLogo(<http://weblogo.threepiusone.com/create.cgi>)(Crooks *et al*, 2004),蛋白质序列基序分析程序为 MEME(<http://meme-suite.org/tools/meme>)(Bailey *et al*, 2009)。

1.2 分泌蛋白的预测

针对 EHP 全基因组的蛋白质编码序列,运行 EuSecPred 2.0 在线流程筛选分泌蛋白(Druzhinina *et al*, 2012)。由于本程序的局限性,只能预测到经典途径分泌的蛋白质,可能会漏掉部分非经典途径分泌的蛋白质。运行 Tmhmm 程序去除跨膜蛋白,运行 Nucpred、NLStradamus.pl 和 PredictNLS 程序剔除其中的细胞核蛋白质,然后,通过 SignalP-4.0 过滤掉其中无信号肽序列的蛋白质,通过 MitoProt 程序过滤掉定位于线粒体的蛋白,利用 TargetP 和 WoLF PSORT 对具有信号肽的蛋白质进行亚细胞定位预测,保留定位于细胞膜外的蛋白质,最后,利用 Kohgpi 程序剔除具有 GPI 锚定位点的蛋白质,最终得到 EHP 的分泌蛋白集合。

1.3 分泌蛋白的功能分析

利用 BLASTP 程序,将预测获得的 EHP 分泌蛋白序列分别与数据库 Nr 和 Swiss-Prot 数据库进行比对,将比对结果中得分最高的同源序列的功能作为对应分泌蛋白的功能参考(李田等, 2013; 罕园园等, 2014)。

1.4 分泌蛋白的序列特征分析

根据 SignalP-4.0 信号肽预测的结果,截取所有分泌蛋白的信号肽序列,并统计分泌蛋白和信号肽序列的长度以及各类氨基酸组成,同时,截取信号肽剪切位点前后各 3 个氨基酸,并利用 WebLogo 对其组成进行统计,然后,利用 MEME 程序预测分泌蛋白和信号肽中的基序。

2 结果与分析

2.1 预测的虾肝肠胞虫分泌蛋白

利用 EuSecPred 流程对获得的 2548 个虾肝肠胞虫蛋白质序列进行预测,发现具有跨膜域的蛋白有 2018 个,1741 个蛋白质具有细胞核定位信号,144 个蛋白质具有信号肽序列,126 个蛋白具有线粒体定位信号,119 个蛋白具有亚细胞定位信号,筛选后最终获得了 109 个分泌蛋白(表 1)。

2.2 虾肝肠胞虫分泌蛋白的功能分类

预测获得 109 个分泌蛋白,对其进行功能注释得出,其中 60 个蛋白质(占预测所得分泌蛋白的 54.05%)为无明确功能信息的蛋白质,另外 49 个分泌蛋白在数据库中检索到了同源蛋白,其中酶类数目最多,包括具有调控宿主细胞增殖和免疫系统功能的酶类:多肽 N-乙酰氨基半乳糖转移酶和泛素羧基末端水解酶;同时,还预测到 1 个孢壁蛋白:SWP7,孢壁蛋白被认为在粘附、侵染以及致病等方面扮演着重要角色;另外,预测到 2 种蛋白抑制因子: α -胰蛋白酶抑制剂重链 H1、亮氨酸拉链假定肿瘤抑制因子 2,这 2 种蛋白抑制因子可能会抑制宿主免疫系统、调控细胞增殖;还预测到 2 种糖蛋白(图 1)。

2.3 虾肝肠胞虫分泌蛋白的序列特征

2.3.1 分泌蛋白与信号肽的长度和氨基酸组成 虾肝肠胞虫 109 个分泌蛋白序列长度范围为 30~700 aa,主要集中在 30~400 aa,平均长度为 219.34 aa,中值为 181.00 aa(图 2);信号肽长度范围在 9~32 aa,主要长度集中在 15~20 aa,平均长度为 18.09 aa,中值为 18.00 aa(图 3)。

表 1 预测的虾肝肠胞虫分泌蛋白基因及序列分析
Tab.1 Coding genes and sequence information of the predicted EHP secreted proteins

基因编号 Gene ID	基因名称 Gene name	基因长度 Length of gene (bp)	GC 含量 Content of GC (%)	蛋白质长度 Length of protein (aa)	信号肽长度 Length of signal peptide (aa)
EHP00_2619	Hypothetical protein	192	30	63	20
EHP00_2497	Ubiquitin carboxyl-terminal hydrolase 7	249	34	82	16
EHP00_2643	Hypothetical protein	264	20	87	23
EHP00_634	Leucine zipper putative tumor suppressor 2 homolog	942	25	313	17
EHP00_1709	Hypothetical protein	369	21	122	15
EHP00_1921	Hypothetical protein	246	22	81	17
EHP00_1834	Hypothetical protein	366	21	121	18
EHP00_1831	Hypothetical protein	849	18	282	21
EHP00_2132	Polypeptide N-acetylgalactosaminyltransferase 12	108	22	35	24
EHP00_1676	Hypothetical protein	720	24	239	15
EHP00_1895	ATP-dependent rRNA helicase spb4	573	21	190	11
EHP00_1099	Serine hydroxymethyltransferase	1398	34	465	25
EHP00_1094	Glucosamine 6-phosphate N-acetyltransferase	543	29	180	17
EHP00_2309	V-type proton ATPase subunit a, vacuolar isoform	597	29	198	15
EHP00_1540	DNA-directed RNA polymerase subunit alpha 2	1122	24	373	18
EHP00_732	Hypothetical protein	951	18	316	18
EHP00_2316	Hypothetical protein	363	21	120	17
EHP00_738	Hypothetical protein	1059	24	352	18
EHP00_739	Prosolanapyrone synthase	228	35	75	22
EHP00_743	C-terminal-binding protein 2	528	24	175	20
EHP00_750	Hypothetical protein	555	34	184	19
EHP00_751	ATP-dependent protease ATPase subunit HslU	627	29	208	26
EHP00_1904	Hypothetical protein	1038	28	345	32
EHP00_1639	Inter-alpha-trypsin inhibitor heavy chain H1	546	24	181	14
EHP00_1636	Gap junction alpha-5 protein	966	30	321	19
EHP00_1635	Hypothetical protein	402	23	133	1
EHP00_2130	Protein C2	210	40	69	28
EHP00_696	Nuclear pore complex protein Nup153	750	25	249	18
EHP00_711	Hypothetical protein	867	21	288	16
EHP00_1530	Hypothetical protein	1794	22	597	19
EHP00_879	Zinc finger protein 778	387	20	128	20
EHP00_878	Protein NETWORKED 4A	444	23	147	16
EHP00_877	Hypothetical protein	480	21	159	20
EHP00_868	Hypothetical protein	1155	18	384	19
EHP00_1803	Uncharacterized protein YggL	600	38	199	24
EHP00_2112	Glycoprotein	174	19	57	20
EHP00_508	Hypothetical protein	459	25	152	17
EHP00_506	Hypothetical protein	420	26	139	19
EHP00_2251	Demethylmenaquinone methyltransferase	765	21	254	21
EHP00_772	ATP-dependent protease ATPase subunit HslU	561	22	186	19
EHP00_773	Linear gramicidin synthase subunit C	1236	24	411	18

续表

基因编号 Gene ID	基因名称 Gene name	基因长度 Length of gene (bp)	GC 含量 Content of GC (%)	蛋白质长度 Length of protein (aa)	信号肽长度 Length of signal peptide (aa)
EHP00_1718	Protocadherin Fat 4	843	28	280	12
EHP00_1408	Probable cell cycle serine/threonine-protein kinase CDC5 homolog	1776	29	591	18
EHP00_168	Hypothetical protein	588	24	195	19
EHP00_2110	Hypothetical protein	429	23	142	23
EHP00_964	Phosphopantetheine adenylyltransferase	423	18	140	17
EHP00_962	Large tegument protein deneddylase	366	19	121	15
EHP00_1331	Hypothetical protein	687	21	228	19
EHP00_2121	Hypothetical protein	243	26	80	19
EHP00_2122	Hypothetical protein	171	24	56	9
EHP00_1752	Translation machinery-associated protein 22	1380	23	459	15
EHP00_944	SWP7	753	34	250	21
EHP00_2111	Protein FAM178B	189	26	62	16
EHP00_268	Hypothetical protein	207	28	68	17
EHP00_266	RNA polymerase II C-terminal domain phosphatase-like 3	684	23	227	23
EHP00_1549	Hypothetical protein	546	23	181	17
EHP00_1547	Hypothetical protein	519	23	172	19
EHP00_1546	Hypothetical protein	654	20	217	18
EHP00_1545	Trigger factor	498	22	165	17
EHP00_1717	Hypothetical protein	216	16	71	14
EHP00_1449	Phosphate acyltransferase	450	25	149	23
EHP00_613	Uncharacterized Sell-like repeat-containing protein C1B3.10c	1755	28	584	15
EHP00_2114	Hypothetical protein	258	29	85	19
EHP00_2574	Hypothetical protein	1257	21	418	16
EHP00_2577	Hypothetical protein	393	20	130	22
EHP00_1194	Hypothetical protein	456	16	151	19
EHP00_1185	Hypothetical protein	438	20	145	22
EHP00_1182	Hypothetical protein	1257	21	418	16
EHP00_2131	UPF0182 protein SACE_1102	180	21	59	19
EHP00_1483	Protein Teyrha-meyrha	639	36	212	19
EHP00_2119	Hypothetical protein	177	21	58	26
EHP00_1595	30S ribosomal protein S17	987	22	328	17
EHP00_2586	Decapping and exoribonuclease protein 1	471	29	156	22
EHP00_2123	Hypothetical protein	141	35	46	13
EHP00_477	Hypothetical protein	840	28	279	22
EHP00_480	Hypothetical protein	1200	26	399	15
EHP00_1976	Histone transcription regulator 3 homolog	366	25	121	18
EHP00_1942	Nebulin-related-anchoring protein	285	22	94	20
EHP00_1612	Hypothetical protein	1035	23	344	15
EHP00_372	Coatomer subunit beta'-2	2286	28	761	23
EHP00_393	Hypothetical protein	516	25	171	20

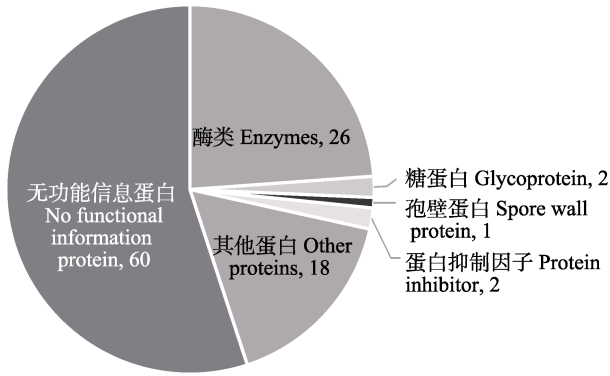


图 1 虾肝肠胞虫分泌蛋白的功能分类

Fig.1 Functional classification of EHP secreted proteins

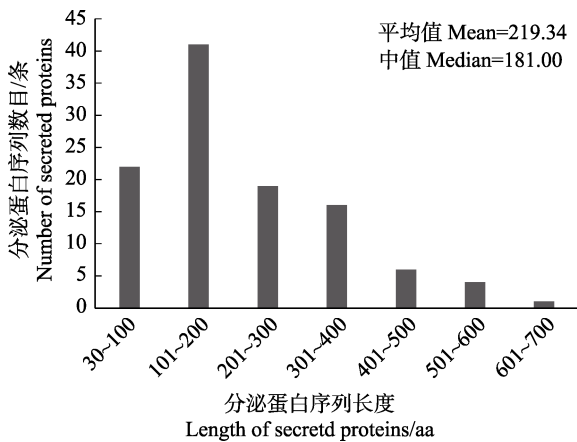


图 2 虾肝肠胞虫分泌蛋白氨基酸长度分析

Fig.2 Analysis of the lengths of EHP secreted proteins

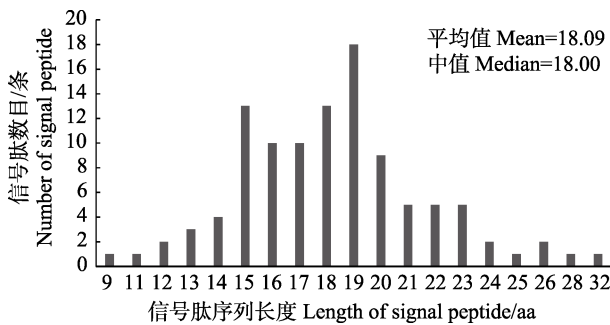


图 3 虾肝肠胞虫分泌蛋白信号肽长度分析

Fig.3 Analysis of the signal peptides length of secreted proteins in EHP

对分泌蛋白和信号肽的氨基酸组成统计分析发现,分泌蛋白和信号肽均以疏水性氨基酸为主(图 4)。分泌蛋白序列中疏水性氨基酸占比为 42.57%,其次是亲水性氨基酸,占比为 29.30%。而信号肽同样以疏水性氨基酸为主,占比高达 68.93%,而亲水性氨基酸比例仅为 22.42%。

2.3.2 信号肽剪切位点处氨基酸组成 统计信号肽剪切位点前后各 3 个氨基酸的组成(图 5),分析发

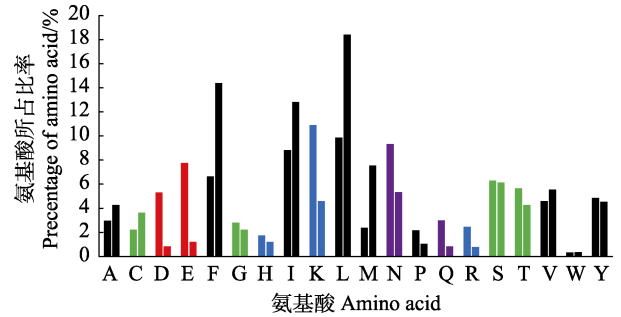


图 4 虾肝肠胞虫分泌蛋白的氨基酸组成

Fig.4 Amino acid composition of EHP secreted proteins

A: 丙氨酸 Ala; C: 半胱氨酸 Cys; D: 天冬氨酸 Asp; E: 谷氨酸 Glu; F: 苯丙氨酸 Phe; G: 甘氨酸 Gly; H: 组氨酸 His; I: 异亮氨酸 Ile; K: 赖氨酸 Lys; L: 亮氨酸 Leu; M: 甲硫氨酸 Met; N: 天冬酰胺 Asn; P: 脯氨酸 Pro; Q: 谷氨酰胺 Gln; R: 精氨酸 Arg; S: 丝氨酸 Ser; T: 苏氨酸 Thr; V: 缬氨酸 Val; W: 色氨酸 Trp; Y: 酪氨酸 Tyr。每种氨基酸上方左侧柱为在所有分泌蛋白序列中占有的比率,右侧柱为在所有信号肽序列中占有的比率。黑色为疏水性氨基酸,绿色为亲水性氨基酸,红色为酸性氨基酸,蓝色为碱性氨基酸,紫色为中性氨基酸。下同 Left column above each amino acid is the ratio in all secreted protein sequences. Right column above each amino acid is the ratio in all signal peptide sequences. Black column: Hydrophobic amino acid; Green column: Hydrophilic amino acid; Red column: Acid amino acid; Blue column: Basic amino acid. The same as below

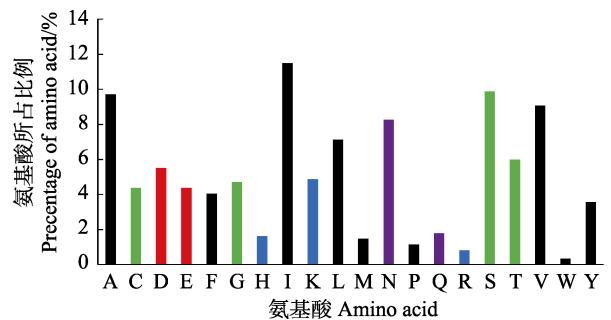


图 5 虾肝肠胞虫分泌蛋白信号肽剪切位点处氨基酸的组成

Fig.5 The composition of amino acids flanking the signal peptide splice site in the secretory proteins of EHP

现,与整个信号肽氨基酸组成类似,信号肽剪切位点氨基酸组成同样以疏水性氨基酸为主,其占比为 47.90%,其中,异亮氨酸(I)和丙氨酸(A)含量最丰富;亲水性氨基酸占比为 34.95%,其中,丝氨酸(S)和天冬酰胺(N)所占比例较高。从各个位置的氨基酸组成分析,上游的-3 位主要为异亮氨酸(I),-2 位主要为苯丙氨酸(F),紧邻信号肽剪切位点的-1 和 1 位主要为丙氨酸(A),下游的 3 位同样主要为异亮氨酸(I)(图 6)。

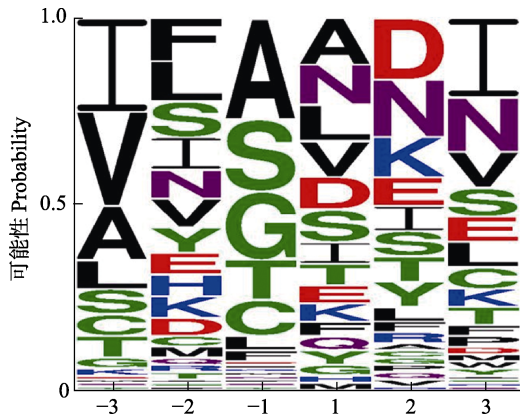


图6 虾肝肠胞虫分泌蛋白信号肽剪切位点处氨基酸组成
Fig.6 Composition patterns of amino acids flanking the signal peptide splice site in the secreted proteins of EHP

2.3.3 分泌蛋白和信号肽中的基序 对虾肝肠胞虫信号肽区域的基序分析,发现了1种氨基酸组成模式: NV[VT][IK]CA[ED][SA](图7)。对其非信号肽区分析发现3种基序: VAYDMFLSRCILHMLDVMMPLYVE

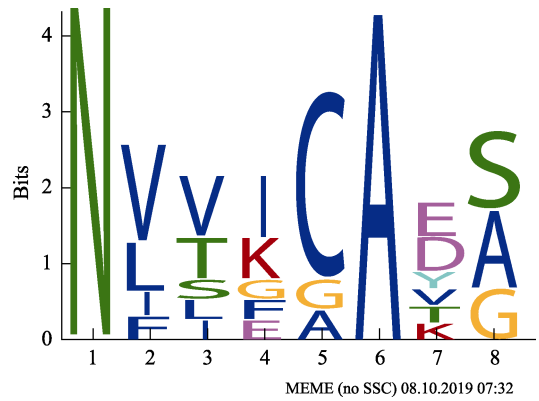


图7 虾肝肠胞虫分泌蛋白信号肽基序
Fig.7 Motif in signal peptide of EHP secreted proteins

NESFMDQVAEVFCFNWTATEFYNSIIR, [QM]L[RK][FA]QN[TR][ND]G[SC][NK]D[NE][KE]I[SK][QE]L[KF][IE]KIK[TE][MI]C[NK][RK]I[ED][LA][IF][VI][NF][MK][VI][VE]Q[QA][QM][TQ][YN][QE][LI]K[ME]DD[PH]和 SCFGKFSFPISNRSAEYFKTVYDQWNETLVKIPV KIYRRTL(图8)。

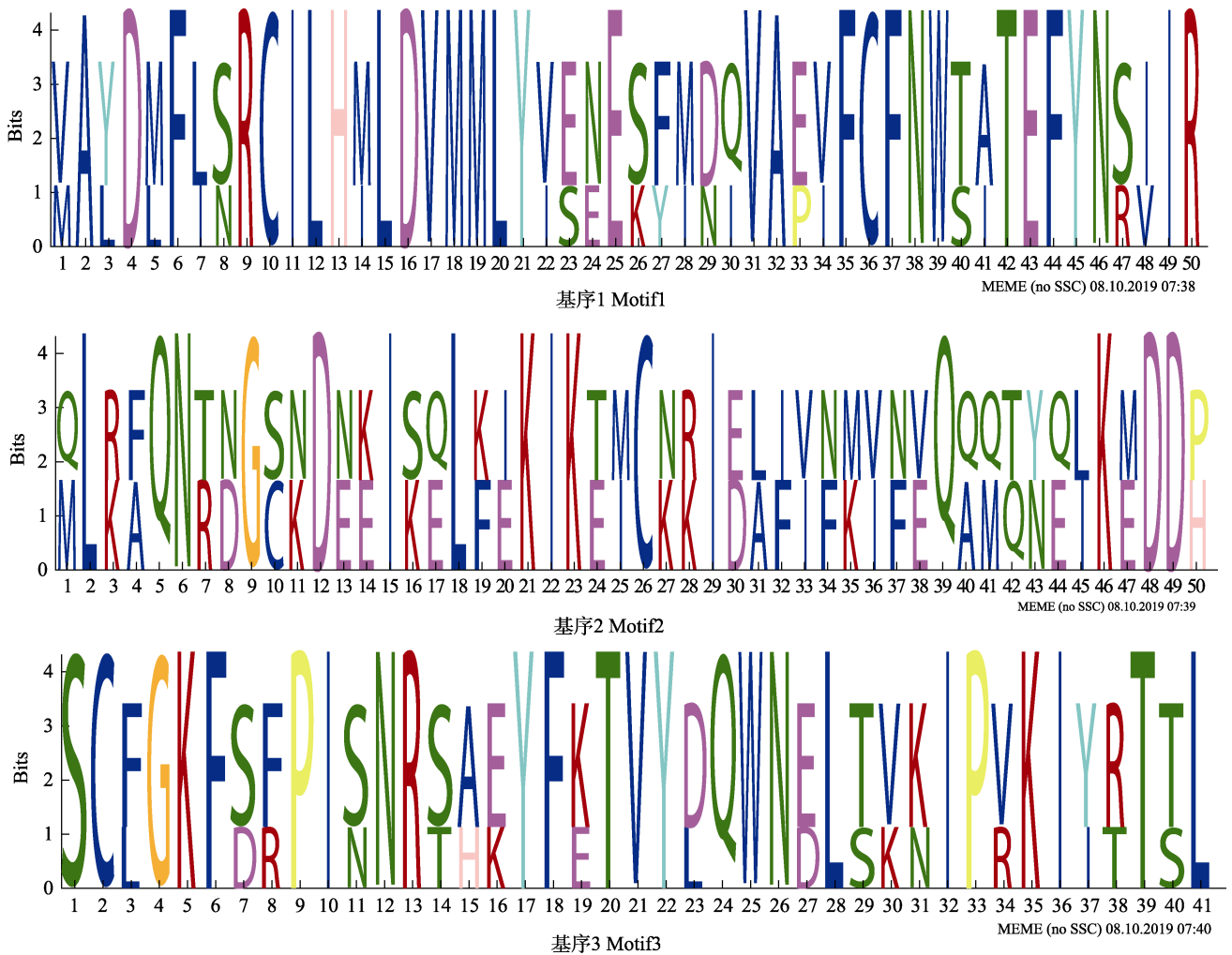


图8 虾肝肠胞虫非信号肽区基序
Fig.8 Motif in non-signal peptide of EHP secreted proteins

3 讨论

对于专性胞内寄生的微孢子虫来说,目前,仅有一些关于孢壁蛋白分离方法的报道(吴正理等,2007),尚没有分泌蛋白的分离方法。生物信息学算法的快速发展为成功预测微孢子虫的分泌蛋白提供了可能。本研究基于虾肝肠胞虫基因组数据库,运行 EuSecPred 2.0 在线流程在全基因组 2548 条蛋白质中进行筛选和预测,最终获得了 109 条分泌蛋白,这些分泌蛋白组成以疏水性氨基酸为主,这与其信号肽和剪切位点前后各 3 个氨基酸组成相一致。

本研究预测获得的 109 条分泌蛋白,其中,60 个是无功能注释的蛋白质,这部分蛋白质可能是虾肝肠胞虫分泌的特有蛋白质。在有明确功能注释的 49 个分泌蛋白中,有些是与微孢子虫黏附和入侵相关的蛋白,如孢壁蛋白(Spore wall protein, SWP)、钙粘蛋白(Cadherin)和多肽 N-乙酰氨基半乳糖转移酶 12 (Polypeptide N-acetylgalactosaminyltransferase 12, GALNT12)。孢壁(Spore wall)是微孢子虫最外层的结构,在侵染过程中,孢壁应最先与宿主细胞接触。孢壁蛋白作为孢壁的主要成分,对维持孢壁稳固结构和保护胞内原生质起着重要作用,同时直接参与侵染过程中与宿主细胞的特异性识别过程。Jaroenlak 等(2018)鉴定并分离了 EHP 的第 1 个孢壁蛋白(EhSWP1),在其 N 末端发现了 3 个肝素结合基序,进一步的体外结合实验、竞争实验和诱变研究表明, EhSWP1 能够和肝素结合进而引发入侵。同样, Yang 等(2018)研究证明,家蚕微孢子虫(*Nosema bombycis*)的部分孢壁蛋白在粘附和侵染过程中起关键作用,本研究中筛选的 SWP7 也可能与 EHP 的入侵相关。钙粘蛋白是 1 种 Ca 依赖的细胞粘着糖蛋白,可能与宿主细胞的靶向信号识别并粘附;同时, Frixione 等(1994)研究表明, Ca^{2+} 会以某种方式触发孢子萌发。GALNT12 是合成 O-糖链的起始酶,而粘蛋白主要由 O-糖链组成,参与很多与粘附相关的过程, O-糖链还可调控细胞表面受体的表达,进而影响细胞的生长和凋亡(刘可人等,2006)。因此,推测 EHP 侵染过程中分泌的 GALNT1 可能参与到宿主细胞的识别和粘附。

在这些分泌蛋白中还有一些抑制宿主免疫功能的分泌蛋白。 α -胰蛋白酶抑制剂重链 H1(Inter-alpha-trypsin inhibitor heavy chain H1, ITIH1)可调控细胞外基质稳定并抑制 C5 转化酶活性(Zhuo *et al*, 2008),进而抑制宿主免疫系统。泛素羧基末端水解酶(Ubiquitin carboxyl-terminal hydrolase, UCH)可水解被泛素(Ubiquitin)标记的需要被分解的蛋白质,也可

以标记跨膜蛋白,然后通过酶促反应相互连接,进而介导靶蛋白的降解或跨膜运输。Ciechanover 等(2000)研究表明,泛素蛋白水解系统在广泛的细胞基本过程中发挥重要作用。其中,包括调节细胞周期、调节免疫和炎症反应、控制信号转导途径、发育和分化、DNA 修复等。这些复杂的过程是通过蛋白质的单个或子集的特异性降解来控制的。虾肝肠胞虫 UCH7 在侵染过程中的信号传递及调控宿主免疫方面行使何种功能尚需研究进行明确。

另外,还有调控细胞增殖的蛋白,比如亮氨酸拉链假定肿瘤抑制因子 2 (Leucine zipper putative tumor suppressor, LZTS2),拥有特定的亮氨酸拉链结构域(LZ),参与 Wnt/ β -catenin 信号通路调控 β -catenin 的表达及细胞内分布,与 NF- κ B 拥有广泛的相互作用,从而调控细胞增殖和凋亡(Peng *et al*, 2011; 王泉雄等, 2014)。在微管系统中, LZTS2 也有重要的调节作用,可抑制细胞的有丝分裂和迁移(张轶等, 2016)。EHP 分泌的 LZTS2 是否能抑制宿主细胞的增殖和迁移从而抑制宿主细胞行使其功能尚需进一步研究。

分泌蛋白在多种寄生虫系统中介导粘附和侵袭,对于明确胞内寄生虫的感染机制至关重要。本研究初步筛选了 109 个分泌蛋白,接下来进一步对这些分泌蛋白进行功能研究对于理解虾肝肠胞虫的入侵机理具有重要意义。

参 考 文 献

- Nguyen Ba A, Pogoutse A, Provart N, *et al*. NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. BMC bioinformatics, 2009, 10(1): 202
- Bailey TL, Boden M, Buske FA, *et al*. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Research, 2009, 37(S2): W202–W208
- Brameier M, Krings A, MacCallum RM. NucPred—Predicting nuclear localization of proteins. Bioinformatics, 2007, 23(9): 1159–1160
- Ciechanover A, Orian A, Schwartz AL. Ubiquitin-mediated proteolysis: Biological regulation via destruction. BioEssays, 2000, 22(5): 442–451
- Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. European Journal of Biochemistry, 1996, 241(3): 779–786
- Crooks GE, Hon G, Chandonia JM, *et al*. WebLogo: A sequence logo generator. Genome research, 2004, 14(6): 1188–1190
- Druzhinina IS, Shelest E, Kubicek CP. Novel traits of *Trichoderma* predicted through the analysis of its secretome. FEMS Microbiology Letters, 2012, 337(1): 1–9
- Emanuelsson O, Nielsen H, Brunak S, *et al*. Predicting subcellular

- localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 2000, 300(4): 1005–1016
- Fankhauser N, Mäser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*, 2005, 21(9): 1846–1852
- Frixione E, Ruiz L, Undeen AH. Monovalent cations induce microsporidian spore germination *in vitro*. *Journal of Eukaryotic Microbiology*, 1994, 41(5): 464–468
- Gupta N, Hartmann A, Lucius R, *et al.* The obligate intracellular parasite *Toxoplasma gondii* secretes a soluble phosphatidylserine decarboxylase. *Journal of Biological Chemistry*, 2012, 287(27): 22938–22947
- Han YY, Ma KL. A genome-wide prediction and analysis of secreted proteins of *Tupaia chinensis*. *Chinese Journal of Comparative Medicine*, 2014, 24(4): 33–37 [罕园园, 马开利. 树鼩全基因组分泌蛋白的预测分析. *中国比较医学杂志*, 2014, 24(4): 33–37]
- Horton P, Park KJ, Obayashi T, *et al.* WoLF PSORT: Protein localization predictor. *Nucleic acids research*, 2007, 35(S2): W585–W587
- Jaroenlak P, Boakye DW, Vanichviriyakit R, *et al.* Identification, characterization and heparin binding capacity of a spore-wall, virulence protein from the shrimp microsporidian, *Enterocytozoon hepatopenaei* (EHP). *Parasites and Vectors*, 2018, 11(1): 177
- Katinka MD, Duprat S, Cornillot E, *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, 2001, 414(6862): 450–453
- Krogh A, Larsson B, von Heijne G, *et al.* Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *Journal of Molecular Biology*, 2001, 305(3): 567–580
- Li T, Liu XL, Han B, *et al.* A genome-wide prediction and analysis on secreted proteins of *Nosema bombycis*. *Science of Sericulture*, 2013, 39(2): 295–301 [李田, 刘显林, 韩冰, 等. 家蚕微孢子虫全基因组分泌蛋白的预测分析. *蚕业科学*, 2013, 39(2): 295–301]
- Liu KR, Jin MF, Wu SL. Prediction of O-glycosylation site and catalytic characteristics of glycosylation enzyme. *Chemistry of Life*, 2006, 26(1): 25–27 [刘可人, 金美芳, 吴士良. O-糖基化位点预测及糖基化酶催化特点. *生命的化学*, 2006, 26(1): 25–27]
- Liu Z, Zhang QL, Wan XY, *et al.* Development of real-time PCR assay for detecting microsporidian *Enterocytozoon hepatopenaei* and the application in shrimp samples with different growth rates. *Progress in Fishery Sciences*, 2016, 37(2): 119–126 [刘珍, 张庆利, 万晓媛, 等. 虾肝肠胞虫 (*Enterocytozoon hepatopenaei*) 实时荧光定量 PCR 检测方法的建立及对虾样品的检测. *渔业科学进展*, 2016, 37(2): 119–126]
- McKerrow JH, Sun E, Rosenthal PJ, *et al.* The proteases and pathogenicity of parasitic protozoa. *Annual Review of Microbiology*, 1993, 47(1): 821–853
- Peng Y, Clark C, Luong R, *et al.* The leucine zipper putative tumor suppressor 2 protein LZTS2 regulates kidney development. *Journal of Biological Chemistry*, 2011, 286(46): 40331–40342
- Petersen TN, Brunak S, Heijn GV, *et al.* SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, 2011, 8(10): 785–786
- Qiao Y, Shen H, Wan XH, *et al.* Preliminary isolation and morphological observation of *Enterocytozoon hepatopenaei* infecting the shrimp *Penaeus vannamei*. *Journal of Fishery Sciences of China*, 2018, 25(5): 1051–1058 [乔毅, 沈辉, 万夕和, 等. 南美白对虾肝肠胞虫的分离及形态学观察. *中国水产科学*, 2018, 25(5): 1051–1058]
- Song ZL, Dong X, Zhao RH, *et al.* Evaluation on the detection of *Enterocytozoon hepatopenaei* in pooled DNA samples of *Litopenaeus vannamei* based on TaqMan qPCR. *Progress in Fishery Sciences*, 2019, 40(3): 122–132 [宋增磊, 董宣, 赵若恒, 等. 基于 TaqMan qPCR 检测凡纳滨对虾样品中虾肝肠胞虫并样检测方法的评价. *渔业科学进展*, 2019, 40(3): 122–132]
- Tourtip S, Wongtripop S, Stentiford GD, *et al.* *Enterocytozoon hepatopenaei* sp. nov. (Microsporida: Enterocytozoonidae), a parasite of the black tiger shrimp *Penaeus monodon* (Decapoda: Penaeidae): Fine structure and phylogenetic relationships. *Journal of Invertebrate Pathology*, 2009, 102(1): 21–29
- Wang XX, Yang G, Zhang DM, *et al.* LZTS2 tumor suppressor gene and advancements in tumor research. *Chinese Journal of Clinical Oncology*, 2014, 41(20): 1328–1332 [王泉雄, 杨光, 张大明, 等. LZTS2 抑癌基因与肿瘤相关性的研究进展. *中国肿瘤临床*, 2014, 41(20): 1328–1332]
- Wang Y, Dang X, Ma Q, *et al.* Characterization of a novel spore wall protein NbSWP16 with proline-rich tandem repeats from *Nosema bombycis* (microsporidia). *Parasitology*, 2015, 142(4): 534–542
- Wu ZL, Tan XH, Pan GQ, *et al.* Extraction of spore wall proteins of *Nosema bombycis* with improved methods. *Science of Sericulture*, 2007, 33(1): 62–66 [吴正理, 谭小辉, 潘国庆, 等. 家蚕微孢子虫 (*Nosema bombycis*) 孢壁蛋白提取方法的优化研究. *蚕业科学*, 2007, 33(1): 62–66]
- Yang D, Pan L, Chen Z, *et al.* The roles of microsporidia spore wall proteins in the spore wall formation and polar tube anchorage to spore wall during development and infection processes. *Experimental Parasitology*, 2018, 187: 93–100
- Zhang Y, Yu WW, Tao RY, *et al.* LZTS2: A novel tumor suppressor play an important role in tumor. *Progress in Modern Biomedicine*, 2016, 16(12): 2372–2375 [张轶, 余稳稳, 陶瑞雨, 等. LZTS2 抑癌基因在肿瘤中的作用机制. *现代生物医学进展*, 2016, 16(12): 2372–2375]
- Zhuo L, Kimata K. Structure and function of inter- α -trypsin inhibitor heavy chains. *Connective Tissue Research*, 2008, 49(5): 311–320

Genome-Wide Prediction and Analysis of Secreted Proteins of *Enterocytozoon hepatopenaei*

NING Zijian, JIANG Hongbo, LIU Qi, BAO Jie^①

(College of Animal Husbandry and Veterinary Medicine, Shenyang Agricultural University, Shenyang 110866)

Abstract Secreted proteins are synthesized in cells and secreted to function outside the cell. They play an important role in the manipulation of host cells and the virulence of many eukaryotic protozoan parasites. *Enterocytozoon hepatopenaei* (EHP) is an obligate intracellular parasite that can infect a variety of economically valuable shrimp species, and is one of the most serious diseases affecting global shrimp production. Here, we predict the secreted proteins of the EHP genome with the EuSecPred2.0 pipeline and analyzed the length of these secreted proteins, the length of the signal peptide, and the amino acid distribution at the cleavage site, and annotated the function of secreted proteins. The results show that the length of the secreted proteins ranged from 30 to 400 amino acid residues, the signal peptides covered approximately 9~32 amino acids, and the cleavage sites of the signal peptides were mainly composed of hydrophobic amino acids. Motif analysis revealed a NV[VT][IK]CA[ED][SA] motif in signal peptides. Functional annotation of proteins revealed that a variety of key proteins are involved in adhesion and infection of microsporidia, regulation of cell cycle, and immune response. The results shed light on the infection mechanism of EHP and provide a theoretical basis for further defining the pathogenic factors of EHP.

Key words *Enterocytozoon hepatopenaei*; Genome; Secreted protein; Signal peptide; Motif

① Corresponding author: BAO Jie, E-mail: baojie@syau.edu.cn