

预测和分析血吸虫含有 EF-hand 结构域的表膜蛋白的功能

于复东^{①②③} 康 斌^{①③} 李园园^{①③*} 李亦学^{①③*}

(① 中国科学院上海生命科学研究院系统生物学重点实验室生物信息中心, 上海 200031; ② 中国科学院研究生院, 上海 200031;

③ 上海生物信息技术研究中心, 上海 200235. * 联系人, E-mail: yxli@sibs.ac.cn, yyli@scbit.org)

摘要 血吸虫病仍然是严重危害人类健康的重要寄生虫疾病之一。在血吸虫中发现一类特殊的表膜蛋白, 它们含有 EF-hand 结构域, 但是对它们功能的了解除了可以作为抗原之外几乎一无所知。蛋白-蛋白相互作用、位点特异性变异和糖基化位点修饰等 *in silico* 分析的结果表明, 这类表膜蛋白不仅可以与具有免疫调节功能的宿主蛋白进行相互作用来帮助血吸虫主动调节宿主的免疫反应, 还可以通过位点特异性变异、糖基化修饰自身那些可以被宿主免疫攻击所识别的位点, 以帮助血吸虫被动调节宿主免疫攻击。此外, 对表膜蛋白 C 端区域的分析表明, 这些蛋白还可以协助血吸虫抑制宿主多核细胞的细胞趋化和宿主的 IgG4 免疫反应来调节和逃避宿主的免疫攻击。总之, 分析结果暗示, 这些表膜蛋白可以协助血吸虫在宿主-寄生虫相互作用过程中调节和逃避宿主的免疫攻击, 起到自我保护的作用。

关键词 血吸虫 宿主-寄生虫相互作用 表膜蛋白 EF-hand 结构域 生物信息学分析

血吸虫病仍然是热带亚热带国家和地区严重威胁人类健康的寄生虫疾病之一, 其中日本血吸虫、曼式血吸虫和埃及血吸虫是引起血吸虫病的主要寄生虫^[1]。血吸虫有复杂的生活史, 经历发生在终宿主体内的有性世代和中间宿主螺内的无性繁殖世代过程, 包括尾蚴、童虫、成虫、卵和毛蚴等重要的发育阶段。除毛蚴之外的发育阶段, 血吸虫都有保护其免受宿主免疫系统攻击的表膜结构^[2]。这种表膜结构是一种类似三明治的结构, 包括最外层的与宿主相互作用的体表膜、中间位置的合胞体基质层和最里层的基底膜结构^[3]。在该结构中镶嵌了许多蛋白, 其中大多数是与细胞的能量代谢, 线粒体基质、质膜以及细胞骨架的跨膜转运, 以及囊泡的转运等功能相关, 这些说明该表膜结构是拥有多种功能的结构^[4,5]。

有趣的是, 在这些表膜蛋白中发现一类含有 EF-hand 结构域的蛋白^[4,6], 这类蛋白很少在胞外的结构中被发现。EF-hand 结构域是一种具有钙结合位点的螺旋-中央环-螺旋的结构, 其中两螺旋几乎是垂直的, 这有利于它们结合钙离子^[7]。常见的含有 EF-hand 结构域的蛋白按照功能可以分为调节型蛋白和结构组成型蛋白两类, 前者在结合钙离子之后产生构象的变化, 并且有酶活性的改变; 后者结合钙离子后并

没有产生类似的变化, 它们似乎仅是起到胞内钙离子水平的缓存作用^[7]。目前研究发现, 含有 EF-hand 结构域的蛋白几乎都定位在胞质和细胞膜中^[8], 很少在表膜结构中表达。到目前为止只在一些包括日本血吸虫^[9,10]、曼式血吸虫^[6]、埃及血吸虫^[11]、肝片吸虫^[12]、大片吸虫^[13]和华支睾吸虫等吸虫的表膜中有这类蛋白的表达。由此看来, 这类特殊的 EF-hand 结构域的表膜蛋白可能在吸虫中发挥其特殊的作用, 但目前除了可以作为抗原诱导宿主的免疫反应之外的功能几乎没有报道^[6,9,11,14]。

为了更好地预测和分析这类 EF-hand 结构域的表膜蛋白的更多功能, 本研究对这类表膜蛋白进行了生物信息学分析。首先, 用 *in silico* 的方法找到含有 EF-hand 结构域的表膜蛋白序列, 利用系统发育分析把各个蛋白划分到其合适的亚家族中, 这种分类有助于对每一亚家族内蛋白进行进一步的特异性变异位点的分析; 其次, 利用蛋白-蛋白相互作用和蛋白结构域等分析得到这些表膜蛋白更多的可能功能; 最后, 结合特异性变异位点、不同发育阶段的表达模式(pattern)以及糖基化模式等分析, 发现这些表膜蛋白可以帮助血吸虫逃避或者调节宿主的免疫攻击。这些结果有利于扩大对血吸虫的表膜蛋白在宿主和

寄生虫相互作用过程中的功能理解.

1 材料和方法

(i) 数据收集. 2006年5月10日, 从NCBI数据库里直接获得6个物种(日本血吸虫、曼式血吸虫、埃及血吸虫、肝片吸虫、大片吸虫和华支睾吸虫)中所有含有EF-hand结构域的表膜蛋白序列. 序列收集的原则是: 全长的蛋白序列同时有全长的编码区序列(coding DNA sequence, CDS); 在表膜中表达并且有EF-hand结构域(或类似的EF-hand结构域). 此外, 具有相同编码区序列的只保留一条, 其余的都作为冗余序列被去除掉.

本研究所需的日本血吸虫和曼式血吸虫的表达标签序列(expressed sequence tags, EST)分别从国家人类基因组南方研究中心(Chinese Human Genome Center at Shanghai, CHGC, <http://schistosoma.chgc.sh.cn/sj-proteome/>)和巴西圣保罗大学(University of Sao Paulo, USP, <http://verjo18.iq.usp.br/schisto/>)获得.

(ii) 蛋白分子的分类. 本研究所需的多序列联配分析都是使用缺省值条件的ClustalW (版本号是1.83); 聚类分析使用Mega (3.1 版本)软件^[15]中的Neighbor-joining算法和JTT (Jones-Taylor-Thornton)矩阵进行系统发育分析; 进化树的显著性分析利用重复1000次的bootstrap算法进行统计分析.

(iii) 蛋白特异性变异位点分析. 所有亚家族蛋白成员都进行配对的多序列比对, 根据蛋白多序列比对的结果得到对应核酸序列的多序列比对结果, 然后使用PAML^[16](3.14 版本)中codeml的branch-site和M2a模型以及去掉配对间隔(gap)的方法对亚家族中每个家族成员进行位点特异性的选择压力分析.

(iv) 结构域-结构域的相互作用. 利用Pfam和Prosite数据库的同源性对表膜蛋白结构域进行特征分析. 针对表膜蛋白的C端在这两个数据库中没有同源结构域的情况, 我们进一步使用BLASTP和PSI-BLASTP对这类表膜蛋白在NCBI的NR(非冗余数据库)作进一步的同源搜索(E 的初始值是 10^{-5}).

为了进一步提高找出与这些表膜蛋白的相互作用蛋白成员的敏感度, 我们开发了一套整合Pfam和DIP数据库的蛋白相互作用信息的方法. 首先, 从Pfam数据库中抽提出可以与EF-hand结构域(或者EF-hand的类似结构域)相互作用的蛋白序列, 从DIP数据库(蛋白相互作用数据库)中找到这些抽提出来的蛋白序列的同源序列, 同时保证这些同源序列的物

种可以作为血吸虫的宿主, 这样就可以建立这些同源序列蛋白与吸虫EF-hand结构域表膜蛋白的可能的相互作用关系, 把这些吸虫的宿主蛋白序列存放到EFInterDatabase的数据库中. 其次, 从DIP数据库中找到与EFInterDatabase的数据库存储蛋白的相互作用蛋白, 用BLASTP(E 的初始值是0.1)对这些得到的相互作用蛋白进行第2次同源搜索, 并从中找出吸虫表膜蛋白的同源蛋白(含有EF-hand或与其类似的结构域). 最后, 利用DIP数据库中蛋白间的相互作用关系, 以及第2次搜索结果得到的吸虫表膜蛋白的同源蛋白的桥梁作用, 建立了EFInterDatabase数据库中吸虫宿主蛋白与本文分析的吸虫表膜蛋白的相互作用关系, 把具有这种相互作用关系的蛋白对(包括吸虫EF-hand结构域的表膜蛋白和与之相互作用的宿主蛋白)以及第2次搜索结果得到的蛋白存放到SecondDatabase数据库中.

2 结果

2.1 吸虫中含有EF-hand结构域的表膜蛋白的识别、描述和分类

为了更好地识别和确认现有数据库中所有EF-hand结构域的表膜蛋白, 用Sm22.6亚家族的一个成员序列(GenBank登录号为: AAA29856.1)对NCBI的NR数据库进行同源性搜索(E 的初始值是 10^{-5}), 得到21条非冗余的同源蛋白序列, 并发现这些序列只存在于吸虫, 包括日本血吸虫、曼式血吸虫、埃及血吸虫、肝片吸虫和大片吸虫, 以及华支睾吸虫(网络版附表1)中, 说明它们可能是吸虫特有的表膜蛋白. 多序列比对结果表明, 这些表膜蛋白的N端序列富含天冬氨酸和谷氨酸, 说明它们具有亲水性的特点. TMHMM软件^[17]的预测结果显示, 这些表膜蛋白C端可能具有类似跨膜结构而使其具有疏水性特点. SignalP (3.0 版本)^[18]的信号肽预测结果显示, Sja56755894基因具有信号肽结构, 这暗示了Sja567-55894蛋白可以在胞内表达之后分泌到表膜结构上. 根据Pfam和Prosite数据库的同源搜索结果, 这些吸虫表膜蛋白序列在N端有两个EF-hand或者类似的结构域以及一个没有结构域同源信息的C端区域. 这种表膜蛋白结构的特殊性及其只在吸虫表膜中表达的特异性说明这类表膜蛋白具有不同于那些在胞质和胞膜中具有EF-hand结构域蛋白的功能.

目前, 对这些蛋白的注释只是简单的根据其蛋白分子的大小, 这种简单的注释所代表的分类结果

并不能很好地体现各个基因间的相互进化关系. 为了更好地展现各个含有 EF-hand 结构域表膜蛋白之间的进化关系, 并进一步对各个亚家族内的成员进行位点特异性变异的分析, 需要对这些表膜蛋白做相应的多序列比对和进化分析. 进化分析的结果显示, 这些表膜蛋白可以划分为几个亚家族, 即 Sm20.8, Sm22.6, Sm21.7, Fasciolidae 亚家族和一个包含 Sja56754688 和 Sja56756302 的新的没有基因注释信息的亚家族, Sja56757349 序列在进化上更靠近 Sm22.6 亚家族(图 1). 从进化分析的结果中可以看出, 没有基因注释信息的新的亚家族(包含 Sja56754688 和 Sja56756302)在进化上介于 Sm21.7 和 Fasciolidae 亚家族之间, 另一没有注释信息的成员 Sja56755894 属于 Sm20.8 亚家族. 通过这种系统发育的分类分析, 没有注释的基因和已经注释过的基因之间的关系被很好地体现出来, 同时也是对未注释基因的一种功能注释. 这种分类结果有利于进一步对其每个亚家族成员进行深入的分析.

2.2 血吸虫通过主动调节宿主免疫反应来阻碍宿主的免疫攻击

血吸虫能够在宿主体内发育成熟说明血吸虫有

主动调节宿主的免疫攻击从而达到其生存的能力. 本研究利用蛋白-蛋白相互作用和保守的基序(motif)两方面的功能分析来探索血吸虫自身主动调节宿主免疫攻击的策略.

通常, 蛋白的功能可以根据其同源蛋白或者与其相互作用的蛋白的功能来推断. 为了更好地了解这些吸虫表膜蛋白在吸虫感染过程中可能的功能, 我们开发了一套整合 Pfam 和 DIP 数据库的蛋白功能信息的方法, 该方法的结果存放在 SecondDatabase 数据库中, 该数据库存放着吸虫-宿主的相互作用的蛋白对, 在这些蛋白作用对中至少有一个蛋白对成员是吸虫表膜蛋白的同源蛋白. 根据这些相互作用蛋白对的功能可以推测 SecondDatabase 数据库中吸虫表膜蛋白的功能, 蛋白间的功能相互作用结果表明, 这些表膜蛋白可能通过与其宿主中同源蛋白的相互作用来体现宿主-寄生虫相互作用(网络版附表 2). 在 SecondDatabase 数据库中有几组相互作用蛋白对是值得关注的: MRP-8 (DIP:1165N)和MRP-14 (DIP:1166N)蛋白对 [19]在调节炎症活化过程中起重要作用 [20]; P2BA_HUMAN (DIP: 6095N)和CALB_HUMAN (DIP: 6096N)在调节 T 细胞活化过程中发挥重要作用 [21].

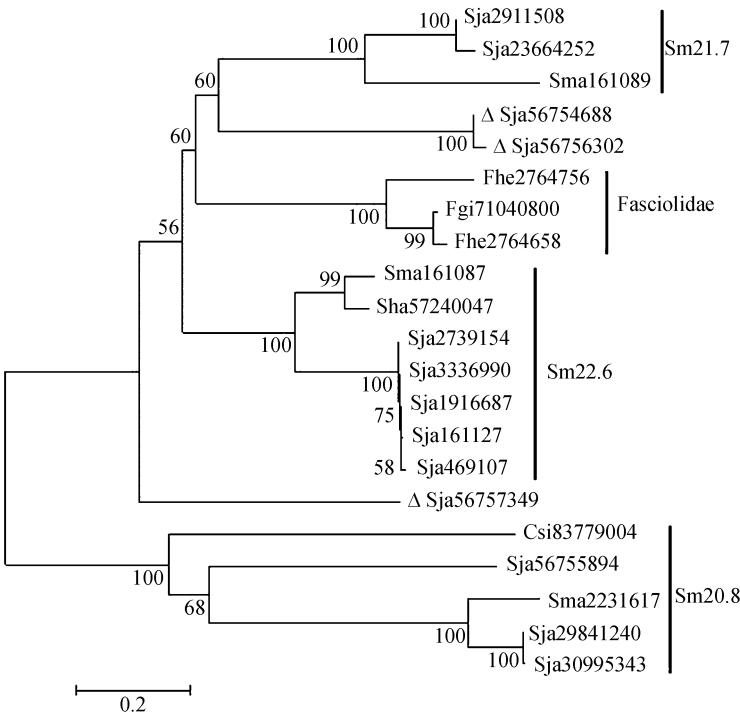


图 1 含有 EF-hand 结构域的吸虫表膜蛋白的分类结果

在该结果中基因名字是物种的缩写后缀 NCBI 中该基因的 GI 号码, 如基因 Sja2911508 是日本血吸虫的缩写 Sja 和 GI 号码(2911508)组成

这两组相互作用对所对应的吸虫的同源表膜蛋白可能通过与其宿主蛋白相互作用调节宿主的免疫反应, 这样有利于吸虫在感染过程中自身的生存. 除了这两组涉及免疫反应的蛋白作用对之外, 仍有一些涉及其他重要反应的蛋白对. 例如, CALM_HUMAN (DIP:538N)和ATB4_HUMAN (DIP:6128N)蛋白对参与调控内源钙离子水平^[22], 这暗示了吸虫中对应的同源蛋白有助于胞内钙离子水平调控, 这个过程可能与其他一些位于表膜上的蛋白, 如钙蛋白酶(calpain)、钙网织蛋白(calreticulin)和钙调磷酸酶(calcineurin)^[23]起协同作用; TPCS_RABIT(DIP: 6059N)和TRIF_RABIT (DIP: 6060N)涉及肌肉收缩过程的调控^[24]. 这些蛋白-蛋白相互作用的结果暗示了血吸虫可以间接或者直接利用其宿主同源蛋白参与宿主的免疫反应过程和其他发生在表膜的相关细胞过程来完成血吸虫生存的需求.

这类表膜蛋白除了N端有两个EF-hand结构域之外, 其C端区域可能具有尚未注释的功能. 为了进一步研究其C端区域的功能, 用BLASTP和PSI-BLAST对这类表膜蛋白在NCBI的非冗余蛋白库(NR)中做同源性搜索. 结果显示, 一条细粒棘球绦虫(*Echinococcus granulosus*)表膜蛋白(EgTeg, GenBank登录号: AAX20156.1)可以与这些吸虫表膜蛋白的C端匹配得很好. EgTeg蛋白在细粒棘球绦虫中可以通过抑制宿主的细胞趋化、诱导宿主的慢性炎症反应来逃避宿主的免疫攻击^[25]. 其中, EgTeg蛋白的两肽段(肽段1和肽段2)发挥了这些重要功能, 而且这两肽段与吸虫表膜蛋白的C端区域非常保守(图2). 这两段肽段的保守性表明, 这类血吸虫的表膜蛋白可能具有与EgTeg蛋白类似的功能. 这些表膜蛋白的C端的功能和N端的EF-hand结构域的相关功能说明, 吸虫可以通过这些表膜蛋白来主动调控宿主的免疫攻击.

2.3 血吸虫通过较小幅度的序列变异和不同阶段的基因表达来调节宿主的免疫监督

抗原变异有助于血吸虫在宿主体内的生存, 多种抗原变异对寄生虫逃避宿主的免疫监督非常重要. 在抗原变异中, 基因突变、基因重组和基因转变(switching)起主要的作用, 而且抗原的翻译后修饰可能掩盖宿主免疫系统识别的位点, 从而有助于寄生虫的存活. 血吸虫抗原蛋白的糖基化修饰可以很好地体现这种抗原变异的特点^[26,27]. 为了更好地检测

血吸虫中这类抗原变异, 我们从位点特异性抗原变异、不同抗原的差异表达的组和抗原修饰三方面进行研究.

(i) 位点特异性抗原变异. 正选择压力下表面蛋白(surface protein)降低宿主的抗原识别能力常发生在宿主-病毒相互作用系统中^[28], 而寄生虫系统中寄生虫基因的抗原位点特异性突变有利于寄生虫对宿主特异的适应性改变^[29]. 为了找到更多引起抗原变异的抗原基因突变位点, 我们使用PAML软件(3.14版本)^[16]中codeml模块的branch-site和M2a模型以及去除配对缺失(gap)的方法对这类表膜蛋白做相应的位点特异性选择突变分析. 结果显示, Sm22.6, Sm20.8和Sm21.7的表膜蛋白亚家族成员具有多个正选择位点(网络版附表3). 这些表膜蛋白中的正选择位点引起的抗原变异可能有助于血吸虫成功逃避或者调节宿主的免疫攻击.

(ii) 表膜抗原蛋白的修饰. 血吸虫抗原的糖基化修饰, 尤其是那些被宿主血细胞抗原的修饰是血吸虫逃避宿主免疫攻击的重要机制之一^[26,27]. 用DictyOGlyc^[30]和YinOYang (1.2版本)软件对吸虫表膜抗原蛋白进行糖基化位点的预测, 同时在Prosit数据库中进行搜索. 结果显示, 这些吸虫表膜蛋白上具有一些特殊的糖基化修饰位点或者模式(pattern)(网络版附表4), 这些糖基化位点或者模式有助于血吸虫掩盖宿主免疫攻击的位点. 此外, 我们还在吸虫表膜蛋白上发现了一些十四烷基化模式(myristoylation pattern), 这些烷基化模式暗示, 吸虫表膜蛋白可能具有其他蛋白类似的功能, 如帮助蛋白定位到胞膜上^[31]及与其他胞膜蛋白(如钙调蛋白^[32,33])发生特殊的蛋白相互作用. 总之, 这类吸虫表膜蛋白的糖基化修饰的结果提示, 它们有利于血吸虫逃避或调节宿主的免疫攻击. 此外, 这些表膜蛋白很可能通过十四烷基化修饰而定位到吸虫表膜上.

(iii) 吸虫表膜蛋白的表达差异分析. 血吸虫中循环抗原有助于血吸虫的宿主免疫逃避^[34], 这些循环抗原包括肠相关抗原(gut-associated antigens)、表膜相关抗原(membrane-associated antigens)和可溶性虫卵抗原(soluble egg antigens)^[35]. 而组成血吸虫表膜抗原的一些蛋白同时也是循环抗原的重要组成部分, 同时这些循环抗原也是血吸虫的不同发育阶段各种抗原的差异表达的不同组合, 因此这种抗原表达模

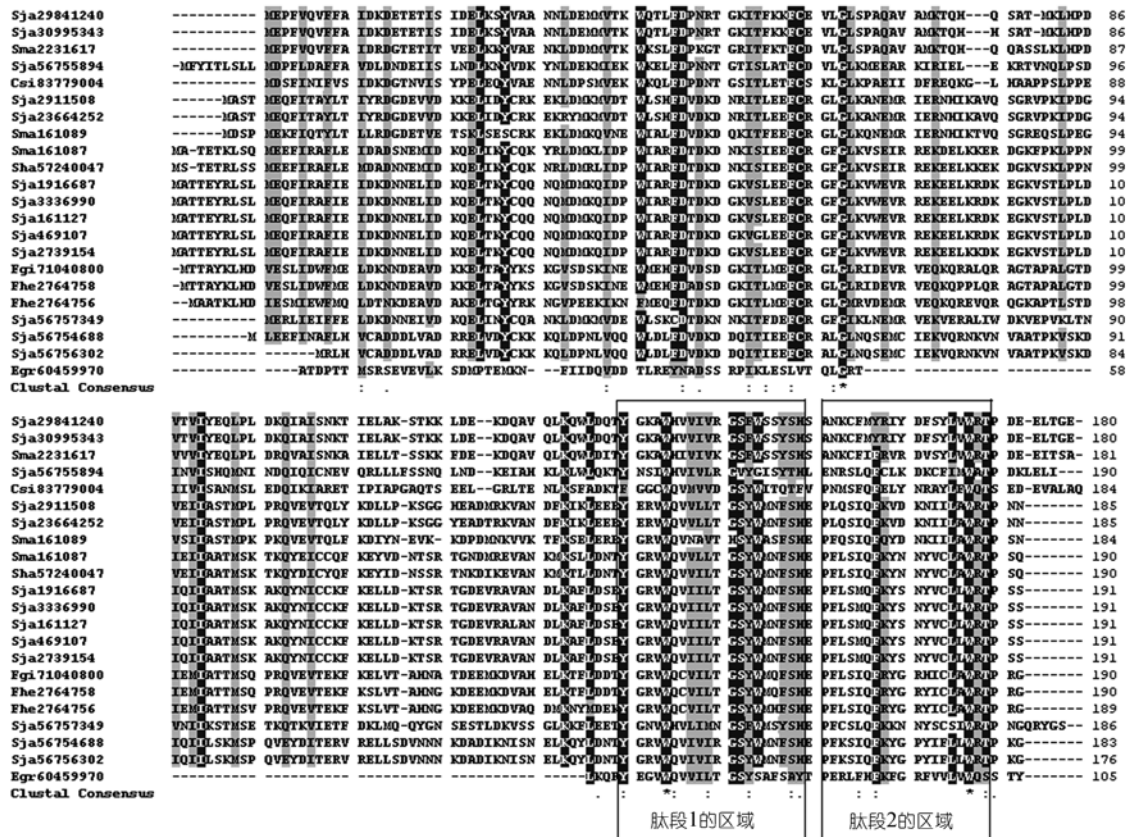


图 2 21 条吸虫表膜蛋白序列和一条细粒棘球绦虫的 EgTeg 蛋白序列(Egr60459970)的多序列比对结果

所有含有 EF-hand 结构域的表膜蛋白跟那些可以调节宿主免疫反应的 EgTeg 蛋白的肽段 1 和肽段 2 具有很高的同源性。加黑部分为完全相同的氨基酸, 灰色为相似性高达 90% 以上的氨基酸。多序列比对结果中, “*”代表完全相同的一组氨基酸; “:”代表有很强相似性的一组氨基酸; “.”代表较弱相似性的一组氨基酸

式变化引起的抗原转变(switching)有助于血吸虫利用类似循环抗原的抗原组合去调节宿主的免疫攻击。为了更好地体现这些含有 EF-hand 结构域的表膜蛋白的表达差异情况, 我们利用日本血吸虫和曼式血吸虫的不同发育阶段中各个抗原基因的 EST 匹配数来定性地表示抗原蛋白的差异表达, 匹配的 EST 筛选标准是小于 10^{-20} 的 E 值以及 90% 以上匹配的同源性。结果发现, 在日本血吸虫中 Sm21.7 亚家族成员的表达明显比其他亚家族成员低很多, 这些表膜蛋白在童虫、成体和雄虫等发育阶段相对表达较高(图 3(a)), 同样在曼式血吸虫中表膜蛋白的表达也出现类似的阶段依赖的表达模式(pattern)(图 3(b)). 这些 EST 匹配结果可以定性说明, 血吸虫表膜蛋白在不同发育阶段呈现各种差异表达模式, 从而在各个发育阶段形成各种表膜抗原蛋白的表达模式的不同组合方式。这种特殊吸虫表膜蛋白抗原的组合方式有助于理解宿主对日本血吸虫和曼式血吸虫产生的

不同程度的免疫反应, 从另一角度也能说明血吸虫对宿主免疫攻击的调节。

3 讨论

结果表明, 这些含有 EF-hand 结构域的血吸虫表膜蛋白可以帮助血吸虫主动地和被动地调节并逃避宿主的免疫攻击。血吸虫可以利用这类表膜蛋白 C 端的肽段调节宿主的巨噬细胞和 T 细胞的活化以及抑制多核细胞趋化来主动地调节并逃避宿主的免疫反应。此外, 血吸虫还通过表膜蛋白的抗原变异、抗原的糖基化修饰和抗原位点特异性选择以及不同抗原表达模式的组合所引起的抗原转变等方式被动调节宿主免疫攻击。

嗜中性粒细胞的随机移动和趋化实验结果显示, 细粒棘球绦虫的 EgTeg 的肽段 1 可以显著地抑制宿主的多核细胞的趋化。免疫印迹实验结果表明, EgTeg 的肽段 2 强烈抑制 IgG4 对 EgTeg 的免疫反应。外周血单核细胞内的干扰素($IFN-\gamma$)和白介素(IL-4)的胞内

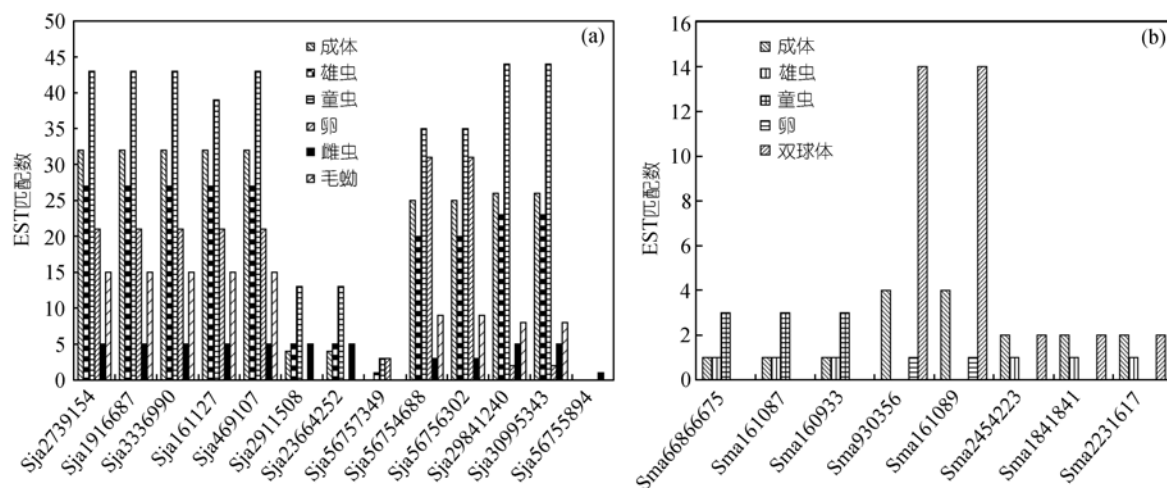


图3 日本血吸虫和曼氏血吸虫在成体、雄虫、雌虫、童虫、卵、毛蚴和双球体(gemeball)不同发育阶段的代表性基因表达差异

(a) 日本血吸虫(*S. japonicum*); (b) 曼氏血吸虫(*S. mansoni*)

免疫荧光实验结果说明, EgTeg可以专一性地诱导分泌白介素(IL-4)的T淋巴细胞的活化^[25]。此外, 根据多序列比对的结果(图2), 这类血吸虫表膜蛋白的C端肽段与细粒棘球绦虫的EgTeg肽段有很高的同源性。这种序列的保守性表明, 血吸虫的表膜蛋白可能有EgTeg肽段的类似功能来抑制宿主细胞趋化, 并诱导Th1依赖的弱的免疫反应来参与宿主免疫逃避。细粒棘球绦虫体内的免疫荧光定位实验说明, EgTeg可以定位到体表膜结构上^[25], 这种寄生虫抗原蛋白的表膜的定位信息暗示具有这类表膜抗原的寄生虫可主动地调节宿主的免疫逃避。这类吸虫表膜蛋白C端两保守肽段的功能暗示这两肽段是可能成为治疗血吸虫病的候选疫苗。

抗原变异引起的宿主免疫监督的逃避对血吸虫在宿主体内的生存起重要作用, 表膜蛋白上的正选择位点可以增强抗原变异的效果, 还可能帮助血吸虫像病毒那样逃避宿主的免疫攻击, 这些不同血吸虫中的特异性抗原变异位点也可能导致物种差异的抗原性。尽管我们不能预测这些用作修饰的糖基是否来源于宿主血细胞表膜抗原^[36], 但吸虫表膜蛋白的糖基化修饰位点和模式的结果仍然可以说明, 它们有助于血吸虫掩盖那些可以被宿主免疫系统识别的攻击位点。此外, 血吸虫不同发育阶段的表膜抗原蛋白表达的不同组合模式引起的基因转变, 有助于我们对日本血吸虫和曼氏血吸虫引起的宿主免疫反应差异的理解, 这种不同抗原组合的模式可能像循环抗

原那样诱导宿主产生弱的免疫反应, 而不是直接的细胞毒杀伤的免疫反应。因此, 抗原的糖基化修饰、位点特异性抗原变异和不同抗原表达差异组合引起基因转变的综合效应可以进一步帮助血吸虫逃避宿主的免疫攻击。

在血吸虫表膜抗原蛋白中, 血吸虫22.6 kD的表膜抗原是主要的表膜抗原之一, 同时也是在血吸虫感染过程中引起宿主IgE免疫反应的主要抗原^[11], 因此它们对于血吸虫在宿主体内的生存发育非常重要。而在本研究系统发育分类分析结果中, Sm22.6亚家族成员的埃及血吸虫基因(Sha57240047)和曼氏血吸虫基因(Sman161087)聚在一起, 而这两物种的基因在序列上与191个氨基酸的日本血吸虫基因相比较有一个氨基酸的缺失。而且位点特异性变异结果表明, Sm22.6亚家族的基因有18个处于正选择压力下的变异位点, 这18个变异位点呈现出在曼氏血吸虫和埃及血吸虫中相同而与日本血吸虫不同的核苷酸的物种差异(网络版附表3和网络版附图)。此外, 由于日本血吸虫相对于曼氏血吸虫和埃及血吸虫而言会引起更严重的血吸虫病, 这18个在Sm22.6亚家族基因中的特异性突变位点可能在不同血吸虫引起的宿主免疫反应差异中起重要的作用。

总之, 这种*in silico*的生物信息学分析可以帮助那些几乎跟其他物种没有同源性的吸虫表膜蛋白获得更多的功能, 还能为进一步确认他们的功能提供直接或者间接的帮助。

参 考 文 献

- Engels D, Chitsulo L, Montresor A, et al. The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Trop*, 2002, 82(2): 139—146[DOI]
- McLaren D J, Hockley D J. Blood flukes have a double outer membrane. *Nature*, 1977, 269(5624): 147—149
- Van Hellemond J J, Retra K, Brouwers J F, et al. Functions of the tegument of schistosomes: Clues from the proteome and lipidome. *Int J Parasitol*, 2006, 36(6): 691—699[DOI]
- Braschi S, Curwen R S, Ashton P D, et al. The tegument surface membranes of the human blood parasite *Schistosoma mansoni*: A proteomic analysis after differential extraction. *Proteomics*, 2006, 6(5): 1471—1482[DOI]
- Perez-Sanchez R, Ramajo-Hernandez A, Ramajo-Martin V, et al. Proteomic analysis of the tegument and excretory-secretory products of adult *Schistosoma bovis* worms. *Proteomics*, 2006, 6(Suppl 1): S226—236[DOI]
- Mohamed M M, Shalaby K A, LoVerde P T, et al. Characterization of Sm20.8, a member of a family of schistosome tegumental antigens. *Mol Biochem Parasitol*, 1998, 96(1-2): 15—25[DOI]
- Taylor D A, Sack J S, Maune J F, et al. Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution. *J Biol Chem*, 1991, 266(32): 21375—21380
- Kuo H J, Tran N T, Clary S A, et al. Characterization of EHD4, an EH domain-containing protein expressed in the extracellular matrix. *J Biol Chem*, 2001, 276(46): 43103—43110[DOI]
- Hafalla J C, Alamares J G 2nd, Acosta L P, et al. Molecular identification of a 21.7 kDa schistosoma japonicum antigen as a target of the human IgE response. *Mol Biochem Parasitol*, 1999, 98(1): 157—161[DOI]
- Santiago M L, Hafalla J C, Kurtis J D, et al. Identification of the *Schistosoma japonicum* 22.6-kDa antigen as a major target of the human IgE response: Similarity of IgE-binding epitopes to allergen peptides. *Int Arch Allergy Immunol*, 1998, 117(2): 94—104[DOI]
- Fitzsimmons C M, Stewart T J, Hoffmann K F, et al. Human IgE response to the *Schistosoma haematobium* 22.6 kDa antigen. *Parasite Immunol*, 2004, 26(8-9): 371—376[DOI]
- Ruiz de Eguino A D, Machin A, Casais R, et al. Cloning and expression in *Escherichia coli* of a *Fasciola hepatica* gene encoding a calcium-binding protein. *Mol Biochem Parasitol*, 1999, 101(1-2): 13—21[DOI]
- Vichasri-Grans S, Subpipattana P, Sobhon P, et al. An analysis of the calcium-binding protein 1 of *Fasciola gigantica* with a comparison to its homologs in the phylum Platyhelminthes. *Mol Biochem Parasitol*, 2006, 146(1): 10—23[DOI]
- Francis P, Bickle Q. Cloning of a 21.7-kDa vaccine-dominant antigen gene of *Schistosoma mansoni* reveals an EF hand-like motif. *Mol Biochem Parasitol*, 1992, 50(2): 215—224[DOI]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*, 2004, 5(2): 150—163[DOI]
- Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 1997, 13(5): 555—556
- Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 2001, 305(3): 567—580[DOI]
- Bendtsen J D, Nielsen H, von Heijne G, et al. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 2004, 340(4): 783—795[DOI]
- Youssef P, Roth J, Frosch M, et al. Expression of myeloid related proteins (MRP) 8 and 14 and the MRP8/14 heterodimer in rheumatoid arthritis synovial membrane. *J Rheumatol*, 1999, 26(12): 2523—2528
- Marti T, Erttmann K D, Gallin M Y. Host-parasite interaction in human onchocerciasis: Identification and sequence analysis of a novel human calgranulin. *Biochem Biophys Res Commun*, 1996, 221(2): 454—458[DOI]
- Kissinger C R, Parge H E, Knighton D R, et al. Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature*, 1995, 378(6557): 641—644[DOI]
- Elshorst B, Hennig M, Forsterling H, et al. NMR solution structure of a complex of calmodulin with a binding peptide of the Ca²⁺ pump. *Biochemistry*, 1999, 38(38): 12320—12332[DOI]
- Liu F, Lu J, Hu W, et al. New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog*, 2006, 2(4): e29[DOI]
- Vassilyev D G, Takeda S, Wakatsuki S, et al. Crystal structure of troponin C in complex with troponin I fragment at 2.3-Å resolution. *Proc Natl Acad Sci USA*, 1998, 95(9): 4847—4852[DOI]
- Ortona E, Margutti P, Delunardo F, et al. Screening of an *Echinococcus granulosus* cDNA library with IgG4 from patients with cystic echinococcosis identifies a new tegumental protein involved in the immune escape. *Clin Exp Immunol*, 2005, 142(3): 528—538
- Velupillai P, dos Reis E A, dos Reis M G, et al. Lewis(x)-containing oligosaccharide attenuates schistosome egg antigen-induced immune depression in human schistosomiasis. *Hum Immunol*, 2000, 61(3): 225—232[DOI]
- Simpson A J. Tegumental proteins of *Schistosoma mansoni*: Complex biomolecules and potent antigens. *Mem Inst Oswaldo Cruz*, 1992, 87(Suppl 4): 11—17
- Tomescu C, Law W K, Kedes D H. Surface downregulation of major histocompatibility complex class I, PE-CAM, and ICAM-1 following de novo infection of endothelial cells with Kaposi's sarcoma-associated herpesvirus. *J Virol*, 2003, 77(17): 9669—9684[DOI]
- Forsberg R, Christiansen F B. A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. *Mol Biol Evol*, 2003, 20(8): 1252—1259[DOI]
- Gupta R, Jung E, Gooley A A, et al. Scanning the available *Dicystostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks. *Glycobiology*, 1999, 9(10): 1009—1022[DOI]
- Maurer-Stroh S, Gouda M, Novatchkova M, et al. MYRbase: Analysis of genome-wide glycine myristoylation enlarges the functional spectrum of eukaryotic myristoylated proteins. *Genome Biol*, 2004, 5(3): R21[DOI]
- Hayashi N, Matsubara M, Jinbo Y, et al. Nef of HIV-1 interacts directly with calcium-bound calmodulin. *Protein Sci*, 2002, 11(3): 529—537[DOI]
- Hayashi N, Izumi Y, Titani K, et al. The binding of myristoylated N-terminal nonapeptide from neuro-specific protein CAP-23/NAP-22 to calmodulin does not induce the globular structure observed for the calmodulin-nonmyristoylated peptide complex. *Protein Sci*, 2000, 9(10): 1905—1913
- Qian Z L, Deelder A M. *Schistosoma japonicum*: Immunological characterization and detection of circulating polysaccharide antigens from adult worms. *Exp Parasitol*, 1983, 55(2): 168—178[DOI]
- van Lieshout L, Polderman A M, Deelder A M. Immunodiagnosis of schistosomiasis by determination of the circulating antigens CAA and CCA, in particular in individuals with recent or light infections. *Acta Trop*, 2000, 77(1): 69—80[DOI]
- Goldring O L, Clegg J A, Smithers S R, et al. Acquisition of human blood group antigens by *Schistosoma mansoni*. *Clin Exp Immunol*, 1976, 26(1): 181—187

补充材料

附表1 收集的吸虫 21 条表膜蛋白的基本信息

id_GI	Trematodeid	Proteins	Species	Reference
gi:161087	Sma161087	MATETKLSQMEEFIRAFLEIDADSNEIDMKQELIKYCQKYRLDMKLIDPWIAFRDT DKDNKISIEEFCRGLKLVSEIRREKDELKKERDGGKFKLPNIEIIAATMSKTKQYE ICCFKEYYVDNTSRTGNDMREVANKMKSLLDNTYGRVWQVVLLTGSYWMNFSHE PFLSIQFKYNNYVCLAWRTPSQ	<i>Schistosoma mansoni</i>	PubMed:1852171
gi:161089	Sma161089	MDSPMEKFIQTYLTLRDGDETVETSKLSESCRKEKLDMDKQVNEWIALFDVDDKQ KITFEFCRGLGLKQNEIRIERNHIKTQSGREQSLPEGVSIIASTMPKPKQVEVTQ LFKDIYNEVKKDPDMNKVVVKTFKSELERRYGRVWQVNAVTHSYWASFSEHPFQSI QFQYDNKILAWRTPSN	<i>Schistosoma mansoni</i>	PubMed:1371327
gi:2231617	Sma2231617	MEPFVQVFAIDRDGTETITVEELKKYVAENKLDMMVTKWKSFLDPKGTGRITF KTFCDLVGLSPAQAVAMKTQHQAASSLKLHPDVVVIYEQLPLDRQVAISNKAIELL TSSKKFDEKDAQVQLKQWLDITYGKAHWIVVKGFSWSSYSHSANKCFIFVRDVS YLVWRTPDEEITSA	<i>Schistosoma mansoni</i>	NA
gi:57240047	Sha57240047	MSTETRLSSMEEFIRAFLEMDADNNEMIDKQELIKYCQKNRLDMRLIDPWIAFRDT DKDNKISIEEFCRGLKLVSEIRREKEELKKEKDGKVSCLPPNVEIIAATMSKTKQY DICYQFKEYIDNSSRTNKDIKEVANKMKTLLDNTYGRVWQVVILTGSYWMNFSHE PFLSIQFKYNNYVCLAWRTPSQ	<i>Schistosoma haematobium</i>	PubMed:15679635
gi:2739154	Sja2739154	MATTEYRLSLMEQFIRAFIEIDKDNNELIDKQELTKYCQQNQMDMKQIDPWIAFRD TDKDGKVSLEEFRCRGLKLVWEVRREKEELKKDKEGKVSTLPLDIQIIAATMSKA KQYNICCKFKELLDKTSRTGDEVRAVANDLKAFLDSEYGRVWQVIILTGSYWMNF SHEPFLSMQFKYSNYVCLLWRTPS	<i>Schistosoma japonicum</i>	PubMed:9784652
gi:1916687	Sja1916687	MATTEYRLSLMEQFIRAFIEIDKDNNELIDKQELTKYCQQNQMDMKQIDPWIAFRD TDKDGKVSLEEFRCRGLKLVWEVRREKEELKRDEGKVSTLPLDIQIIAATMSKAK QYNICCKFKELLDKTSRTGDEVRAVANDLKAFLDSEYGRVWQVIILTGSYWMNFS HEPFLSMQFKYSNYVCLLWRTPS	<i>Schistosoma japonicum</i>	NA
gi:3336990	Sja3336990	MATTEYRLSLMEQFIRAFIEIDKDNNELIDKQELTKYCQQNQMDMKQIDPWIAFRD TDKDGKVSLEEFRCRGLKLVWEVRREKEELKRDEGKVSTLPLDIQIIAATMSKAK QYNICCKFKELLDKTSRTGDEVRAVANDLKAFLDSEYGRVWQVIILTGSYWMNFS HEPFLSMQFKYSNYVCLLWRTPS	<i>Schistosoma japonicum</i>	PubMed:8194761
gi:161127	Sja161127	MATTEYRLSLMEQFIRAFIEIDKDNNELIDKQELTKYCQQNQMDMKQIDPWIAFRD TDKDGKVSLEEFRCRGLKLVWEVRREKEELKRDEGKVSTLPLDIQIIAATMSKAK QYNICCKFKELLDKTSRTGDEVRAVANDLKAFLDSEYGRVWQVIILTGSYWMNFS HEPFLSMQFKYSNYVCLLWRTPS	<i>Schistosoma japonicum</i>	PubMed:1852171
gi:469107	Sja469107	MATTEYRLSLMEQFIRAFIEIDKDNNELIDKQELTKYCQQNQMDMKQIDPWIAFRD TDKDGKVSLEEFRCRGLKLVWEVRREKEELKRDEGKVSTLPLDIQIIAATMSKA KQYNICCKFKELLDKTSRTGDEVRAVANDLKAFLDSEYGRVWQVIILTGSYWMNF SHEPFLSMQFKYSNYVCLLWRTPS	<i>Schistosoma japonicum</i>	NA
gi:2911508	Sja2911508	MASTMEQFITAYLTIYRDGDEVVDKKEIDYCRKEKLDMMKMDVDTWLSHFDVDDK NRITLLEFCRGLGLKANEMRIERNHIKAVQSGRVPKIPDGVEIIASTMPLPRQVEVT QLYKDLLPKSGGHEADMRRKVANDFKIKLEERYERVWQVVLLTGSYWMNFSHEPL QSIQFKVDKNILAWRTPNN	<i>Schistosoma japonicum</i>	PubMed:10029318
gi:23664252	Sja23664252	MASTMEQFITAYLTIYRDGDEVVDKKEIDYCRKEKRYMKMDVDTWLSHFDVDDK NRITLLEFCRGLGLKANEMRIERNHIKAVQSGRVPKIPDGVEIIASTMPLPRQVEVT QLYKDLLPKSGGYEADTRKVANDFKIKLEERYERVWQVVLLTGSYWMNFSHEPLQ SIQFKVDKNILAWRTPNN	<i>Schistosoma japonicum</i>	NA
gi:56757349	Sja56757349	MERLIEIFFELDKDNNEIVDKQELINYCQANKLDMKMDVDEWLSKCDTDKNNKITF DEFRCRGLGLKNEIRVEKVERALIWDKVEPVKLTNNIHKSTMSKTKTKVIETFD KLMQYQGNSESTLDKVSSGLKKFLEETYGNVWHVLMNGSFWMSYSHEPFCSLQ FKKNYSYSCSIWRTPNGQRYGS	<i>Schistosoma japonicum</i>	PubMed:16617374
gi:56754688	Sja56754688	MLEEFINAFHLVCADDLVDARRELVDYCKKKQLDPNLVQQWLDLFDVDDKDDQI TIEEFCRALGLNQSEMCIKQVRNKNVAATPKVSKDIQILSKMSPQVEYDITERV RELLSDVNNKADIKNISNELKQYLDNTYGRVWQVIVIRGSYWMNFSHEPFSIQ FKYGPYIFLLWRTPKG	<i>Schistosoma japonicum</i>	PubMed:16617375
gi:56756302	Sja56756302	MRLHVCADDDLVDARRELVDYCKKKQLDPNLVQQWLDLFDVDDKDDQITIEEFCR ALGLNQSEMCIKQVRNKNVAATPKVSKDIQILSKMSPQVEYDITERVRELLSDV NNNKDADIKNISNELKQYLDNTYGRVWQVIVIRGSYWMNFSHEPFSIQFKYGPYI FLLWRTPKG	<i>Schistosoma japonicum</i>	PubMed:16617376
gi:56755894	Sja56755894	MFYITLSLLMDPFLDAFFAVDLNDNEIISLNDLKNYVDKYNLDEKMIKWKELFDP NNTGTISLATFCDVLGLKMEEARIRIELEKRTVNQLPSDINVISHQMNINDIQICN EVQRLLLFSSNQLNDKEIAHKLKLWLQKTYNSIWHVIVLRGVYGISYTHLENRSQ FCLKDKCFIMWATPDKLELI	<i>Schistosoma japonicum</i>	PubMed:16617377

续表

id_GI	TrematodeId	Proteins	Species	Reference
gi:29841240	Sja29841240	MEPFVQVFFAIDKDETTISIDELKSYVAANNLDEMVTWKWQTLFDPNRTGKITFK KFCEVLGLSPAQAVAMKTQHQSATMKLHPDVTVIYEQLPLDKQIAISNKTIELAKST KKLDEKDQAVQLKQWLDQTYGKAHWVIVIRGSFWSSYSHSANKCFMYRIYDFS LVWRTPEELTGE	<i>Schistosoma japonicum</i>	PubMed:12973349
gi:30995343	Sja30995343	MEPFVQVFFAIDKDETTISIDELKSYVAANNLDEMVTWKWQTLFDPNRTGKITFK KFCEVLGLSPAQAVAMKTQHHSATMKLHPDVTVIYEQLPLDKQIAISNKTIELAKST KKLDEKDQAVQLKQWLDQTYGKAHWVIVIRGSFWSSYSHSANKCFMYRIYDFS LVWRTPEELTGE	<i>Schistosoma japonicum</i>	NA
gi:71040800	Fgi71040800	MTTAYKLHDVESLIDWFMELDKNNDEAVDKKELTAYYKSKGVSDSKINEWMEHF DVSDDGKITLMEFCRGLGLRIDEVRVEQKQRALQRAGTAPALGTDIEMIATTMSQP RQVEVTEKFKELVTAHNATDEEMKDVAHELKTFLDDTYGRVWQCIVLTGSYWMQ FSHEPFLSIQFRYGRHICLAWRTPRG	<i>Fasciola gigantica</i>	PubMed:16297461
gi:2764758	Fhe2764758	MTTAYKLHDVESLIDWFMELDKNNDEAVDKKELTAYYKSKGVSDSKINEWMEHF DADSDGKITLMEFCRGLGLRIDEVRVEQKQPPLQRAGTAPALGTDIEMIATTMSQP RQVEVTEKFKSLVTAHNGKDEEMKDVAHELKTFLDDTYGRVWQCIVLTGSYWMQ FSHEPFLSIQFRYGRYICLAWRTPRG	<i>Fasciola hepatica</i>	PubMed:10413039
gi:2764756	Fhe2764756	MAATKLHDIESMIEWFMQLDTNKAEDAKELTGYRKNGVPEEKIKNFMEQFDT DKDGGKITLMEFCRGLGMRVDEMRVEQKQREVQRQKAPTLDIEMIATTMSVPR QVEVTEKFKSLVTAHNGKDEEMKDVAQDMKNYMDEKYGRVWQCIVLTGSYWM HFSHEPFLSIQFRYGRYICLAWRTPRG	<i>Fasciola hepatica</i>	NA
gi:83779004	Csi83779004	MDSFINIFVSIDKDGTVISYPELEQYVAENNLDPMSVEKWKQLFDPDNTGSITLET FCSKLGLKPAEIHDFREQGLHAAPSLPPEIIVISANMSLEDQIKIARETIPIAPGAQT SEELGRLTENLKSFADKTFGGCWQVMVVDGSYWITQTFVPNMSQFELYNRAYLF WQTSEDEVALAQ	<i>Clonorchis sinensis</i>	NA

id, 序列的 GI 号码; TrematodeId, 吸虫表膜蛋白对应的 id; proteins, 吸虫基因对应的蛋白序列; species, 吸虫基因对应的物种; reference, 释放吸虫基因对应的文献. 基因名字是物种的缩写后缀 NCBI 中该基因的 GI 号码, 如基因 Sja2911508 由日本血吸虫的缩写 Sja 和 GI 号码(2911508)组成

附表 2 吸虫 21 条表膜蛋白的蛋白-蛋白相互作用结果

trematodeId	dipHomologueId	dipInteractionId	references	functions
Sja2911508-Sja23664252-Fgi71040800-Fh e2764758	DIP:1165N	DIP:1166N	(PMID: 10606357)	indicating host-parasite interaction in human
Sja2911508-Sja23664252-Sja56754688-Sja 56756302-Sja56755894-Fhe2764756	DIP:1166N	DIP:1165N	(PMID: 8619876) (PMID: 10606357)	onchocerciasis by a novel human calgranulin Altered activation and differentiation of lining layer macrophages at the site of maximal cartilage destruction in rheuma- toid arthritis (RA)
Sma161087-Sma2231617-Sha57240047-Sja 2739154-Sja1916687-Sja3336990-Sja1611 27-Sja469107-Sja2911508-Sja23664252-Sj a56757349-Sja56754688-Sja56756302-Sja2 9841240-Sja30995343-Fhe2764756	DIP:6096N	DIP:6095N	(PMID: 8524402)	To regulate T-cell activation (as Cal- cineurin)
Fgi71040800-Fhe2764758-Fhe2764756	DIP:6059N	DIP:6060N-DIP:6059N	(PMID: 9560191)	To regulate of muscle contraction
Sma161087-Sma161089-Sma2231617-Sha5 7240047-Sja2739154-Sja1916687-Sja3336 990-Sja161127-Sja469107-Sja2911508-Sja 56757349-Fgi71040800-Fhe2764758-Fhe2 764756-Csi83779004	DIP:538N	DIP:6128N	(PMID: 10493800)	To interact with Ca ²⁺ -transporting AT- Pase 3, plasma membrane, control the Ca ²⁺ level in the cell by Ca ²⁺ pump
Sma161087-Sma161089-Sma2231617-Sha5 7240047-Sja2739154-Sja1916687-Sja3336 990-Sja161127-Sja469107-Sja2911508-Sja 56757349-Fgi71040800-Fhe2764758-Fhe2 764756-Csi83779004	DIP:538N	DIP:383N	(PMID: 9009191)	The activity of NMDA (N-methyl-D- as- partate) receptors which neurotransmitter receptors suggesting a functionally impor- tant interaction between NMDA receptors and the postsynaptic cytoskeleton.
Sma161087-Sma161089-Sma2231617-Sha5 7240047-Sja2739154-Sja1916687-Sja3336 990-Sja161127-Sja469107-Sja2911508-Sja 56757349-Fgi71040800-Fhe2764758-Fhe2 764756-Csi83779004	DIP:538N	DIP:452N	(PMID: 8872303)	To contribute to the growth and plasticity of neuronal connections
Sma2231617	DIP:27560N	DIP:27522N-DIP:27526 N-DIP:27527N-DIP:275 30N-DIP:27528N-DIP:2 7520N-DIP:27521N-DI P:27523N-DIP:27529N	(PMID: 14743216)	To mediate and regulate TNF-alpha/ NF-kappa B signal transduction pathway.

dipHomologueId, DIP 数据库中对应的吸虫表膜蛋白的同源蛋白 id; dipInteractionId, DIP 数据库中与 dipHomologueId 蛋白相互作用的蛋白 id; references, 涉及该相互作用蛋白对的文献 id(对应的 PubMed id); functions, 对应的相互作用可能的功能; 多条 trematodeId 的蛋白 id 用“-”符号连接

附表 3 吸虫表膜蛋白上处于正选择压力下的位点

subfamily	Sm22.6	Sm22.6	Sm21.7	Sm21.7	Sm20.8	Sm20.8	Sm20.8	Sm20.8
trematodeId	Sja2739154-Sja1916687-Sja3336990-Sja161127-Sja469107	Sha57240047-Sma161087	Sja2911508-Sja23664252	Sma161089	Sma2231617	Sja28317707-Sja30995342	Sja56755893	Csi83779003
sites	6	5	2	2	101	100	101	102
	23	22	10	10	132	131	133	134
	35	34	15	15				
	40	39	16	16				
	42	41	22	22				
	47	46	26	26				
	61	60	31	31				
	63	62	41	41				
	77	76	44	44				
	99	98	46	46				
	123	122	47	47				
	136	135	55	55				
	140	139	56	56				
	141	140	70	70				
	147	146	90	89				
	148	147	96	95				
	185	184	104	103				
	191	190	118	117				
			119	118				
			121	120				
			122	121				
			129	128				
			134	133				
			141	140				
			150	149				
			151	150				
			154	153				
			158	157				
			172	171				

附表 4 吸虫表膜蛋白用 DictyOGlyc 和 YinOYang 软件预测的糖基化修饰位点以及对 Prosite 数据库搜索的糖基化模式

software Prediction			Prosite			
trematodeId	DictyOGlyc	YinOYang	glycosylationId	positionPep	myristoylationId	myristoylationPositionPep
Sma161087	187(T)-189(S)	187(T)-189(S)	PS00001	125-128:(NTSR) 164-167:(NFSH)	PS00008	72-77:(GLkvSE)
Sma161089	183(S)	22(T)-25(T)-96(S)-152(T)-154(S)-181(T)-183(S)	NA	NA	PS00008	67-72:(GLkqNE)
Sma2231617	65(S)	39(T)-74(T)-146(S)-150(S)-179(T)-180(S)	NA	NA	PS00008	63-68:(GLspAQ) 145-150:(GSfwSS)
Sha57240047	187(T)-189(S)	2(S)-3(T)-187(T)-189(S)	PS00001	125-128:(NSSR) 164-167:(NFSH)	PS00008	72-77:(GLkvSE)
Sja2739154	190(S)	4(T)-188(T)-190(S)-191(S)	PS00001	165-168:(NFSH)	NA	NA
Sja1916687	190(S)	4(T)-188(T)-190(S)-191(S)	PS00001	165-168:(NFSH)	NA	NA
Sja3336990	190(S)	4(T)-188(T)-190(S)-191(S)	PS00001	165-168:(NFSH)	NA	NA
Sja161127	190(S)	4(T)-188(T)-190(S)-191(S)	PS00001	165-168:(NFSH)	NA	NA
Sja469107	190(S)	4(T)-188(T)-190(S)-191(S)	PS00001	165-168:(NFSH)	NA	NA
Sja2911508	182(T)	4(T)-85(S)-111(T)-121(S)-182(T)	PS00001	159-162:(NFSH)	PS00008	67-72:(GLkaNE) 122-127:(GGheAD)
Sja23664252	182(T)	4(T)-85(S)-111(T)-121(S)-182(T)	PS00001	159-162:(NFSH)	PS00008	67-72:(GLkaNE) 122-127:(GGyeAD)
Sja56757349	NA	128(S)-186(S)	PS00001	149-152:(NGSF) 170-173:(NYSC)	PS00008	63-68:(GiklNE) 118-123:(GNseST) 181-186:(GQryGS)
Sja56754688	180(T)	85(T)-89(S)	PS00001	66-69:(NQSE) 128-131:(NISN) 157-160:(NFSH)	PS00008	64-69:(GLnqSE)

续表

software Prediction		Prosite			
Sja56756302 173(T)	78(T)-82(S)	PS00001	59-62:(NQSE) 121-124:(NISN) 150-153:(NFSH)	PS00008	57-62:(GLnqSE)
Sja56755894 NA	59(T)-95(S)	PS00001	57-60:(NNTG) 166-169:(NRSL)	PS00008	158-163:(GI syTH)
Sja29841240 65(S)	74(T)-88(T)-149(S)-178(T)	PS00001	48-51:(NRTG) 104-107:(NKTI)	PS00008	63-68:(GLspAQ) 144-149:(GSfwSS)
Sja30995343 65(S)	74(T)-88(T)-149(S)-178(T)	PS00001	48-51:(NRTG) 104-107:(NKTI)	PS00008	63-68:(GLspAQ) 144-149:(GSfwSS)
Fgi71040800 NA	2(T)-39(S)-108(S)-187(T)	PS00001	127-130:(NATD)	PS00008	41-46:(GVsdSK) 91-96:(GTapAL)
Fhe2764758 NA	2(T)-39(S)-108(S)-116(T)-187(T)	NA	NA	PS00008	41-46:(GVsdSK) 91-96:(GTapAL)
Fhe2764756 NA	97(T)-115(T)-186(T)	NA	NA	NA	NA
Csi83779004 36(S)	50(T)-84(S)-117(T)-176(S)	PS00001	95-98:(NMSL) 158-161:(NMSF)	PS00008	63-68:(GLkpAE) 114-119:(GAqtSE)

DictyOGlyc, 使用 DictyOGlyc 软件预测的表膜蛋白糖基化修饰位点和对应的氨基酸; YinOYang, 使用 YinOYang 软件预测的表膜蛋白糖基化修饰位点和对应的氨基酸; glycosylationId, Prosite 数据库中对应的糖基化修饰模式 id; postionPep, 表膜蛋白的糖基化修饰模式对应的位置和模式; myristoylationId, Prosite 数据库中对应的十四烷基化模式 id; myristoylationPositionPep, 十四烷基化模式对应的位置和肽段

Sja2911508	-----KAST	HEQFITAYLT	TYRGGDEVVD	KKEIIDCRK	EKLDQGVDT	LSKTFVSKD	NRITLLEECR	GLSDQNEMR	IERNHIAVQ	SGRVPKIPDG	94
Sja23664252	-----KAST	HEQFITAYLT	TYRGGDEVVD	KKEIIDCRK	EKKYQGVDT	LSKTFVSKD	NRITLLEECR	GLSDQNEMR	IERNHIAVQ	SGRVPKIPDG	94
Sma161089	-----KASP	HEKFIQTLYT	LLRGGDETVV	TSKSESCKR	EKLDMQVNE	IALTFVSKD	QKITFEETCR	GLSDQNEMR	IERNHIAVQ	SGREQSLPEG	94
Sma161087	MA-TETKLSQ	HEEFIRAELE	IDADSNKICD	KQELINCCQ	YRLDMQLIDP	VIARFTBKO	NKISIEECR	GFGLKVEIR	REKEELKCR	DGKFPKLPN	99
Sha57240047	M-STETRLSS	HEEFIRAELE	IDADSNKICD	KQELINCCQ	NRLDMQLIDP	VIARFTBKO	NKISIEECR	GFGLKVEIR	REKEELKCR	DGKVPKLPN	99
Sja2739154	MATTEYRLSL	HEQFIRAFIE	IDKNNELID	KQELTKCQO	NQMDMQIDP	VIARFTBKO	CKVSLLEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Sja3336990	MATTEYRLSL	HEQFIRAFIE	IDKNNELID	KQELTKCQO	NQMDMQIDP	VIARFTBKO	CKVSLLEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Sja1916687	MATTEYRLSL	HEQFIRAFIE	IDKNNELID	KQELTKCQO	NQMDMQIDP	VIARFTBKO	CKVSLLEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Sja161127	MATTEYRLSL	HEQFIRAFIE	IDKNNELID	KQELTKCQO	NQMDMQIDP	VIARFTBKO	CKVSLLEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Sja469107	MATTEYRLSL	HEQFIRAFIE	IDKNNELID	KQELTKCQO	NQMDMQIDP	VIARFTBKO	CKVSLLEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Fgi71040800	M-TTAYLKMD	VESLIWEME	LDKNDEAVD	KCEITATYS	KGVSDSKINE	VDGKTFVSD	CKITLMECR	GLGLRIDEVR	VEQKQPPLO	AGTAPALGT	99
Fhe2764758	M-TTAYLKMD	VESLIWEME	LDKNDEAVD	KCEITATYS	KGVSDSKINE	VDGKTFVSD	CKITLMECR	GLGLRIDEVR	VEQKQPPLO	AGTAPALGT	99
Fhe2764756	--MARTKLM	IESLICWEMQ	LDKNDEAVD	AKELTGYRK	NGVPEEKIKN	VDGKTFVSD	CKITLMECR	GLGLRIDEVR	VEQKQPPLO	AGTAPALGT	99
Sja56757349	-----	HEEFIRAELE	IDKNNELID	KQELINCCQ	NQMDMQIDP	VIARFTBKO	NKISIEECR	GFGLKVEVR	REKEELKCR	EGKVSTPLD	100
Sja56754688	-----M	LEEFIAEELH	VCADBDLVAD	REELVDCCK	KQLDPNLVQO	VLBLTFVSKD	DOITIEECR	ALNDQSEMC	IEKVORNVKN	VAMPKVSCK	91
Sja56756302	-----MRLH	VCADBDLVAD	REELVDCCK	KQLDPNLVQO	VLBLTFVSKD	DOITIEECR	ALNDQSEMC	IEKVORNVKN	VAMPKVSCK	91	
Sma2231617	-----	HEPFPVVEFA	IDRGTETIT	VEELDKVRE	NKLDMDQVTK	NKSLTFPKGT	GRITETECR	VLGLSPAQAV	RIKTHQMSAT	S---LKLMD	87
Sja29841240	-----	HEPFPVVEFA	IDRGTETIT	VEELDKVRE	NKLDMDQVTK	NKSLTFPKGT	GRITETECR	VLGLSPAQAV	RIKTHQMSAT	S---LKLMD	87
Sja30995343	-----	HEPFPVVEFA	IDRGTETIT	VEELDKVRE	NKLDMDQVTK	NKSLTFPKGT	GRITETECR	VLGLSPAQAV	RIKTHQMSAT	S---LKLMD	87
Sja56755894	---MFYITLSL	HEPFLDAEFA	VDLNDDEIS	LDNDQNVVK	YNLDEKMKIE	NKSLTFPKGT	GRITETECR	VLGLSPAQAV	RIKTHQMSAT	VN---QLSDP	86
Csi83779004	-----	NBSFINIEVS	IDKNGTNIYS	YPELQGVRE	NKLDMDQVTK	NKSLTFPKGT	GRITETECR	VLGLSPAQAV	RIKTHQMSAT	AP---PSLPPE	88
Clustal Consensus		::	::	::	::	::	::	::	::		
Sja2911508	VEIETASTMPL	PROVEVTOLY	KDILLP-KSGG	YEADTRKVN	DFEIKLEERY	ERVQVLLT	GSYDNFSHE	PLQSIQKVD	KNILAMRT	NN-----	185
Sja23664252	VEIETASTMPL	PROVEVTOLY	KDILLP-KSGG	YEADTRKVN	DFEIKLEERY	ERVQVLLT	GSYDNFSHE	PLQSIQKVD	KNILAMRT	NN-----	185
Sma161089	VSIEIATHPK	PKVEVTOLF	KDIYN-EVK-	KDPDQNVVK	TFSELEERY	GRVQVNAV	MSYDASFHE	PFQSIQKVD	KNILAMRT	SN-----	184
Sma161087	IEIETATHSK	TKVEYCCQF	KEYID-NSSR	TGNDKEVAN	KQSLDNTY	GRVQVLLT	GSYDNFSHE	PFQSIQKYN	NYVCLART	SQ-----	190
Sha57240047	VEIETATHSK	TKVEYCCQF	KEYID-NSSR	TGNDKEVAN	KQSLDNTY	GRVQVLLT	GSYDNFSHE	PFQSIQKYN	NYVCLART	SQ-----	190
Sja2739154	IQIETATHSK	AKYINCCCK	KELLD-KTSR	TGDEVRVAN	DLAFDSEY	GRVQVILL	GSYDNFSHE	PFQSIQKYS	NYVCLART	SS-----	191
Sja3336990	IQIETATHSK	AKYINCCCK	KELLD-KTSR	TGDEVRVAN	DLAFDSEY	GRVQVILL	GSYDNFSHE	PFQSIQKYS	NYVCLART	SS-----	191
Sja1916687	IQIETATHSK	AKYINCCCK	KELLD-KTSR	TGDEVRVAN	DLAFDSEY	GRVQVILL	GSYDNFSHE	PFQSIQKYS	NYVCLART	SS-----	191
Sja161127	IQIETATHSK	AKYINCCCK	KELLD-KTSR	TGDEVRVAN	DLAFDSEY	GRVQVILL	GSYDNFSHE	PFQSIQKYS	NYVCLART	SS-----	191
Sja469107	IQIETATHSK	AKYINCCCK	KELLD-KTSR	TGDEVRVAN	DLAFDSEY	GRVQVILL	GSYDNFSHE	PFQSIQKYS	NYVCLART	SS-----	191
Fgi71040800	IEIETATHSQ	PROVEVTEK	KSLVT-AMNG	KDEEKQVAV	ELATFDDT	GRVQCVILL	GSYDNFSHE	PFQSIQKYG	RYICLART	RG-----	190
Fhe2764758	IEIETATHSQ	PROVEVTEK	KSLVT-AMNG	KDEEKQVAV	ELATFDDT	GRVQCVILL	GSYDNFSHE	PFQSIQKYG	RYICLART	RG-----	190
Fhe2764756	IEIETATHSQ	PROVEVTEK	KSLVT-AMNG	KDEEKQVAV	ELATFDDT	GRVQCVILL	GSYDNFSHE	PFQSIQKYG	RYICLART	RG-----	190
Sja56757349	VNIETATHSE	TKVEYVTEF	KDLMQ-QYGN	SESLDKVSS	GLKFLDETY	GNDVILLDN	GSYDNFSHE	PFQSIQKYN	NYVCLART	NGQYGS-	186
Sja56754688	IQIETATHSK	QVEYDITERV	RELLSDVNNN	KDADIKQIN	ELQYDNTY	GRVQVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	KG-----	183
Sja56756302	IQIETATHSK	QVEYDITERV	RELLSDVNNN	KDADIKQIN	ELQYDNTY	GRVQVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	KG-----	176
Sma2231617	VVVYEQQLPL	DEIVASINKA	IELLT---SS	KKLEKQDQV	QLQWDDTY	GKGDVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	DELTGE-	180
Sja29841240	VVVYEQQLPL	DEIVASINKA	IELLT---SS	KKLEKQDQV	QLQWDDTY	GKGDVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	DELTGE-	180
Sja30995343	VVVYEQQLPL	DEIVASINKA	IELLT---SS	KKLEKQDQV	QLQWDDTY	GKGDVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	DELTGE-	180
Sja56755894	INVSQNDICI	NQDQCNVEV	QRELL--FSS	NQNDQKQIN	KLQWDDTY	GKGDVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	DKLELI--	190
Csi83779004	IEIETATHSK	QVEYDITERV	RELLSDVNNN	KDADIKQIN	ELQYDNTY	GRVQVIVR	GSYDNFSHE	PFQSIQKYG	PYIFLART	DELTGE-	180
Clustal Consensus	::*::	::	::	::	::	::	::	::	::	::	34

附图 吸虫 21 条表膜蛋白序列的多序列比对结果

*, 完全相同的一组氨基酸; :, 有很强相似性的一组氨基酸; ·, 较弱相似性的一组氨基酸