

中图分类号: R363

单位代码: 13705

UDC: 616-095

密 级: 公开



成都医学院
CHENGDU MEDICAL COLLEGE

硕士学位论文

纤细裸藻剪接系统及转化方法研究

学科专业: 病原生物学

研究方向: Pre-mRNA 剪接

学位类型: 学术学位

指导教师: 孙成副 副研究员

学生姓名: 高萍蔚

学 号: 2133010008

二〇二四年五月

Chengdu Medical College

Master Degree Thesis

Research on the splicing system and transformation
methods in *Euglena gracilis*

Major: Medical Microbiology

Research field: Splicing of pre-mRNA

Degree type: Academic degree

Adviser: Chengfu Sun

Name of Postgraduate: Pingwei Gao

Postgraduate Number: 2133010008

May, 2024

基金资助

本论文在以下基金资助下完成：

- 1.国家自然科学基金项目（批准号：31872717）
- 2.成都医学院研究生科研创新基金项目（项目编号：YCX2023-01-06）

目 录

前言	1
第一章 纤细裸藻剪接体组分鉴定	3
1 目的	3
2 材料与方法	3
2.1 细胞株	3
2.2 主要试剂	3
2.3 主要仪器及耗材	4
2.4 本实验相关引物	4
2.5 常规实验试剂配置	7
2.6 实验方法	8
3 结果与分析	10
3.1 纤细裸藻剪接体蛋白基因的生信分析	10
3.2 纤细裸藻 U2 与 U6 snRNA 的分析	13
3.3 纤细裸藻剪接体组分的生化鉴定	14
3.4 SL 介导的反式剪接在纤细裸藻中的普遍性	15
3.5 纤细裸藻与其它物种剪接体蛋白质组的比较	15
3.6 纤细裸藻对剪接抑制剂不敏感的分子基础	16
4 结论	19
5 讨论	20
第二章 纤细裸藻核 pre-mRNA 内含子的鉴定与分析	21
1 目的	21
2 材料与方法	21
2.1 细胞株	21
2.2 实验方法	21
3 结果与分析	22
3.1 纤细裸藻细胞核 pre-mRNA 内含子的鉴定	22
3.2 细胞核 pre-mRNA 常规内含子 5'ss 区域的特征分析	25
3.3 Extended U6/5'ss 螺旋的形成	27
3.4 细胞核 pre-mRNA 常规内含子 3'ss 区域的特征分析	28
3.5 细胞核 pre-mRNA 非常规内含子末端碱基配对	30
3.6 纤细裸藻细胞核 pre-mRNA 内含子的选择性剪接	32
3.7 纤细裸藻细胞核 pre-mRNA 内含子串联重复序列	33
4 结论	35
5 讨论	35

第三章 纤细裸藻 Dbr1 蛋白抑制剂的筛选及功能验证	37
1 目的	37
2 材料与方法	37
2.1 细胞株	37
2.2 主要试剂	37
2.3 主要仪器及耗材	38
2.4 常规实验试剂配置	39
2.5 本实验相关引物序列	44
2.6 实验方法	45
3 结果与分析	49
3.1 纤细裸藻 Dbr1 基因的确定	49
3.2 纤细裸藻 Dbr1 蛋白的纯化	50
3.3 纤细裸藻 Dbr1 蛋白结构预测	51
3.4 小分子化合物的筛选	52
3.5 小分子化合物的体内活性检测	53
4 结论	54
5 讨论	54
第四章 纤细裸藻生物学转化方法的构建	55
1 目的	55
2 材料与方法	55
2.1 细胞株	55
2.2 主要试剂	55
2.3 主要仪器及耗材	56
2.4 常规实验试剂配置	56
2.5 本实验相关引物序列	56
2.6 实验方法	57
3 结果与分析	60
3.1 Pep-1-GFP 融合蛋白及相关蛋白的纯化	60
3.2 PGH 蛋白的细胞内转化	61
3.3 Pep-1 的位置转化效率的影响	63
3.4 二甲亚砜(DMSO)介导的细胞内传导	63
4 结论	65
5 讨论	65
参考文献	68
文献综述 反式剪接研究进展	72
参考文献	78
附录	81

摘 要

研究背景与目的:

早期分化的病原生物,如锥虫、线虫等,均具有传染性,对人体的健康造成了巨大伤害。目前为止对这些生物的剪接生化研究仅限于 SL 介导的反式剪接。顺式剪接的研究大部分是以人类或者酵母为模型,相比之下其他物种则非常之少。因此,我们选取了与锥虫同属一界的纤细裸藻 (*Euglena gracilis*) 为研究对象。

纤细裸藻是一种单细胞真核生物,又称眼虫,体内含有叶绿体,并利用其独特的鞭毛来运动。纤细裸藻的基因组大小与人类基因组大小相当,由于纤细裸藻的基因组中存在较多的重复序列,现在完整的基因组序列还未测序完成。与后鞭毛界的人和酵母不同,纤细裸藻属于古虫界,其剪接类型更具有多样化。在纤细裸藻基因组测序中只发现了少数基因及其内含子,大多数还是属于未知状态。

我们首先利用生信分析鉴定纤细裸藻剪接体基因以及内含子序列,后续计划建立可行的遗传操作方法,为纤细裸藻体内剪接体的纯化奠定基础。以纤细裸藻为模型的研究可以更加深化我们对剪接体的理解,拓展我们对剪接方式的认知。

材料与方法:

1.从纤细裸藻总 RNA 中富集带 polyA 的 RNA 并进行建库。随后进行三代全长转录组以及二代无参转录组测序,结合现有 SRA 数据库中的纤细裸藻转录组测序数据,分别通过 BLASTX 与人剪接体蛋白进行比对。将三个 RNA-seq 结果所得的剪接基因进行整合,并且利用数据库中纤细裸藻基因组数据,获得了纤细裸藻剪接体组分基因序列。最后对鉴定出来的纤细裸藻剪接体组分进行 PCR 验证。此外,我们还利用 BLAST 预测出与纤细裸藻同界的锥虫、利什曼原虫等物种的剪接基因。最后,我们将多种已知的人类剪接抑制剂作用于纤细裸藻细胞,检测纤细裸藻对剪接抑制剂的敏感性。

2.我们利用三代转录组测序数据,以及从 SRA 数据库中获得的转录组数据,共同用于纤细裸藻内含子的鉴定分析。首先,使用 BLAST 程序将叶绿体、线粒体、细胞核 rDNA 这些序列都从 RNA-seq 数据中删除;然后使用 cd-hit 消除冗余序列,从而分离出细胞核 mRNA 序列。使用 exonerate 将 mRNA 序列与基因组进

行对比；最后，对提取到的内含子进行手动筛选并分类。为了鉴定纤细裸藻的 intron 序列，我们从细胞核 mRNA 中提取含有 SL 外显子的序列，并与基因组序列进行 BLAST 比对。我们使用 Seqkit 对提取出的序列进行分析。

3.根据前期工作所预测的纤细裸藻 Dbr1 基因序列，通过 RT-PCR、RACE 等方法获得 Dbr1 全长编码序列。通过 Alphafold 和 Robetta 等 AI 程序预测获得 Dbr1 三维结构。根据 PDB 蛋白质数据库中溶组织内阿米巴虫 Dbr1 (PDB:4PEH) 晶体结构中 Dbr1 蛋白与核酸结合位点确定 Dbr1 上可能的小分子结合区域。利用 AutoDock Vina 软件从 TargetMol 天然产物库中筛选与 Dbr1 亲和力较高的小分子。接下来分别通过蛋白表达体外检测 Dbr1 与小分子的结合能力；以及通过细胞核前体 mRNA 内含子的剪接情况体内检测小分子抑制 Dbr1 的能力。

4.采用无缝克隆方法，基于细菌表达质粒 pET21-EGFP，构建含有跨膜肽标签 Pep-1 序列的相关质粒 PGH (Pep-1-GFP-His)、HPG (His-Pep-1-GFP)、HGP (His-GFP-Pep-1) 和 GPH (GFP-Pep-1-His)。通过化学方法转化细菌，并用 IPTG 诱导，通过 SDS-PAGE 和考马斯亮蓝染色观察蛋白诱导情况。采用 Ni-NTA 亲和层析柱纯化获得相应的融合表达蛋白。含有跨膜肽的 GFP 融合蛋白分别透析至 PBS 中，然后与纤细裸藻细胞在 37℃ 下共同培养，并使用 BX63 荧光显微镜进行观察。另外，我们使用二甲亚砜(DMSO)处理纤细裸藻细胞后，培养基中分别加入 GFP 蛋白、带有荧光标记的 PCR 扩增的 DNA 片段，然后同样在荧光显微镜下进行观察。我们对质粒分子进行了转化观察，通过 PCR 方法检查细胞中是否转入相应的质粒。

结果:

1.生信分析鉴定出 166 个剪接蛋白编码基因和 U2、U6 snRNA 基因。经过与人类剪接体进行对比，我们发现纤细裸藻中缺失的蛋白大多数是来自 SR、hnRNP、B*/C、EJC/TREX 复合物以及 snRNP 蛋白。其余的剪接蛋白相对比较保守。我们对鉴定出的部分剪接基因进行了 PCR 验证。29 个剪接蛋白编码基因以及 U2 和 U6 snRNA 的 PCR 结果表明预测出的基因序列是正确的。这进一步提示我们所鉴定出来的纤细裸藻剪接基因存在且序列正确。

通过对纤细裸藻 166 个剪接蛋白编码基因和 212 条序列分析得知, 含有 SL 外显子的剪接基因概率为 91%, 进一步说明了 SL 介导的反式剪接在纤细裸藻剪接基因中存在的普遍性。对古虫界生物(纤细裸藻、锥虫等)剪接体蛋白进行比较, 我们发现纤细裸藻与布氏锥虫、杜氏利什曼原虫之间剪接蛋白组成存在明显差异。利用人类剪接抑制剂, 直接作用于纤细裸藻细胞, 我们发现纤细裸藻细胞对这些剪接抑制剂不敏感。

2. 我们通过 *exonerate* 程序共鉴定出 28337 个细胞核前体 mRNA 内含子, 包括 11527 个 GT-AG 型内含子, 659 个 GC-AG 型内含子, 11653 个 ss-variant 型内含子, 4498 个 ss-invariant 型内含子。通过鉴定纤细裸藻中的大量内含子, 包括顺式剪接的常规内含子和非常规内含子, 以及 SL 介导的反式剪接。我们发现纤细裸藻的常规内含子和 *outtron* 表现出独特的富含胞嘧啶核苷(C)的多聚胞苷序列和常规的 CAG-3'。此外, 常规内含子具有 5'-GTGTG 共同序列, 这与 *spliced leader RNA (SL-RNA)* 中的 5'剪接位点(5'ss)序列(GTATA)不同。非常规内含子的两末端序列能够形成一个互补配对结构。此外, 我们还描述了另一种剪接模式, 称为重叠的新型剪接形式。

3. 我们利用 RACE 方法以及重叠 PCR 得到了完整的纤细裸藻 Dbr1 序列。*AlphaFold* 和 *Robetta* 程序预测的纤细裸藻 Dbr1 三维蛋白结构基本一致。纤细裸藻 Dbr1 与溶组织内阿米巴虫 Dbr1 晶体结构(PDB:4PEH)进行叠合比对分析发现其 RMSD 值为 1.024, 表明两者结构非常相近, 预测模型可靠性较高。我们根据溶组织内阿米巴虫 Dbr1 晶体结构中小分子 RNA 所结合的位置, 获得纤细裸藻 Dbr1 蛋白质与抑制剂小分子结合的空间坐标位置。通过 *AutoDock Vina* 程序对从 *TargetMol* 天然产物库虚拟筛选后, 我们获得了一个具有潜在结合活性的小分子 Ca。我们通过去除 Dbr1 蛋白表面柔性序列及密码子优化, 并采用 MBP 融合标签方法在大肠杆菌中成功表达了可溶性的 Dbr1 蛋白。我们利用淀粉直链亲和层析柱、离子交换纯化到 MBP-Dbr1 融合蛋白, 在 4℃ 或 33℃ 酶切得到了纤细裸藻 Dbr1 蛋白。我们通过核磁共振化学位移扰动分析(NMR-CSP)实验观察蛋白与配体结合情况。最后我们进行小分子抑制剂的体内活性检测, qPCR 结果表明, 经

过 Ca 处理后的纤细裸藻其内含子表达量要显著高于正常细胞。

4.通过无缝克隆方法成功构建含有 Pep-1 序列的相应质粒。质粒转化 BL21 (DE3) 细胞后进行 IPTG 诱导, SDS-PAGE 和考马斯亮蓝染色显示相应蛋白成功被诱导出。与不含 Pep-1 的对照 GFP 蛋白相比, 培养基中加入 100 μ M 的 PGH 蛋白并 37°C 孵育 12 小时后, 我们发现蛋白进入细胞的效率可达 70%。通过对 Pep-1 不同位置的融合蛋白进行转化研究, 我们观察到 PGH 和 HPG 的转化效率远远高于 HGP 和 GPH, 表明蛋白 N 端的 Pep-1 能更有效地促进细胞内转化。对于 DMSO 介导的转化, 10% DMSO 处理细胞 1 小时, 加入 GFP 蛋白并孵育 12 小时后, 可观察到大约 80% 的细胞都含有荧光。同样 DMSO 也能促进一个含有 Cy5 标记的 1.7 kb 的 DNA 片段进入纤细裸藻细胞, 且跨膜效率与 GFP 蛋白类似。我们同样测试了一个 10 kb 大小的质粒分子, 通过 DMSO 处理细胞后, 培养基中加入质粒分子, 纤细裸藻细胞培养 3 天后, 对提取的细胞 DNA 进行 PCR, 凝胶电泳发现同样能够检测到质粒上特有的一段 736 bp 的片段。

结论:

1.我们将二代和三代 RNA-seq 结果所得剪接基因进行整合对比, 最终我们在比对结果中获得了 166 个纤细裸藻核心剪接基因。利用数据库中纤细裸藻基因组数据, 与人类以及锥虫等生物进行同源序列比较发现了纤细裸藻 U2 和 U6 snRNA 的存在。我们发现布氏锥虫、杜氏利什曼原虫以及 *D. papillatum* 的剪接蛋白数目分别为 78 个、76 个、121 个, 提示古虫界生物在剪接体组成上具有明显差异。通过人类剪接抑制剂对纤细裸藻细胞的作用, 我们发现纤细裸藻细胞对目前报道过的这些抑制剂均不敏感。

2.我们成功鉴定了纤细裸藻中四种不同类型内含子, 即 GT-AG、GC-AG、ss-invariant 和 ss-variant 型。我们发现纤细裸藻 SL-RNA 的 5'ss 区域和 U6 间存在 extended U6/5'ss 螺旋。以及在纤细裸藻的 3'ss 区域富含丰富的胞苷(C), 这与人类和酵母等后鞭毛界生物存在显著差异。本研究为这些内含子的剪接机制以及裸藻的基因组学和进化提供了新的见解。

3.通过虚拟筛选以及实验验证成功鉴定出能够抑制纤细裸藻 Dbr1 酶活性的小

分子抑制剂 Ca，从而为将来研究 Dbr1 功能及内含子种类及数量的鉴定奠定了基础。

4.我们开发了跨膜肽和 DMSO 介导的将外源物质传递到纤细裸藻的转化方法。与农杆菌介导的核转化、电穿孔和微注射等方法相比，本方法快速、方便、高效。

关键词：RNA-seq；内含子；SL-RNA；虚拟筛选；脱支酶；跨膜肽；DMSO；胞内转化

Abstract

Research background and purpose:

Early differentiated pathogens, such as *Trypanosoma* and *Elegans*, are infectious and cause great harm to human health. Biochemical investigation of splicing in these organisms has so far been explored only in SL *trans*-splicing. Most of the *cis*-splicing studies have been conducted on organisms like human or yeast, and scarcely on other species. Therefore, we selected the model organism *Euglena gracilis*, which shares same Excavata groups with *Trypanosoma*.

Euglena gracilis, a unicellular eukaryote, also known as eye worm, contains the organelle chloroplast and utilizes its unique flagellum for locomotion. The genome of *E. gracilis* is comparable to that of the human genome in size, but has more repetitive sequences. Therefore, its complete genome sequence is still unknown at present. Different from humans and yeast, which both belong to the Opisthokonta clade, *E. gracilis* is a Excavata species, and harbors diverse splicing types. To date, only a few genes and their introns have been identified in the draft genome of *E. gracilis*.

Therefore, to lay the foundation for the purification of spliceosomes, we first identified the spliceosomal genes and their introns in *E. gracilis* by bioinformatics analysis and then established a feasible genetic manipulation method. Studies on the model organism *E. gracilis* can give us a better understanding of the spliceosome and expand our knowledge on splicing style.

Materials and methods:

1. RNAs with the polyA tail was enriched from the total RNAs of *E. gracilis*, and library construction was subsequently performed. The full-length transcriptomic and next-generation transcriptomic RNA-seq were combined with available transcriptomic data in the SRA database and were compared by BLASTX. The splicing genes from three RNA-seq results were integrated. Using the genomic data of *E. gracilis* from the NCBI database and comparing with those of humans and *T. brucei*, we obtained the gene sequences of the spliceosomal components of *E. gracilis*. Finally, PCR was performed to validate the identified *E. gracilis* spliceosomal components. In addition, splicing genes from *T. brucei* and *L. major*, which belong to the same clade with *E.*

gracilis, were identified by the BLAST program. Several human splicing inhibitors were applied to *E. gracilis* cells to test their sensitivity.

2. The above full-length transcriptomic RNA-seq and the secondary-generation transcriptomic data were used for the identification of *E. gracilis* introns. Initially, we utilized the BLAST program to remove the chloroplast and mitochondria genomic sequences, as well as the nuclear rDNA sequences, from the above RNA-seq data. Redundant sequences were then removed using cd-hit, resulting in only nuclear mRNA sequences for further analysis. Comparison of mRNA sequences to the genome was performed using exonerate. Finally, the extracted introns were manually examined and categorized. For outtron identification in *E. gracilis*, sequences with the SL exon were extracted from the nuclear mRNA sequences. These sequences were subjected to BLAST against the genome sequences. The resulting outtrons were extracted with the seqkit toolkit.

3. Based on the predicted Dbr1 gene sequence of *E. gracilis* in the previous work, the full-length coding sequence of Dbr1 was obtained by RT-PCR and RACE. The three-dimensional structure of Dbr1 is predicted by AI programs including Alphafold and Robetta. The putative small-molecule binding region on Dbr1 is deduced from the crystal structure of *E. histolytica* Dbr1 in complex with a fragment of an RNA molecule (PDB:4PEH). To identify high-affinity inhibitors of Dbr1, small molecules from a TargetMol natural product library are fed to the AutoDock Vina program. The binding of the selected molecule to the bacterially purified Dbr1 protein is further examined *in vitro*. Inhibition of Dbr1 activity by the small molecule is validated by the readout of splicing defects in nuclear pre-mRNA introns *in vivo*.

4. A series of cell-penetrating peptide (CPP) Pep-1 tag sequences containing plasmids, including PGH (Pep-1-GFP-His), HPG (His-Pep-1-GFP), HGP (His-GFP-Pep-1), and GPH (GFP-Pep-1-His), were constructed from pET21-EGFP by seamless cloning and transformed into bacteria. Upon induction by IPTG, the expressions of Pep-1 fusion proteins were examined by SDS-PAGE and coomassie staining. Fusion proteins were purified via a Ni-NTA affinity chromatography column and dialyzed against PBS. Proteins were added to *E. gracilis* cells and observed under a BX63 fluorescence microscope. In addition, dimethyl sulfoxide (DMSO)-treated *E. gracilis*

cells were added with exogenous materials including GFP protein, a PCR-amplified DNA fragment with a Cy5 label, and a plasmid, respectively, and observed under a microscope except for the plasmid, which was examined with PCR.

Result:

1. Bioinformatics analysis identified 166 spliceosome protein-coding genes, and U2 and U6 snRNA genes. In comparison to the human spliceosome, we found that most of the spliceosomal proteins lacking in *E. gracilis* are from the SR, hnRNP, B*/C, EJC/TREX complexes, and snRNP proteins. The remaining spliceosome proteins are relatively conserved. We further examined spliceosome genes by PCR. The results of a total of 29 spliceosomal protein-coding genes and U2 and U6 snRNA indicated that the predicted gene sequences were all correct. This further suggests that the identified spliceosome genes in *E. gracilis* exist and are authentic.

From the analysis of 166 spliceosome protein-coding genes and 212 sequences of *E. gracilis*, the ratio of spliceosome genes containing SL exons in total *E. gracilis* spliceosomal genes is calculated to be 91%, which further indicated that trans-splicing by SL was common in this organism. Comparing the composition of the spliceosome among the Excavata organisms, we found that there are significant differences between *E. gracilis* and two parasitic organisms, *T. brucei* and *L. major*. We found that *E. gracilis* cells are not sensitive to canonical splicing inhibitors.

2. We identified a total of 28,337 nuclear pre-mRNA introns using the exonerate program, including 11,527 GT-AG type introns, 659 GC-AG type introns, 11,653 ss-variant type introns, and 4,498 ss-invariant type introns. By identifying a large number of introns in *E. gracilis*, including conventional and nonconventional introns in cis-splicing and trans-splicing by SL, we found that conventional introns and outtrons in *E. gracilis* exhibit a unique cytidine-rich sequence and canonical CAG-3' splice site (ss). In addition, the conventional intron has a 5'-GTGTG sequence in common, which is distinct from the 5'ss sequence (GTATA) in spliced leader RNA (SL-RNA). The two terminal sequences of the nonconventional intron are able to form a paired structure. In addition, we describe another splicing type, termed overlapping.

3. We used RACE and overlapping PCR to obtain the intact sequence of *E. gracilis* Dbr1. The AlphaFold and Robetta programs predict essentially similar three-dimensional

structures of the deduced Dbr1 protein. The superposition of *E. gracilis* and *E. histolytica* Dbr1 structures reveals an RMSD value of 1.024, indicating the predicted *E. gracilis* Dbr1 structure is highly reliable. The putative location of the small molecule on the *E. gracilis* Dbr1 structure is deduced from the spatial coordinate of the small molecule RNA in the *E. histolytica* Dbr1 structure. A small molecule of Ca with high binding affinity is obtained from a TargetMol natural product library by virtual screening with AutoDock Vina. We successfully expressed the Dbr1 protein in *E. coli* by removing the surface flexible sequence and codon optimization. We purified the MBP-Dbr1 fusion protein using Amylose Resin and ion-exchange chromatography and enzymatically cleaved it to obtain the untagged *E. gracilis* Dbr1 protein. The protein-ligand binding is assayed by nuclear magnetic resonance-chemical shift perturbation (NMR-CSP). Finally, we test the *in vivo* activity of small-molecule inhibitors. qPCR results showed that intron expression of Ca-treated *E. gracilis* was significantly up-regulated than that of normal cells.

4. Plasmids containing the Pep-1 sequence were successfully constructed through the seamless cloning method. The plasmids were transformed into BL21 (DE3) cells, and cells were induced by IPTG. SDS-PAGE and coomassie staining showed that all proteins were successfully expressed. Upon addition of 100 μ M of PGH protein to cells and incubation for 12 h at 37°C, we found that the Pep-1-tagged protein entered the cells with an efficiency of 70% compared with the GFP protein without Pep-1. We further observed that the transformation efficiencies of PGH and HPG are much higher than those of HGP and GPH, indicating that N-terminal Pep-1 promotes intracellular transformation more effectively. For DMSO-mediated transformation, cells were treated with 10% DMSO for one hour. Fluorescence signals were observed in approximately 80% of cells after adding GFP protein. Similarly, DMSO also promoted intracellular transformation of a 1.7 kb DNA fragment containing a Cy5 label with comparable efficiency to that of GFP. We also tested a 10 kb plasmid and found it could also enter cells, as detected by PCR amplification of a 736 bp plasmid-specific fragment.

Conclusion:

1. We combined the next-generation and full-length transcriptomic RNA-seq results and aligned them to human spliceosome genes, and finally obtained 166 core

spliceosome genes in *E. gracilis*. Using the genomic data of *E. gracilis* in the database, we identified the *E. gracilis* U2 and U6 snRNAs by comparing them with those of humans and *T. brucei*. We identified 78, 76, and 121 splice proteins for *T. brucei*, *L. major*, and *D. papillatum*, respectively, indicating significant differences in spliceosome composition within Excavata. We found that *E. gracilis* cells were insensitive to all reported splicing inhibitors.

2. We successfully identified four different types of introns in *E. gracilis*, namely GT-AG, GC-AG, ss-invariant, and ss-variant. We found the presence of an extended U6/5'ss helix in the 5'ss region of SL-RNA and U6 in *E. gracilis*. A wealth of cytosine (C) in the 3'ss region in *E. gracilis* introns is significantly different from that of Opisthokonta organisms such as humans and yeast. This research provides new insights into the splicing mechanism of these introns as well as the genomics and evolution of Euglenozoa.

3. We successfully screened a small-molecule inhibitor, Ca, that can inhibit the Dbr1 enzyme activity in *E. gracilis*, thus laying the foundation for future studies on Dbr1 function and intron identification.

4. We have developed CPP- and DMSO-mediated transformation methods for the delivery of exogenous materials into this algal. Comparing to other available methods, such as agrobacterium-mediated nuclear transformation, electroporation, and microinjection, our methods are fast, convenient, and efficient.

Key words: RNA-seq; Intron; SL-RNA; Virtual screening; Dbr1; Cell-penetrating peptide; DMSO; Intracellular transformation

前言

许多真核病原生物，如非洲昏睡病致病菌布氏锥虫、杜氏利什曼原虫，在基因表达过程中，由顺式剪接介导的内含子去除是非常罕见的，而由 spliced leader (SL) 序列介导的反式剪接是寄生虫锥虫细胞核前体 mRNA 成熟的主要形式^[1]。在这种剪接形式中，SL RNA 外显子上存在 5'剪接位点，前体 mRNA 第一个外显子的 5'端存在分支点腺苷（branch point adenosine, BP-A）和 3'剪接位点，SL RNA 的外显子与前体 RNA 的 5'端外显子先以剪接体依赖的方式进行结合后才能进行后续剪接过程^[2,3]。SL 序列介导的反式剪接同样存在于纤细裸藻中。

纤细裸藻（*Euglena gracilis*）是一种真核原生生物，因其易于分离和培养，常被用作实验生物学的模型。在进化关系上，与致病性生物锥虫亲缘关系较近，同属于古虫界，但与后鞭毛界的酵母和人类相差较远。因此其在基因表达过程中，特别是在前体 mRNA 剪接这一阶段，与进化关系较远的人类和酵母相比有很大的区别。

真核生物内含子的剪接是保证遗传信息正确传递的重要步骤。内含子的去除需要大分子剪接复合体来行使剪接功能。剪接体分为主要剪接体和次要剪接体，主要剪接体存在于所有真核细胞中，而次要剪接体只存在于人类等少数真核细胞中^[4]。主要剪接体包含了五种富含尿嘧啶的核内小 RNA（U1、U2、U4、U5、U6）和几百种剪接体蛋白，次要剪接体包括 U5、U11、U12、U4atac 和 U6atac^[5]。剪接体蛋白通常以三种形式存在：与某种 snRNA 形成相应的 snRNP 复合物（如 U1A、SF3b 及 Prp8 分别与 U1、U2 和 U5 相结合）、相互形成 non-snRNP 复合物（包括 NTC、NTR、RES 和 EJC）或者独自存在（如 Yju2、Spp2、Cwc25、Prp18 等）。在前体 mRNA 剪接过程中，剪接体会依次经历四个阶段：组装、活化、催化、解聚。近年来冷冻电镜技术的持续发展，使得人类和酵母剪接体在不同状态下的三维结构得到大量解析。迄今为止，已经鉴定出了 12 个主要剪接体的剪接复合物（E、A、pre-B、B、pre-B^{act}、B^{act}、B*、C、Ci、C*、P 和 ILS）和 2 个次要剪接体的 B^{act}剪接复合物^[6,7,8,9,10]。通过对酵母和人类的剪接研究发现，虽然两者同为后鞭毛界，但是剪接体在组成上具有物种差异性。比如在人类中主要和

次要剪接体均存在，而酵母中只有主要剪接体；某些特定蛋白也只存在于某些特定物种中；同一蛋白在不同的物种中所属的蛋白复合物不同；酵母和人类剪接构象的差异性。此外，通过对拟南芥、红藻中剪接体的生物信息学鉴定，以及对病原寄生虫锥虫剪接体蛋白的实验鉴定发现不同进化分支的生物，剪接体在组成上存在着较大差异。值得注意的是，布氏锥虫和克氏锥虫中存在着许多不同于人类和酵母的特异性剪接蛋白^[11,12]。

纤细裸藻最突出的特点是其剪接形式的多样性。在纤细裸藻细胞核基因中，例如 *tubA*、*gapC* 等基因，存在着常规与非常规这两种内含子形式。常规内含子与人类和酵母的相同，都为 GT-AG 型内含子，而非常规内含子的边界序列并不是 GT-AG 型。此外，纤细裸藻的叶绿体中存在着大量的 II 型内含子和特殊的 III 型内含子。III 型内含子不仅存在于叶绿体基因编码序列上，也存在于基因之间的非编码区域，其在结构上比 II 型内含子简单，仅包含 II 型内含子的 domain I 和 VI 区域。有研究推测 III 型内含子是从 II 型内含子进化而来的。此外在纤细裸藻叶绿体中 II 型和 III 型内含子还可以相互嵌合形成 *twintron* 这种复杂的内含子形式^[13]。

目前，我们对剪接体的认识主要来自人类和酵母，对其他物种剪接体的研究相对较少。在纤细裸藻中只有 SL RNA、U1 和 U5 snRNA 分子已被鉴定出来，其他 snRNA 分子和蛋白成分仍然未知^[14]。为了拓展我们对剪接体物种差异性的认识，本研究首先对纤细裸藻剪接基因进行生物信息学分析，并进一步通过实验方法来进行分子生物学鉴定。其次，由于目前纤细裸藻细胞核基因中只报道了少数内含子^[15]，内含子特征并不明确。为此，我们还利用生物信息学方法，通过比较纤细裸藻基因组和转录组，大规模鉴定了细胞核常规及非常规内含子，并对其分子特征进行了详细的分析。此外，我们还通过虚拟筛选方法寻找针对剪接体脱支酶 Dbr1 的小分子抑制剂，期望通过抑制剪接来对其内含子进行实验鉴定。最后我们建立了快速将外源物质转入纤细裸藻细胞的转化方法，为将来对纤细裸藻的分子和细胞生物学研究奠定了基础。对纤细裸藻剪接系统的研究将拓展我们对剪接多样性的认识，也将加深我们对剪接机制的理解。

第一章 纤细裸藻剪接体组分鉴定

1 目的

纤细裸藻一个突出而独特的特征是剪接的多样性，包括核前体信使 RNA (precursor mRNA, pre-mRNA) 基因的常规和非常规内含子剪接，以及叶绿体基因的 II 和 III 类内含子剪接。SL-RNA 介导的反式剪接，即用 SL-RNA 序列中 SL 外显子取代 pre-mRNA 分子的 5' 端内含子序列 (称为 outtron)。这种剪接形式在人类和酵母中是不存在的^[16]。然而却存在于纤细裸藻和寄生锥虫等古虫界生物中^[17]。与人类和酵母类似，纤细裸藻 pre-mRNA 常规内含子以及 outtron 和 SL RNA 具有相同的 5' GT 和 3' AG 剪接位点^[18]。因此，这两种内含子的剪接被认为是由剪接体催化的。剪接体由 5 个小 RNA (snRNA, U1, U2, U4, U5 和 U6) 和数百种蛋白质组成^[7]。剪接体蛋白可分为 snRNA 相关蛋白 (包括 Sm/LSm 核心蛋白和各自的 snRNA 相关蛋白) 和非 snRNA 相关蛋白或蛋白复合物。剪接体通过组装、激活、催化和解聚等阶段逐步催化内含子的去除。此前，剪接体 U1 和 U5 snRNA 以及 SL RNA 已在纤细裸藻中被鉴定出来^[19,20,21]。然而，其他的 snRNA 或剪接体蛋白是否存在于纤细裸藻中尚不清楚。因此，我们计划通过对二代、三代全长转录组测序数据进行生物信息学分析，对纤细裸藻剪接相关组分进行了研究。

2 材料与方法

2.1 细胞株

表 1-1 细胞名称、培养条件和来源

Table 1-1 Cell name, culture conditions and source

细胞名称	细胞种类	培养条件
Euglena gracilis	FACHB-848	HUT 培养基

以上细胞株购买于中国科学院水生生物研究所，置于成都医学院科研实验中心恒温培养振荡箱中培养，培养参数为 25℃、100 rpm、2000 lux。

2.2 主要试剂

表 1-2 试剂名称及来源

Table 1-2 Reagent name and source

试剂名称	生产公司
SMARTer PCR cDNA Synthesis Kit	Takara, 日本
FastKing cDNA 第一链合成试剂盒(去基因组)	天根, 中国
Phanta Super-Fidelity DNA Polymerase	诺唯赞, 中国
Trizol	生工, 中国
DEPC	生工, 中国
氯仿	科隆, 中国
异丙醇	科隆, 中国
75%乙醇	科隆, 中国
无水乙醇	科隆, 中国
DNA 分子量标准 Marker D (100~2000 bp)	生工, 中国
二甲亚砜 (DMSO)	MP, 美国
潮霉素	生工, 中国

2.3 主要仪器及耗材

表 1-3 仪器设备名称及来源

Table 1-3 Reagent name and source of instruments and equipment

仪器名称	生产厂家
BX63 正置荧光显微镜	奥林巴斯, 日本
Ultrospec 3100 pro 分光光度计	Amersham Biosciences, 美国

2.4 本实验相关引物

实验中所用到的 PCR 引物均由 Primer6 软件设计, 由擎科生物科技有限公司合成, 见表 1-4。

表 1-4 PCR 引物序列

Table 1-4 PCR primer sequences

基因	序列 (5'-3')
egSmD1_F	GTGTTCAATGAAGCTCGTAAG
egSmD1_R	GCAGTCAAGAGCCGCCACC
egSmE_F	TGGTTGGGCAAACAGCAGC
egSmE_R	GCGGGGCATGTACAGGGTG
egSmF_F	GCTCCACCTATGAGCTTTGAG

续表 1-4

基因	序列 (5'-3')
egSmF_R	TGATTGGGCAGCCAACTATT
egSmG_F	CAACTCACACAAAAAATGGC
egSmG_R	GCTGCCGCACAAAAAACC
egLsm4_F	AGCCCACCCCATATGCTG
egLsm4_R	GAACAGGGCAGGGGATCGG
egLsm5_F	ACTACCACATGGCTGGCGC
egLsm5_R	CAAACGTGGGCAGTGGTGG
egLsm6_F	GGGCTTCACAATGGCCGAC
egLsm6_R	TGTCGTTGAGGAGGAGGGG
egLsm7_F	CTCAACAAGTAATTGAATGCCTGC
egLsm7_R	GTGGAACGACAGCGGAGAA
egLsm8_F	GCCAGCCATGTCTCAGCAA
egLsm8_R	GCCAGTTGGTGGAGGTTGT
egU1C_F	GGGTCGGTGGGATGCCGA
egU1C_R	TGGGGTCGGTCCCAGGCT
egU2A_F	TCGAATGCGGTTGACAGC
egU2A_R	TGATGAGCATCCACCACAG
egU2B_F	GGGTCTTTTTTGGAGGGCG
egU2B_R	GGGATGTGTACCAGCGAGG
egSF3B1_F	ACGAGAAGGTGCAAGAGAATTG
egSF3B1_R	CTGGTGGCTGATGTATATGTTGT
egSF3B5_F	ATGGCCGACCCATTCTGC
egSF3B5_R	GTGTCACCGTGTCACTGTG
egSF3B6_F	CCTCCCCCGCACCATGT
egSF3B6_R	CGCCCCAGAGGGTTCACC
egPHF5A_F	CGCAAACCATGGCAAAGCA
egPHF5A_R	ACGACGTCATGCGGCGAGG
egPrp31_F	CATTATGGCAGAGGCTTTGG
egPrp31_R	GCAAGGAGAGTGAGAAGAGG
egSnu13_F	AAGTTAGGATCTGCTTCTGGTTCAA
egSnu13_R	CGTCATCTGCGTGGCTTCTT
egPrp8_F	CGGCCAAGATGTCCATTCC
egPrp8_R	GGGGGGGAAAGGTTGAGTATG

续表 1-4

基因	序列 (5'-3')
egP8-1.5k_F	GAAGTGCTGGGTGCTCAATG
egP8-1.5k_R	CATTGAGCACCCAGCACTTC
egP8-3k_F	CAACAACCTGACGGACATCTG
egP8-3k_R	CAGATGTCCGTCAGGTTGTTG
egP8-4.7k_F	GGAGAGCATGAAGTACAAGAAGC
egP8-4.7k_R	GCTTCTTGTACTTCATGCTCTCC
egP8-6.3k_F	CTCATCACCAACCTCACC
egP8-6.3k_R	GGTGAGGTGGTGGTGATGAG
egU5-15K_F	TGCTGCCTCATTTGCACAG
egU5-15K_R	GGTCTACAGCTGACCCCCC
egCwc15_F	GCCCCCCCCATGACGACC
egCwc15_R	GGGCACGGCGGCTATTTG
egCTNNBL1_F	CGTGGAGAAGGCGGAGAATGA
egCTNNBL1_R	TCACCGTTGGCTCCCGAAAC
egBud13_F	AGACCAGCCTGACGATGATGAA
egBud13_R	GGGAGTGGTCTTCAGCTCTGTT
egY14_F	TTGACGAGGAAGACGCCACTT
egY14_R	GCAATTCGGTGCAGTCTCACAA
egSUGP1_F	CAGCGTAGAGCGTTCCAGAGTA
egSUGP1_R	TTCCGAGGGTCAGACACATCAG
egUBL5_F	CGATGCCGATGATCGAGGTGA
egUBL5_R	TCAGGTAACGGACAGCAAGAGG
egFAM32A_F	AGCAATGTCATTCCTGGCAAGC
egFAM32A_R	GCACCACCTCGTGAGAAACAAG
egCDK10_F	TGGAAGGAGTTGCTGCTTGCT
egCDK10_R	GGGATGCTCGGCGAAATATGG
egUAP56_F	AGCAAAATGGCGGATGGTG
egUAP56_R	GTACAGAGAGGAGGGCAGG
egU2U1-369779_F	CGTCGGAAGTGTCTGATGGTCT
egU1U2-369779_R	CTGCCGCCGACCTCATAATC
egU2-1198033_F	ACCAGTGCAGTCTGGATGGC
egU2-1198033_R	TGGCTTCGCTTGGTGGCTTC
egU6-1250410_F	CGGTCTCCAAGTCCTCAGTCAG
egU6-1250410_R	CAGTTGCCGTCGGTAGAATGC

2.5 常规实验试剂配置

表 1-5 HUT 培养基

Table 1-5 HUT culture medium

试剂名称	用量
KH ₂ PO ₄	0.02 g
蛋白胨	0.60 g
MgSO ₄ ·7H ₂ O	0.03 g
酵母提取物	0.40 g
乙酸钠	0.40 g
柠檬酸钾	0.04 g
10 ⁻⁴ ×维生素 B12	0.10 mL（现用现加）
10 ⁻⁴ ×维生素 B1	0.10 mL（现用现加）
浓盐酸调节 pH 值	6.40
去离子水定容至	1 L
高压灭菌	

表 1-6 10⁻⁷×维生素 B12Table 1-6 10⁻⁷×vitamin B12

试剂名称	用量
维生素 B12	5 g
去离子水定容至	1 L

表 1-7 10⁻⁴×维生素 B12Table 1-7 10⁻⁴×vitamin B12

试剂名称	用量
10 ⁻⁷ 维生素 B12	1 mL
去离子水定容至	1 L

表 1-8 10⁻⁴×维生素 B1Table 1-8 10⁻⁴×vitamin B1

试剂名称	用量
维生素 B1	4 g
去离子水定容至	1 L

2.6 实验方法

2.6.1 小分子对纤细裸藻的作用

将剪接抑制剂 Pladienolide B (PB) (Adooq Bioscience, 美国)、FR901464 (Adooq Bioscience, 美国)、Herboxidiene (GEX1A) (Cayman, 美国) 以及 OTS964 (TargetMol, 美国) 均溶于二甲亚砜 (DMSO) 中, 并以适当的浓度加入到纤细裸藻细胞中。培养阶段利用 Ultrospec 3100 pro 分光光度计测量细胞浓度, 培养结束后利用 BX63 显微镜观察细胞形态。

2.6.2 纤细裸藻转录组测序

首先, 我们将纤细裸藻细胞送公司进行 RNA-seq 测序。由于我们只对剪接基因进行分析, 只需利用带有 Oligo(dT) 的磁珠从抽取的总 RNA 中富集带 polyA 的 RNA 即可。随后进行 cDNA 文库构建, 利用逆转录试剂盒 (Clontech SMARTer PCR cDNA Synthesis Kit), 以 mRNA 为模板, 合成 cDNA 链, 并将 cDNA 进行末端修复以及连接测序接头, 因此得到了最终的测序文库。利用 AATI 检测文库片段的完整性后立即进行上机测序。本实验使用诺禾致源 (北京, 中国) PacBio Sequel 三代测序系统进行。为了保证数据的可靠性, 我们同时使用 Illumina 进行二代无参转录组测序。

2.6.3 测序数据分析

首先, 使用 Trimmomatic 程序对三代测序获得的 FLNR 原始数据去除测序接头并对其进行直接进行人工 NCBI-BLASTX 分析。随后我们从 SRA 数据库中下载前人的纤细裸藻 RNA-seq 数据 (SRR2094880^[22]; SRR2628535^[15]), 以及通过二代测序获得 Reads 原始数据, 然后将两者分别运行 Fastq-dump 程序进行格式转换和 Trinity 进行无参转录组拼接, 并对 Trinity 拼接文件与 UniProtKB 数据库页面下载 reviewed 的 Fasta 格式的压缩包通过本地化 BLASTX 程序进行比对。然后在比对结果文件中查找核心剪接基因, 再根据比对结果文件中的 Trinity 编号找到 Trinity 拼接文件中对应的序列, 对此序列再进行 NCBI-BLASTX 分析, 以最终确认是否属于剪接基因。最后, 将三个 RNA-seq 结果所得剪接基因进行整合对比, 以获得更多的剪接基因及每个基因更完整的全长序列, 并对可能的剪接基因阅读框进行

预测，获得纤细裸藻完整的核心剪接基因。

根据从 NCBI 网站上检索到的人类剪接体蛋白质。将纤细裸藻与布氏锥虫等原生生物的剪接体蛋白编码基因序列通过 BLAST 程序与人类剪接体蛋白进行人工比对，随后对其开放阅读框进行注释。

2.6.4 PCR 分析鉴定

我们根据生信分析获得的纤细裸藻剪接基因序列设计引物（表 1-4），使用高保真酶 (Phanta Super-Fidelity DNA Polymerase) 进行 PCR 验证，在对 Prp8 进行扩增时，将其全长划分为五个片段进行扩增。将 PCR 产物送至擎科（北京，中国）进行 Sanger 测序分析，进一步验证我们所鉴定出来的纤细裸藻剪接基因是否存在。

2.6.5 纤细裸藻总 RNA 的提取

使用无酶枪头将纤细裸藻细胞液吸取到无核酸酶的 1.5 mL EP 管中；使用离心机 1000 g，1 min 收集细胞，弃上清后，加入 1 mL DEPC 水润洗细胞一次，使用相同的离心条件进行沉淀细胞。随后根据沉淀量加入 Trizol 试剂进行裂解细胞，连续多次吹打混匀，室温静置 5 分钟；加入 200 μ L 氯仿至裂解液中，旋涡混合器震荡混匀 15 秒，室温静置 2-3 分钟；4 $^{\circ}$ C 12000 g，15 min 离心，离心结束后 EP 管中的裂解液将分为三层，小心吸取上层液至新的 EP 管中；再往其中加入 500 μ L 的异丙醇溶液，室温静置 10 分钟；4 $^{\circ}$ C 12000 g，10 min 离心，弃上清，向沉淀中加入 1 mL 75% 无水乙醇，吹打混匀后再次在 4 $^{\circ}$ C 下进行 7500 g，5 min 离心；最后将上清液吸取干净丢弃，沉淀室温放置干燥，待 EP 管中 75% 无水乙醇挥发完毕加入 DEPC 水进行溶解，将水溶液状态的 RNA 立即置于冰上放置。

2.6.6 纤细裸藻 cDNA 的合成

以纤细裸藻总 RNA 为模板，按照 FastKing cDNA 第一链合成试剂盒进行逆转录。于冰上配置下列基因组 DNA 去除反应混合液，见表 1-9。

表 1-9 基因组 DNA 去除反应体系

Table 1-9 gDNA removal reaction system

组成成分	用量
5 $\times$ gDNA Buffer	2 μ L
Total RNA	50 ng-2 μ g
RNase-Free ddH ₂ O	up to 10 μ L

彻底混匀，并简短离心，置于 42℃，孵育 3 分钟，然后置于冰上。于冰上配置反转录反应体系混合液，见表 1-10。

表 1-10 反转录反应体系

Table 1-10 Reverse transcription reaction system

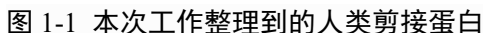
组成成分	用量
10×King RT Buffer	2 μL
FastKing RT Enzyme Mix	1 μL
FQ-RT Primer Mix	2 μL
RNase-Free ddH ₂ O	up to 10 μL

将反转录反应中的 Mix，加到 gDNA 去除步骤的反应液中，充分混匀。42℃，孵育 15 分钟。95℃，孵育 3 分钟之后放在冰上，得到的 cDNA 用于后续实验。

3 结果与分析

3.1 纤细裸藻剪接体蛋白基因的生信分析

根据对人类剪接体复合物中蛋白质组成的研究^[23,24,25,26]，我们整理了人类 257 种剪接体蛋白（图 1-1）。包括 110 种小核糖核蛋白（snRNP）和 79 个非 snRNP 蛋白。snRNP 蛋白包括 14 个 Sm/LSm 蛋白，9 个 U1 相关蛋白，21 个 U2 相关蛋白，5 个 U4 相关蛋白，8 个 U5 相关蛋白，1 个 U6 相关蛋白，5 个 U4/U6.U5 复合物蛋白，19 个 NTC、NTR 以及 IBC 蛋白，3 个 RES 蛋白，15 个 EJC/TREX 蛋白，7 个 ATP 酶，2 个 CBC 蛋白，1 个脱支酶。非 snRNP 为 79 个阶段特异性蛋白，包括复合物 A 中 15 个蛋白，1 个 pre-B 蛋白，复合物 B 中 11 个蛋白，1 个 pre-B^{act} 蛋白，复合物 B^{act} 中 7 个蛋白，复合物 B^{*}/C 中 19 个蛋白，1 个 pre-C^{*}，复合物 C^{*}/P 中 9 个蛋白，复合物 C/C^{*} 中 13 个蛋白和复合物 ILS 中的 2 个蛋白。除上述 189 个蛋白外，还有 12 个 SR 蛋白、23 个 hnRNP 蛋白和 8 个 pre-mRNA 结合蛋白，25 个其它蛋白。



Note: A is thirteen spliceosomal complexes is formed in splicing pathway; B is the 257 human spliceosomal proteins.

由于二代无参转录组 RNA-seq 分析没有基因组序列做指导，拼接过程可能会出现一些错误，因此我们又进行了三代全长转录组测序。在进行全长转录组分析中，我们总共获得了 258944 条测序序列。将全长转录本数目进行统计处理后，

获得了 232051 条全长非嵌合序列。利用层次聚类分析以及 cd-hit 去除冗余序列，得到了 21850 个转录本以及 9669 条基因。随后将三代中得到的全长非嵌合序列与二代分析得到的 166 个剪接蛋白基因和 213 条序列进行 BLAST 比对分析，经过上述相同处理，最终在 21850 个转录本以及 9669 条基因中得到了 130 个剪接蛋白和 168 条序列（图 1-2）。

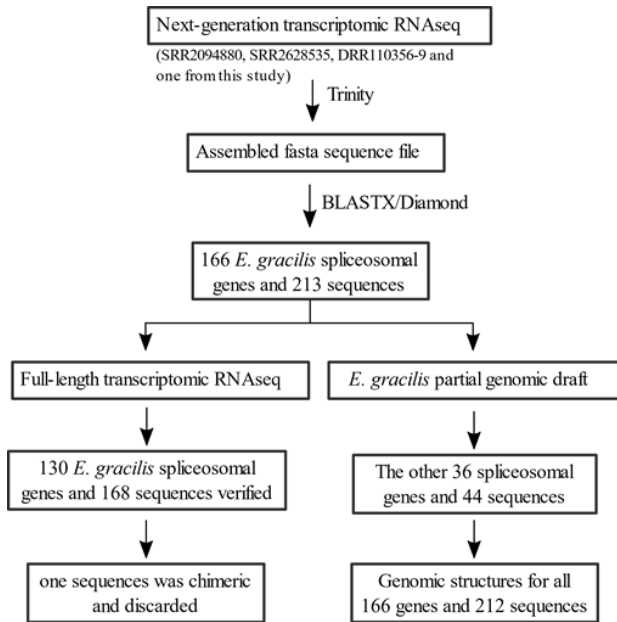


图 1-2 纤毛藻剪接蛋白基因生信分析流程

Fig.1-2 Working procedure for bioinformatic analysis of *E. gracilis* spliceosomal protein gene

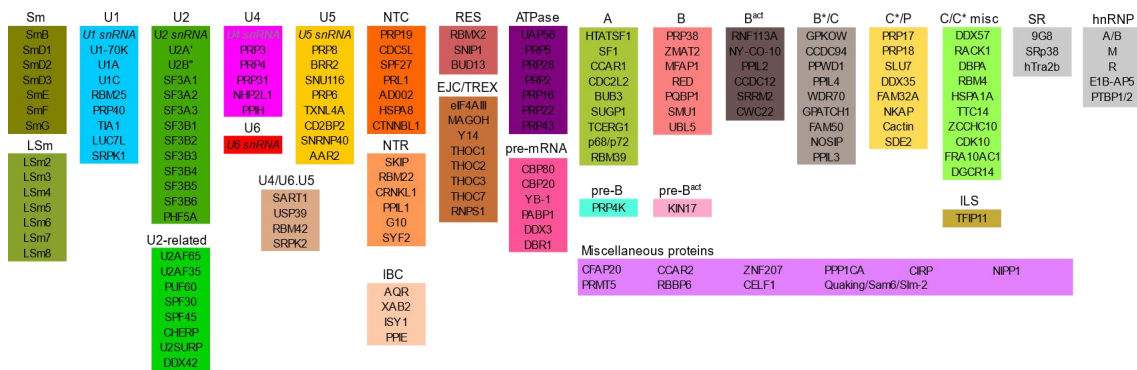


图 1-3 纤毛藻剪接蛋白基因

Fig.1-3 *E. gracilis* spliceosomal protein gene

由上述得知，36 个剪接蛋白基因单独存在于二代转录组分析中。通过对 NCBI 网站下载的 *E. gracilis* 基因组数据（基因组编号 GCA_900893395.1）进行分

析，我们发现这 36 个基因的序列都存在于纤细裸藻基因组上并且均有其相对应的重叠区域。比如，其中的 SF3B5 和 SF3B6 这两个剪接核心蛋白（图 1-4），在基因组上对应区域的序列为 sga_contig_165869 和 sga_contig_1456208。

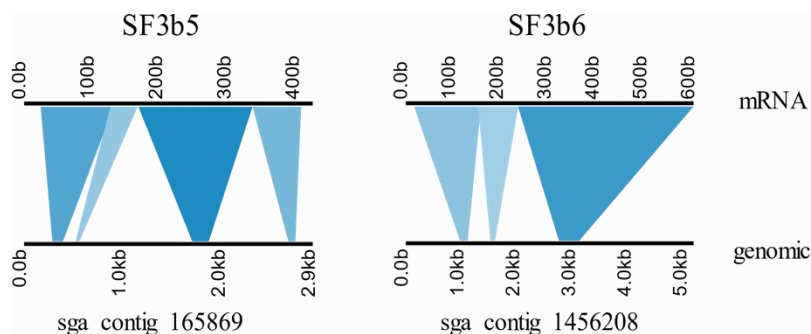


图 1-4 纤细裸藻剪接蛋白基因 SF3B5 和 SF3B6 的基因组鉴定

Fig.1-4 Genomic identification of *E. gracilis* spliceosomal protein genes SF3B5 and SF3B6

注：通过 kablamo 程序分析得到 SF3B5 和 SF3B6 的 mRNA 序列在基因组序列上对应的位置。

Note: mRNA sequences of SF3B5 and SF3B6 were projected to correspond on the genomic sequence by kablamo.

3.2 纤细裸藻 U2 与 U6 snRNA 的分析

在纤细裸藻中，只有 U1 snRNA 和 U5 snRNA 被鉴定过，其余剪接体的 snRNA 都是未知的。*E. sulcatum* 与 *E. gracilis* 进化关系较近，且 *E. sulcatum* 中的 U2 与 U6 snRNA 是已被报道过^[22]。因此，我们以 *E. sulcatum* 中鉴定到的 U2 与 U6 snRNA 为模板，利用 BLAST 程序在纤细裸藻基因组中查找相应的同源序列。我们发现序列 sga_contig_369779 中有与 *E. sulcatum* U2 高度相似的约 200 bp 大小的片段，并且 8 kb 的长度上有 10 个 U2 多拷贝片段（图 1-5）。接下来我们利用同样的方法鉴定了 U6。然而我们根据进化关系较近的 *E. sulcatum* 和 *T. brucei* 的 U6 序列，没有发现纤细裸藻的同源序列^[27]，但根据进化关系较远的人类的 U6 鉴定出了纤细裸藻的同源序列。我们根据对比出来的 U2 (sga_contig_369779) 和 U6 (contig_1250410) 基因组序列（图 1-6）进一步推测出了 U2 与 U6 snRNA 的二级结构（图 1-7）。

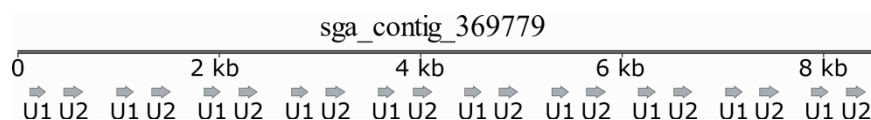


图 1-5 纤细裸藻中 U2 snRNA 的鉴定

Fig.1-5 Genomic identification of *E. gracilis* snRNA genes U2

>U2

```
ATACCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATC
AGCTTAATTTCTGATATATCCAGCACTGCTGGATAAGAATTAATACTAAT
TTCTTGTCGGGGGCCAGGGTTCCTCTGCTTGCACGAGTCCTGCGC
CGCGAGTCGCCCTGCTATTGCACTACCGGCAGGCGTGGCTCACCC
```

>U6

```
GGAGCCCTTCGGGGACATCCGCTAAAATTGGAACGATACAGAGAAGATT
AGCATGGTCCCTGCGCAAGGATGACACGCATTAATCGAGAAAGTATCACC
TTTT
```

图 1-6 纤细裸藻中 U2 和 U6 snRNA 的基因组序列

Fig.1-6 Genomic sequences of *E. gracilis* snRNA genes U2 and U6

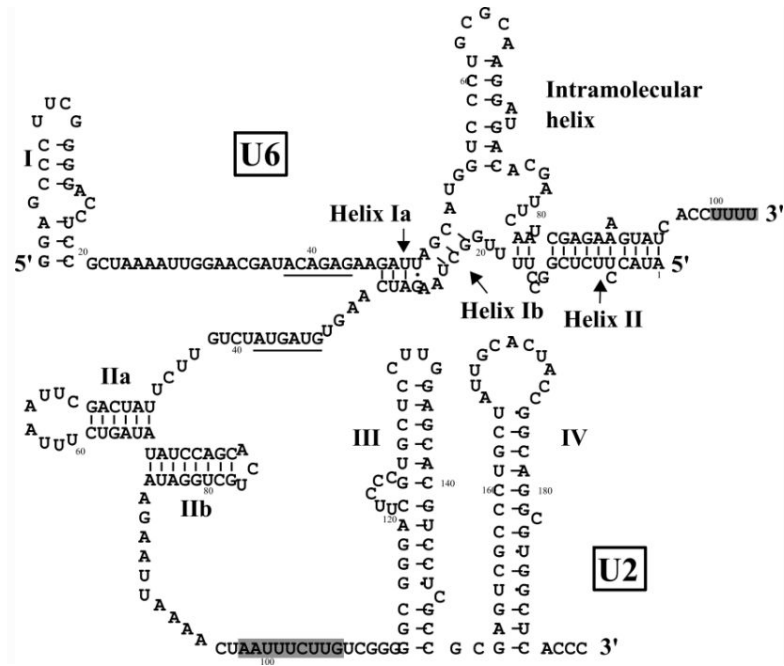


图 1-7 纤细裸藻中 U2 和 U6 snRNA 的二级结构图

Fig.1-7 Secondary structure of *E. gracilis* snRNA genes U2 and U6

注：灰色框表示 U2 的 Sm 结合位点和 U6 的 LSm 结合位点，下划线表示 U2 的分支点 (BP-A) 结合区和 U6 的 5'剪接位点 (5'ss) 结合区。

Note: Gray boxes indicate the Sm binding site of U2 and the LSm binding site of U6, underlines indicate the branch point (BP-A) binding region of U2 and the 5' splice site (5'ss) binding region of U6.

3.3 纤细裸藻剪接体组分的生化鉴定

我们将提取到的纤细裸藻总 RNA 进行逆转录，并以逆转录得到的 cDNA 为模板。随后选取 29 个纤细裸藻剪接蛋白基因进行 RT-PCR 验证，以及 U2 snRNA 和 U6 snRNA 的 PCR 验证。PCR 检测结果表明，这 29 个剪接基因和 U2 与 U6 snRNA 的序列都是正确的(图 1-8)。测序表明扩增序列与拼接序列是相同的，这进一步提示我们所鉴定出来的纤细裸藻剪接基因存在且序列正确。

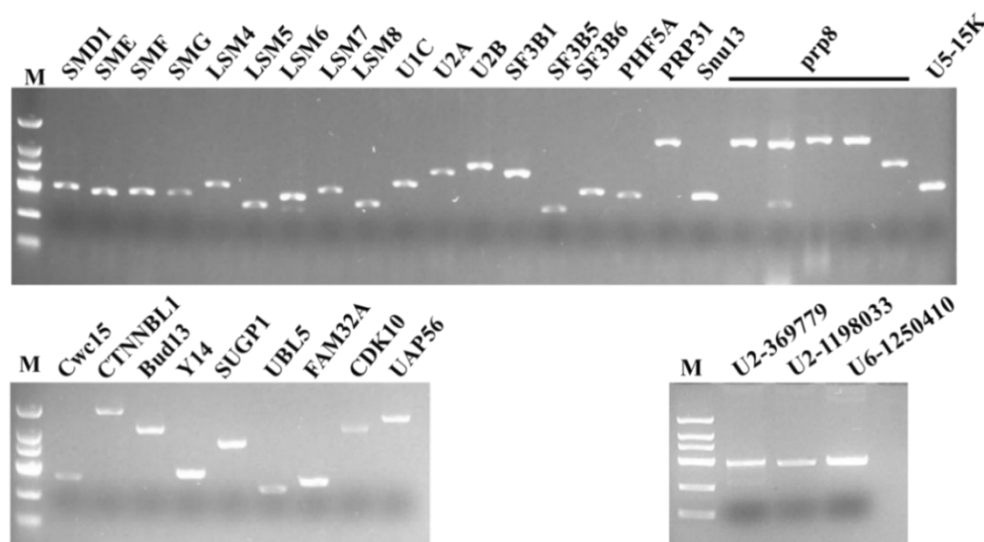


图 1-8 纤细裸藻剪接蛋白基因以及 U2、U6 snRNA 的 PCR 鉴定

Fig.1-8 PCR verification of *E. gracilis* spliceosomal proteins genes and snRNA genes U2 and U6

注：U2 以及 U6 后的编号是由基因组 GCA_900893395.1 中的序列号得来的。

Note: Numbers after U2 and U6 are genomic contig numbers according to GCA_900893395.1.

3.4 SL 介导的反式剪接在纤细裸藻中的普遍性

纤细裸藻一个显著的特点是具有 SL 介导的反式剪接。在鉴定纤细裸藻剪接蛋白编码基因时，我们注意到许多基因上都含有 SL 外显子。在得到的纤细裸藻 166 个剪接蛋白基因和 212 个序列中，我们发现其中 193 个序列均含有 SL 外显子，占 91%。此外，我们还检测了三代全长 RNA-seq 中反式剪接的比例。在最终的 9669 个基因中，有 7028 个基因有 SL 外显子，占 72.7%。这表明，大多数纤细裸藻基因都存在 SL 介导的反式剪接。

3.5 纤细裸藻与其它物种剪接体蛋白质组的比较

纤细裸藻与布氏锥虫、杜氏利什曼原虫在进化关系上都属于古虫界。因此，我们也检测了这些物种的剪接体组成。利用 BLASTP 程序我们共搜索到布氏锥虫

78 个剪接蛋白和杜氏利什曼原虫 76 个剪接蛋白。我们发现这两种病原生物只保留了非常保守的 snRNP 蛋白，包括 Sm/LSm、U2、U4/U6、U5 和 ATP 酶蛋白。布氏锥虫与杜氏利什曼原虫剪接蛋白比较结果表明，两者的剪接体蛋白组成仅有非常微小的差异，其中 SRSF7、hnRNP D 和 hnRNP F/H 仅存在于布氏锥虫中，而 Y14 仅存在于杜氏利什曼原虫中（见附录一）。此外，我们进一步鉴定了同属古虫界的海洋原生生物 *Diplonema papillatum* (*D. papillatum*) 的剪接体组成，发现了 121 个剪接蛋白。综上所述，我们的研究表明，纤毛裸藻与布氏锥虫和杜氏利什曼原虫在剪接体组成上存在明显差异，这种差异可能与它们不同的生存方式密切相关。

3.6 纤毛裸藻对剪接抑制剂不敏感的分子基础

目前已发现许多剪接抑制剂（如 PB 和 GEX1A）可通过作用于特定的剪接体蛋白来干预人类剪接。我们选取了四种剪接抑制剂，分别作用于纤毛裸藻细胞，培养结果表明纤毛裸藻对这些抑制剂均不敏感（图 1-9）。

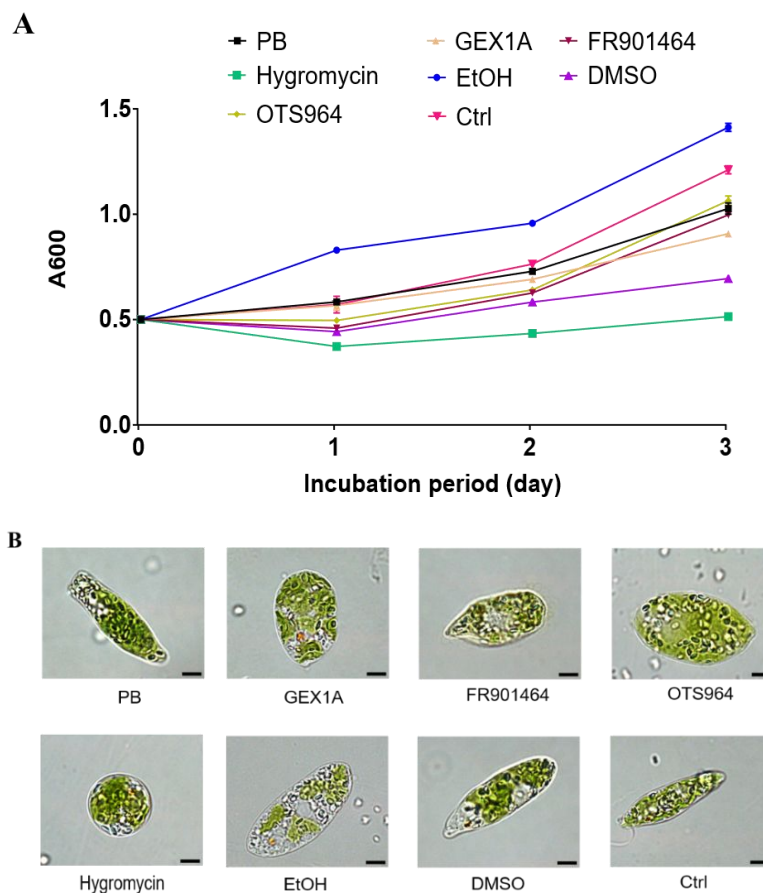


图 1-9 剪接抑制剂对纤细裸藻生长和形态的影响

Fig.1-9 Effect of splicing inhibitors on *E. gracilis* growth and morphology

注：A 为不同的剪接抑制剂对纤细裸藻生长的作用。剪接抑制剂包括 PB（终浓度为 1 μmol ）、GEX1A (1 μmol)、FR901464 (20 μmol) 和 OTS964 (10 μmol)。连续三天每天测量细胞浓度。将促进纤细裸藻细胞生长的无水乙醇作为阳性对照，以 1:1000 的比例添加。潮霉素 (50 $\mu\text{g/mL}$) 可抑制纤细裸藻细胞生长，作为阴性对照。二甲亚砜作为溶剂对照以 1:100 的比例加入到纤细裸藻细胞中；B 为用不同的剪接抑制剂处理纤细裸藻细胞三天后的细胞形态。在 100 倍物镜下拍摄。标尺：10 μm 。

Note: A is growth of *E. gracilis* treated with different splicing inhibitors. Splicing inhibitors including PB (final concentration 1 μmol), GEX1A (1 μmol), FR901464 (20 μmol) and OTS964 (10 μmol) were added to *E. gracilis* cells, and cell density was measured daily for three days. EtOH, which promotes growth of *E. gracilis* cells, was used a positive control and added at a ratio of 1:1000. Hygromycin (50 $\mu\text{g/mL}$) inhibits *E. gracilis* cell growth and was used as a negative control. DMSO was added as a solvent control at a ratio of 1:100 to the *E. gracilis* cell; B is cell morphology of *E. gracilis* after treatment with different splicing inhibitors for three days. The micrographs were taken at 100 $\times$ magnification. Scale bar:10 μm .

由于 PB 类抑制剂作用于 SF3B1/PHF5A 复合物^[28]（图 1-10A），我们比较了人类和纤细裸藻的 SF3B1 和 PHF5A 蛋白序列，SF3B1 在两者之间的相似度为 79%，而 PHF5A 为 92%。两种蛋白质的序列比对结果表明，在人类 SF3B1 蛋白上与 PB 相互作用的 8 个残基中，1078 位的缬氨酸 (Val) 在纤细裸藻中变成了异亮氨酸 (Ile)。人类 PHF5A 蛋白中第 36 位的氨基酸残基变成了半胱氨酸 (Cys)，在纤细裸藻中变成了组氨酸 (His)（图 1-10B）。此前有研究报道过这两个残基的突变使 PB 对 SF3b 结合的能力减弱^[28,29]，因此我们认为纤细裸藻不敏感的原因即在于此。

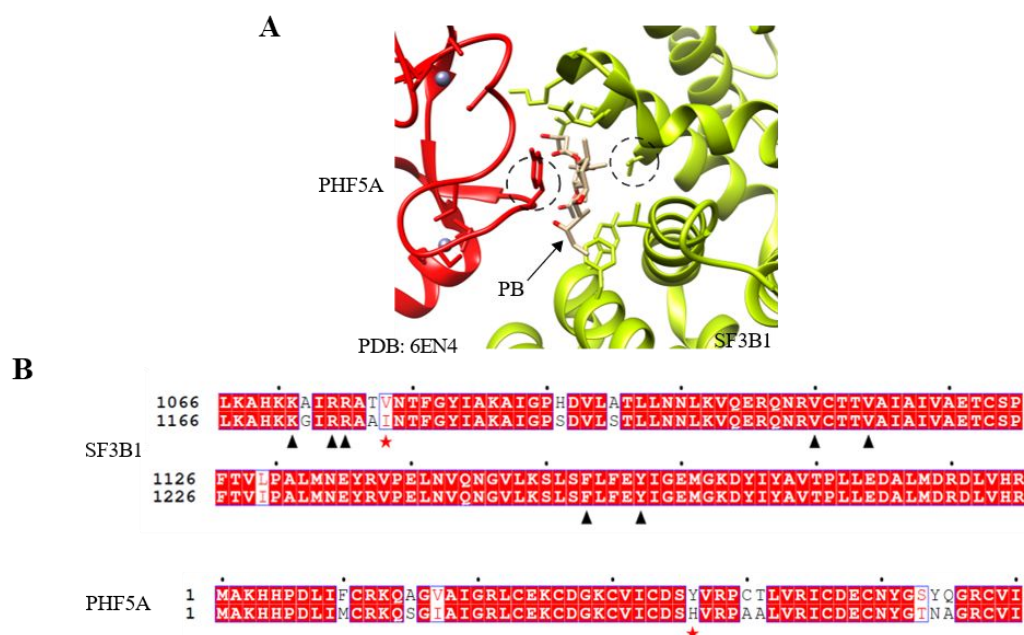


图 1-10 人类与纤细裸藻的 SF3B1 和 PHF5A 蛋白比较

Fig.1-10 Comparison of SF3B1 or PHF5A protein between humans and *E. gracilis*

注：A 为人类 SF3B1 和 PFH5A 蛋白与 PB 作用三维结构图。PHF5A 为红色，SF3B1 为绿色。SF3B1 中 1078 位缬氨酸残基和 PHF5A 中 36 位半胱氨酸残基用虚线圆圈标记。在 SF3B1 和 PHF5A 之间的 PB 用箭头表示；B 为人类与纤细裸藻 SF3B1 和 PHF5A 蛋白的序列比对。每条比对序列中，上面的序列来自人类蛋白质，下面的序列来自纤细裸藻。影响小分子抑制剂与蛋白相互作用的残基用红星标出。SF3B1 蛋白中与 PB 相互作用的其他残基用黑色三角形标出。

Note: A is three-dimensional structure of human SF3B1 or PFH5A with PB. PHF5A is colored red, and SF3B1 is colored green. SF3B1 Val¹⁰⁷⁸ and PHF5A Cys³⁶ are labeled with dashed circles. PB, which is harbored between SF3B1 and PHF5A, is indicated with an arrow; B is sequence alignments of human and *E. gracilis* SF3B1 or PHF5A. For each alignment, the upper sequence was from human proteins, and the lower sequence was from *E. gracilis*. Residues in proteins that influence the interaction with small molecule inhibitors are indicated with red stars. Other residues in SF3B1 that interact with PB are indicated with dark triangles.

此外，我们的实验结果表明，靶向人类剪接体 CDK11B/CDC2L1^[30,31,32]（图 1-11A）的 OTS964 对纤细裸藻生长仍然没有影响。在纤细裸藻中，我们只发现了一种蛋白与 CDK11 的两个同源物 (CDK11A/CDK11B) 的 C 端激酶结构域具有大约 70% 的相似性。人类 CDK11 维持 OTS964 结合活性的两个残基是 572 位的组氨

酸(His)和 579 位的甘氨酸(Gly)。我们发现纤细裸藻 CDK11 蛋白中均变成了其他氨基酸残基(图 1-11B)。从上述分析来看,纤细裸藻的剪接体蛋白与人类的剪接体蛋白存在微小但关键的差异,并使纤细裸藻对剪接抑制剂 OTS964 不敏感。

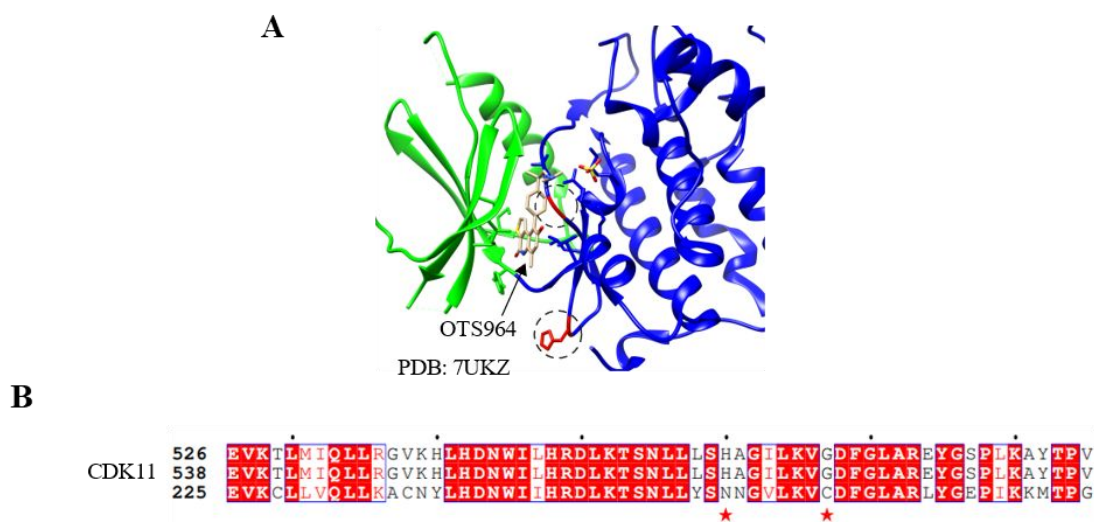


图 1-11 人类与纤细裸藻的 CDK11 蛋白比较

Fig.1-11 Comparison of CDK11 protein between human and *E. gracilis*

注: A 为人类 CDK11B 蛋白与 OTS964 作用三维结构图。OTS964 用箭头表示。OTS964 与 CDK11B 作用的结构域用绿色和蓝色表示。CDK11B 中 572 位组氨酸残基和 579 位甘氨酸残基用虚线圆圈标记; B 为人类与纤细裸藻 CDK11 蛋白的序列比对。上面的两个序列来自人类蛋白质 CDK11A 和 CDK11B, 最下面的序列来自纤细裸藻。影响小分子抑制剂与蛋白相互作用的残基用红星标出。

Note: A is three-dimensional structure of human CDK11B with OTS964. OTS964 is indicated with an arrow. CDK11B domains that contain OTS964 are colored green and blue. CDK11B His⁵⁷² and Gly⁵⁷⁹ are labeled with dashed circles; B is sequence alignments of human and *E. gracilis* CDK11. The two sequences above were from human proteins CDK11A and CDK11B, and the lower sequence was from *E. gracilis*. Residues in proteins that influence the interaction with small molecule inhibitors are indicated with red stars.

4 结论

在本研究中,我们成功鉴定出 166 个纤细裸藻剪接蛋白基因,确定了纤细裸藻中 U2 和 U6 snRNA 的存在。另外我们整理了布氏锥虫的 78 个剪接蛋白和杜氏利什曼原虫的 76 个剪接蛋白,以及 *D. papillatum* 的 121 个剪接蛋白,发现古虫界

生物之间在剪接体组成上具有明显差异。利用四种人类剪接抑制剂处理纤细裸藻细胞，我们发现纤细裸藻对这些人类剪接抑制剂均不敏感。

5 讨论

在本研究中，我们利用二代、三代转录组测序以及结合前人的转录组测序结果进行对比，最终鉴定出了 166 个剪接体蛋白基因以及纤细裸藻中的 U2 和 U6 snRNA。在二代无参转录组与三代全长转录组比较中发现，有 36 种剪接蛋白基因是不存在于三代分析数据中的，但是后续在基因组上仍然能找到这些基因相应的同源序列，大概是这些蛋白的编码基因表达水平太低，导致无法通过三代全长 RNA-seq 检测到。本研究鉴定的纤细裸藻剪接体蛋白基因表明，大部分的 snRNP 及其相关蛋白，以及非 snRNP 的 NTC/NTR/IBC/RES 复合物蛋白在纤细裸藻中是保守的。相比之下，大多数催化阶段的特异性蛋白、SR 蛋白以及 hnRNP 较容易发生进化的改变。此前，Ebenezer 等人通过二代转录组测序数据已经鉴定出了纤细裸藻中的一些剪接体蛋白^[15]，但与我们的分析相比，他们的注释是不完整的，这可能是由于他们的基因组和转录组数据相对较小并且包含一些错误（例如，他们识别的 Prp11/SF3A2 实际上是 SF3A3）。此前估计 SL 介导的反式剪接在纤细裸藻基因中的发生率约为 50%^[22]。本研究中，我们基于我们的全长转录组测序结果评估了这一比率，发现它比之前的估计略高。虽然反式剪接在纤细裸藻中普遍存在，但这种类型的剪接在藻类生物中的意义仍然是未知的。通过我们对纤细裸藻剪接体组分进行完整地鉴定分析，我们明确了这种生物中剪接体的基本特征，为研究纤细裸藻剪接体参与常规内含子顺式剪接和 SL 介导的反式剪接的分子机制奠定了基础。

第二章 纤细裸藻核 pre-mRNA 内含子的鉴定与分析

1 目的

在 pre-mRNA 成熟过程中, 序列中的内含子通过剪接体催化而被移除。剪接体是由 5 个小核 RNA (snRNA) (U1, U2, U4, U5 和 U6) 以及许多相关蛋白组成的大型核糖核蛋白复合物^[5]。在剪接过程中, 内含子内的特定 RNA 元件, 包括 5' 端剪接位点 (5'ss)、3' 端剪接位点 (3'ss)、分支点腺苷序列和多聚嘧啶序列 (Py-tract), 分别被相应的 snRNA 和蛋白质识别^[33]。

典型的剪接位点边界序列是具有 5' GT 和 3' AG 的二核苷酸结构, 这种结构也存在于 SL 介导的反式剪接中。SL 介导的反式剪接是将 spliced leader RNA (SL-RNA) 的外显子添加到 pre-mRNA 上的过程^[34]。在这种反式剪接中, 5' GT 位于 SL-RNA 内, 而 3' AG 位于 pre-mRNA 的 5' 末端内含子区。此外, SL-RNA 在 5'ss 区域内形成的二级结构模拟了 U1 snRNA 与 5'ss 碱基配对, 使得 U1 不需要参与反式剪接^[16,34]。然而, 目前尚不清楚 SL-RNA 参与反式剪接的分子机制。

纤细裸藻 (*Euglena gracilis*), 也称为眼虫, 同时具有顺式剪接和反式剪接这两种剪接形式。初步研究表明, 纤细裸藻中除了常规的 GT-AG 型内含子外, 还含有许多非常规内含子。非常规内含子在一个或两个剪接位点中缺乏 GT/AG 二核苷酸, 但在其末端仍然能够进行碱基互补配对^[16,35]。据推测, 非常规内含子的剪接与常规内含子不同, 常规内含子形成的是套索状 RNA 结构, 而非常规内含子可能会形成环状 RNA^[36]。最近的一项研究中只发现了几百个内含子, 且主要是常规内含子^[15]。目前尚缺乏对纤细裸藻中常规内含子和非常规内含子的分布和性质的全面研究。因此在本研究中, 我们计划利用转录组学和基因组学数据鉴定纤细裸藻的内含子, 进一步揭示了纤细裸藻内含子的独特特征。

2 材料与方法

2.1 细胞株

实验所用纤细裸藻细胞株与第一章相同

2.2 实验方法

2.2.1 内含子的筛选与识别

我们利用纤细裸藻第三代转录组测序(RNA-seq)数据,进行纤细裸藻内含子的鉴定分析。首先,使用 BLAST 程序^[37]将三代 RNA-seq 数据分别与叶绿体^[38]、线粒体^[39]和核糖体 DNA (rDNA)^[40]进行了序列比较。我们将任何符合这些类别的序列都从 RNA-seq 数据中删除,然后使用 cd-hit 消除冗余序列^[41],从而获得细胞核 mRNA 序列。然后,使用 exonerate 程序^[42]将 mRNA 序列与纤细裸藻基因组序列进行比对。通过调整 A/C/G/T 之间的比例(5'ss 的 A/C/G/T 频率为 10/10/50/30,3'ss 的 A/C/G/T 频率为 30/10/50/10),以获得精确比对的外显子序列,随后从基因组序列中通过 exonerate 程序提取内含子。最后,对提取到内含子进行手动筛选并分类,得到 GT-AG、GC-AG、ss-invariant 和 ss-variant 四种类型。

为了鉴定纤细裸藻的 outtron 序列,我们从上述 mRNA 序列中提取出含有 SL 外显子的基因的前 200 bp 序列。用二核苷酸 AG 替换纤细裸藻 SL 外显子后,再次与基因组序列进行 BLAST 比对。使用 Seqkit 提取到基因组上的 outtron 序列^[43]。同样的方法也应用于其它含有 SL 介导的反式剪接的物种,包括秀丽隐杆线虫(*C. elegans*)、布氏锥虫(*T. brucei*)、杜氏利什曼原虫(*L. donovani*)的 outtron 鉴定,利用它们各自的 SL 外显子、RNA-seq 数据(SRA 序列编号:*C. elegans* 为 SRR26536558-73、*T. brucei* 为 SRR15913764-72、*L. donovani* 为 SRR5272520-7)和各自对应的基因组进行鉴定。

2.2.2 内含子序列的分析与处理

使用 Seqkit 对生物序列进行快速处理和分析,包括提取子序列、序列信息统计以及去除冗余序列^[43]。使用 Weblogo3 对内含子末端剪接位点区域的核苷酸相对频率进行分析^[44]。使用 RNAfold 进行 RNA 二级结构分析^[45]。使用 TRF 进行串联重复分析^[46]。使用 R 语言进行序列作图分析。使用 PRISM 7 对串联的内含子数量和碱基含量进行作图分析。

3 结果与分析

3.1 纤细裸藻细胞核 pre-mRNA 内含子的鉴定

我们前期工作进行了细胞总 RNA 的第三代转录组测序(RNA-seq)^[47]。对于细

胞核内含子的鉴定，叶绿体、线粒体基因组以及细胞核 rDNA 基因序列都属于干扰序列，应当被移除，只留下细胞核 pre-mRNA 或 mRNA 基因的序列。我们利用 cd-hit 去除冗余序列后，共获得 18643 个序列，以此作为内含子鉴定的起点。

我们利用 exonerate 程序对 mRNA 序列和基因组数据进行比对，随后提取鉴定出内含子序列(图 2-1)。我们发现，exonerate 程序中位置特异性打分矩阵 (Position-Specific Scoring Matrix, PSSM) 的默认设置在鉴定常规的 GT-AG 型内含子时是有效的，但对非常规内含子不适用。为了克服这个问题，我们调整了 PSSM 的参数，使 PSSM 可以准确地注释非常规内含子。然而，我们发现尽管绝大多数常规内含子被正确识别，但仍有极少数发生注释错误。为此，我们手动检查了鉴定出的所有内含子。

本研究共鉴定出 28337 个内含子。我们将已鉴定出的内含子进一步分为四种类型：GT-AG、GC-AG、ss-invariant 和 ss-variant (图 2-1)。GT-AG 型是传统剪接类型，而 GC-AG 型则是其变体，其余两种内含子的末端为既不是 GT-AG 也不是 GC-AG 的二核苷酸。此外，ss-variant 类型内含子是指内含子的末端核苷酸具有多种可能性，并且不能明确确定(图 2-2)。由于在内含子的一个或两个末端都没有典型的 GT/GC-AG 二核苷酸，因此在本研究中 ss-invariant 和 ss-variant 型内含子均被归类为非常规内含子。

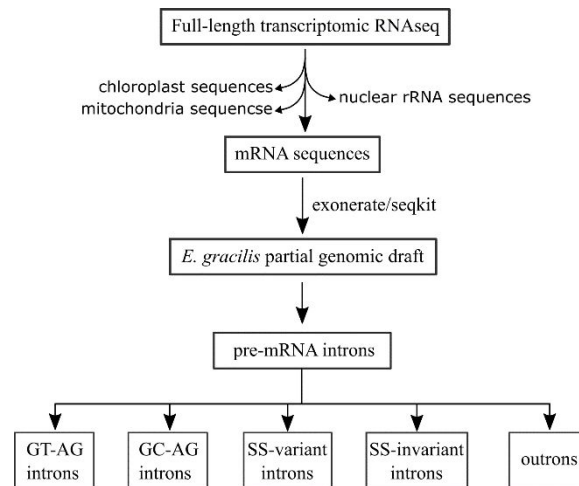


图 2-1 纤细裸藻细胞核 pre-mRNA 内含子的鉴定工作步骤

Fig.2-1 Working procedure for identification of nuclear pre-mRNA introns in *E. gracilis*

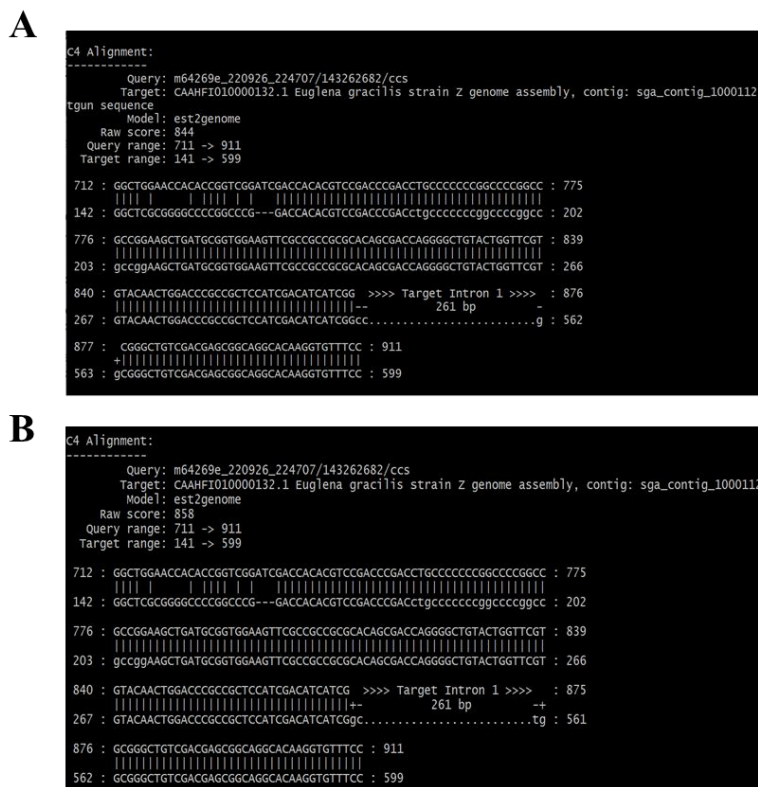


图 2-2 ss-variant 型内含子示意图

Fig.2-2 Example of ss-variant intron

注:Exonerate 程序注释的末端核苷酸。A 为 cc-gg; B 为 gc-tg。

Note: Terminal nucleotides annotated by the exonerate program. A is cc-gg; B is gc-tg.

我们发现, 四种类型内含子的长度分布相似, 最小长度在 50 bp 以下, 最大长度在 4682 bp-8197 bp 之间。值得注意的是, 在 GT-AG 型中, 序列长度小于 60 bp 的短内含子出现了显著的峰值(图 2-3)。短内含子的在每种类型中比例为:GT-AG 为 9.82%(1332/11527), GC-AG 为 1.37%(13/659), ss-invariant 为 0.89%(40/4498), ss-variant 为 1.25%(146/11653)。这表明短内含子在剪接时倾向于采用传统的边界结构。此外, 属于非常规内含子的 ss-invariant 和 ss-variant 的分布曲线高度相似, 表明这两种内含子之间可能存在共同特征。

除上述顺式剪接内含子外, 我们还鉴定出了 11921 个 outtron 序列。由于这些 outtron 的转录起始位点目前无法获得, 因此无法确定它们的长度。在接下来的分析中, 我们只分析 3'端长度仅为 20 bp 的 outtron 序列。

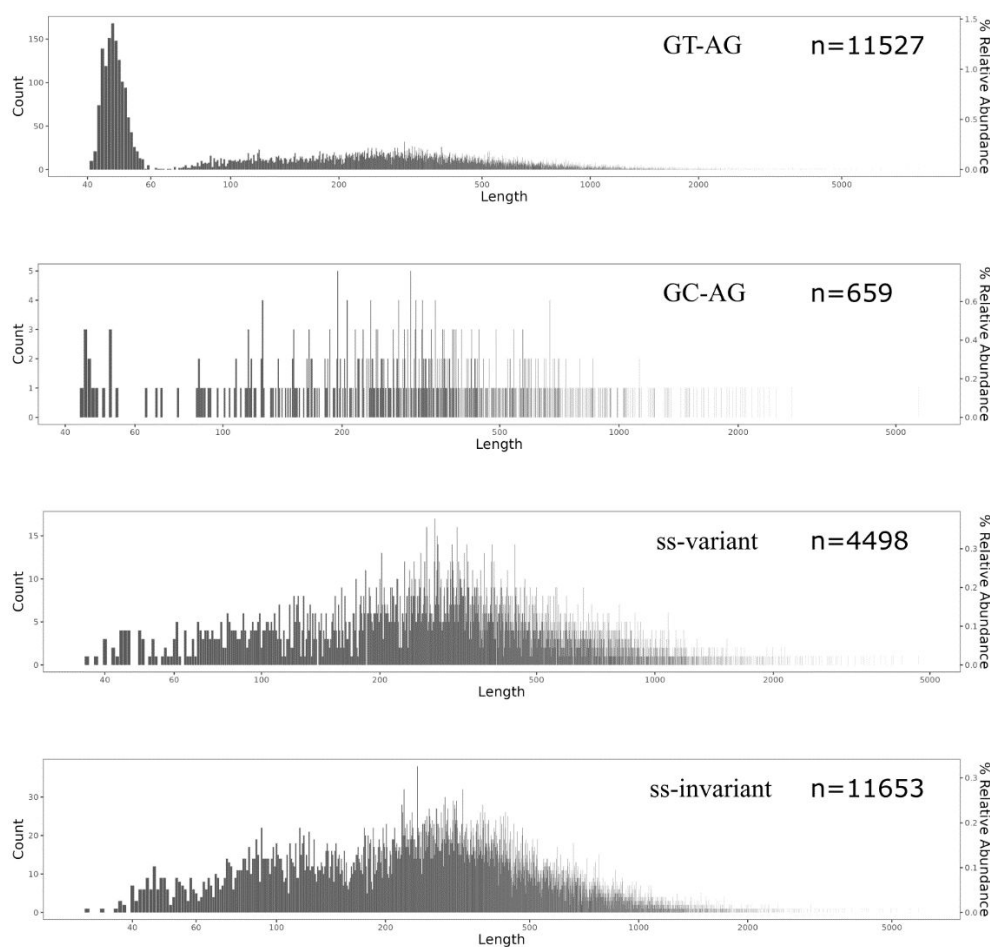


图 2-3 纤细裸藻细胞核 pre-mRNA 内含子的长度分布

Fig.2-3 Length distribution of nuclear pre-mRNA introns in *E. gracilis*

3.2 细胞核 pre-mRNA 常规内含子 5'ss 区域的特征分析

我们首先研究了 GT-AG 和 GC-AG 这两种常规内含子中 5'ss 区域 10 bp 长度的保守性。GT-AG 型内含子的一致性序列为 GTGTG(图 2-4A)。在 GC-AG 型内含子中，只有第 5 位的鸟嘌呤(G)是较为保守的(图 2-4A)。为了与其它物种比较，我们从 IAOD 数据库中检索了人类、拟南芥、裂殖酵母以及酿酒酵母的 U2 内含子，并分析了它们的 5'ss 序列(图 2-4B)。我们观察到，位于第 5 位的 G 在人类和酵母中也是稳定存在。但是，这个 G 在拟南芥中并不稳定。

3.3 Extended U6/5'ss 螺旋的形成

在纤细裸藻 SL-RNA 的 5'ss 区域存在广泛的内部碱基配对，因此我们研究了 SL-RNA 与 snRNA 之间的碱基配对结构。将 SL-RNA 的 5'ss 与 U6 比对，我们观察到 5'ss 的下游区域与 U6 的 ACAGA 序列上游区域之间存在重要的碱基配对(图 2-6)。这种 U6/SL-RNA 螺旋区也被称为 extended U6/5'ss 螺旋，(extended U6/5'ss)，首先发现于人类剪接体 pre-B 复合体的冷冻电镜结构中^[48]。人类中该螺旋内形成 5 个碱基对，然而，纤细裸藻 extended U6/SL-RNA 螺旋由 8 个连续的碱基对组成。



图 2-6 纤细裸藻 U6 ACAGA 上游序列与 SL-RNA 5'ss 下游序列之间存在广泛的碱基配对

Fig.2-6 Extensive base-pairing between sequences upstream of the U6 ACAGA box and downstream of the 5'ss of SL-RNA in *E. gracilis*

注：下划线标记的是 U6 的 ACAGA 区域和 SL-RNA 5'ss 区域。所有序列都以 0-2 位的垂直比例绘制。

Note: The ACAGA box of U6 and the 5'ss of SL-RNA are underlined. All sequence logos are plotted on a vertical scale of 0-2 bits of information.

我们进一步研究了 extended U6/SL-RNA 螺旋在其他进行 SL 介导的反式剪接的物种中的情况(图 2-7A)。对 *C. elegans* U6 和 SL-RNA 的比对发现其也形成了 8 个连续的碱基对。此外，*T. brucei* 和 *L. donovani* 的 U6 和 SL-RNA 也在该区域形成广泛的碱基配对。综上所述，在 SL 介导的反式剪接中，extended U6/5'ss 螺旋的形成是普遍的。

考虑到 pre-mRNA 内含子在 5'ss 下游表现出不同的碱基组成，我们想知道 U6 和某些内含子之间是否也存在 extended U6/5'ss 螺旋。因此，我们通过分析人类、拟南芥、裂殖酵母以及酿酒酵母的 5'ss 下游 10 bp 区域，连同各自的 U6 序列，我们发现这些物种内含子可能会形成 extended U6/5'ss 螺旋(图 2-7B)。例如，酿酒酵母的 Mei4 内含子可以在该区域与 U6 形成 5 个碱基对，而拟南芥 AT2G30650 基因的第二个内含子可以与 U6 形成 8 个碱基对。这一发现表明，extended U6/5'ss 螺旋的形成取决于内含子的序列。

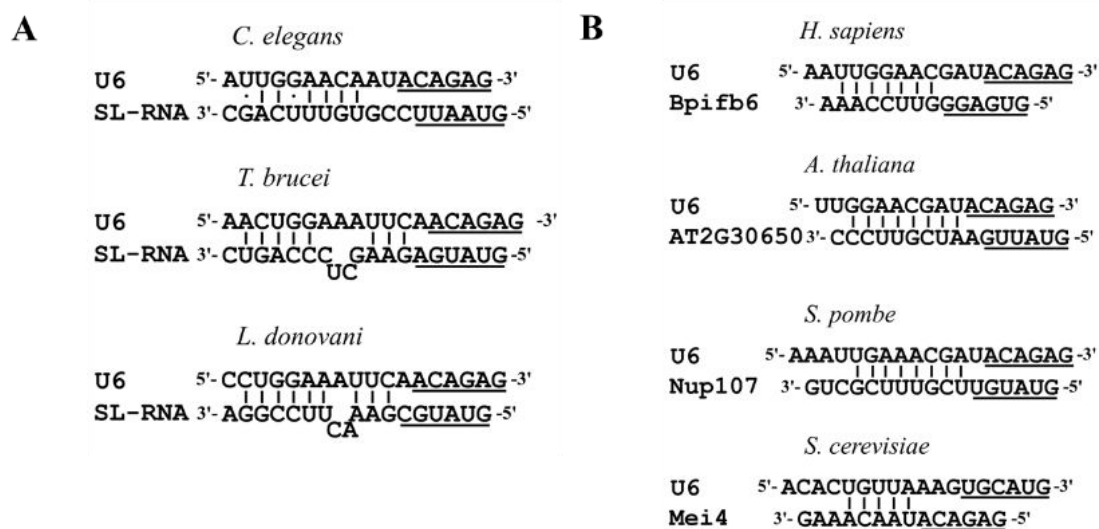


图 2-7 不同物种的 extended U6/5'ss 螺旋

Fig.2-7 The extended U6/5'ss helix in different species

注：A 为秀丽隐杆线虫、布氏锥虫和杜氏利什曼原虫 SL-RNA 的 extended U6/5'ss 螺旋；B 为人、拟南芥、裂殖酵母和酿酒酵母的基因 U6 和 5'ss 区域之间的 extended U6/5'ss 螺旋。下划线标记的是 U6 的 ACAGA 区域和 SL-RNA 5'ss 区域。所有序列都以 0-2 位的垂直比例绘制。

Note: A is the extended U6/5'ss helix of SL-RNA in *C. elegans*, *T. brucei*, and *L. donovani*; B is examples of the extended U6/5'ss helix between the respective U6 and the 5'ss region of genes from human, *A. thaliana*, *S. pombe*, and *S. cerevisiae*. The ACAGA box of U6 and the 5'ss of SL-RNA/intron are underlined. All sequence logos are plotted on a vertical scale of 0-2 bits of information.

3.4 细胞核 pre-mRNA 常规内含子 3'ss 区域的特征分析

我们分析了在纤细裸藻中 GT-AG 型和 GC-AG 型内含子 3'ss 长度 20 bp 区域的序列保守性(图 2-8A)。结果显示这两种内含子形式在这一区域表现相似。值得注意的是，在人类、拟南芥、裂殖酵母以及酿酒酵母等物种的 3'ss 区域我们观察到主要是胸苷(T)的富集组成(图 2-8B)，而纤细裸藻中相同的 3'ss 区域则是丰富的胞苷(C)。

我们进一步检查了所鉴定到的纤细裸藻 outtron 序列 3'ss 区域的序列保守性。我们发现前面提到的顺式剪接内含子的组成，包括富含 C 的多聚胞苷序列和 3'ss 共同序列 CAG，这些特征仍然适用于纤细裸藻的 outtron(图 2-8C)。为了比较 outtron 的组成形式，我们通过转录组和基因组序列分析，从其他具有 SL 介导的反

式剪接的物种中确定了相应的 outtron 的 3'ss 区域。我们分别从 *C. elegans*、*T. brucei* 以及 *L. donovani* 中获得了 4535、976 和 12030 个 outtron 序列。不同物种的 outtron 表现出明显不同的特征。例如，*C. elegans* 具有一个富含腺苷(A)/胸苷(T)的多聚嘧啶序列和一个独特的 3'ss 识别序列 TTTCAG，这在以前被报道过^[49]。在 *T. brucei* 和 *L. donovani* 中，多聚嘧啶区域也有明显的差异，*T. brucei* 的胸苷(T)含量略高于 *L. donovani*。

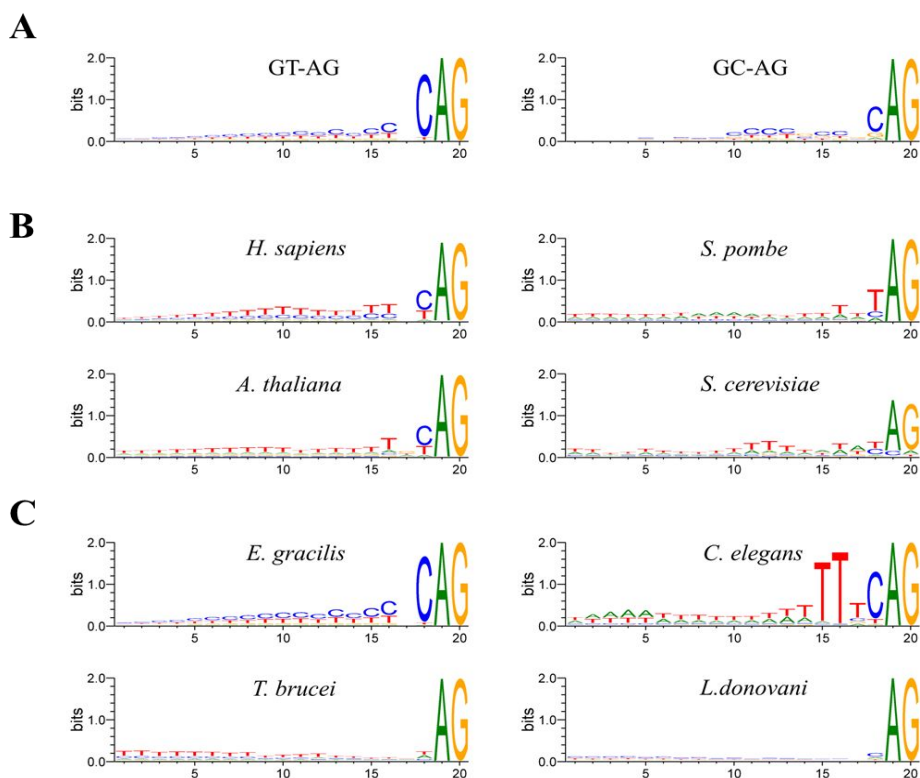


图 2-8 常规内含子 3'ss 区域的组成与序列保守性

Fig.2-8 Sequence conservation and composition of the 3'ss region of conventional introns

注：A 为纤毛裸藻 GT-AG 和 GC-AG 内含子 3'ss 区域的序列保守性；B 为人类、拟南芥、裂殖酵母以及酿酒酵母内含子 3'ss 区域的序列保守性；C 是纤毛裸藻、秀丽隐杆线虫、布氏锥虫、杜氏利什曼原虫 outtron 序列 3'ss 区域的序列保守性。所有序列都以 0-2 位的垂直比例绘制。

Note: A is sequence conservation of the *E. gracilis* 3'ss regions of GT-AG and GC-AG introns; B is sequence conservation of the 3'ss regions of introns in human, *A. thaliana*, *S. pombe*, and *S. cerevisiae*; C is sequence conservation of the 3'ss regions of outtrons in *E. gracilis*, *C. elegans*, *T. brucei*, and *L. donovani*. All sequence logos are plotted on a vertical scale of 0-2 bits of information.

为了直观获得 3'ss 区域中 T/C 碱基分布，我们绘制了这两个碱基在上述物种

的内含子中的比率(图 2-9)。结果表明,大部分纤细裸藻内含子(包括 GT-AG 型和 outtron)是富 C 的,而在其他物种的内含子中,除了 *L. donovani* 外,主要是富 T 的。在 *L. donovani* 中,大多数 outtron 只表现出非常弱的富 C 组成倾向。综上所述,这些发现表明纤细裸藻在多聚嘧啶序列中具有独特的碱基组成。

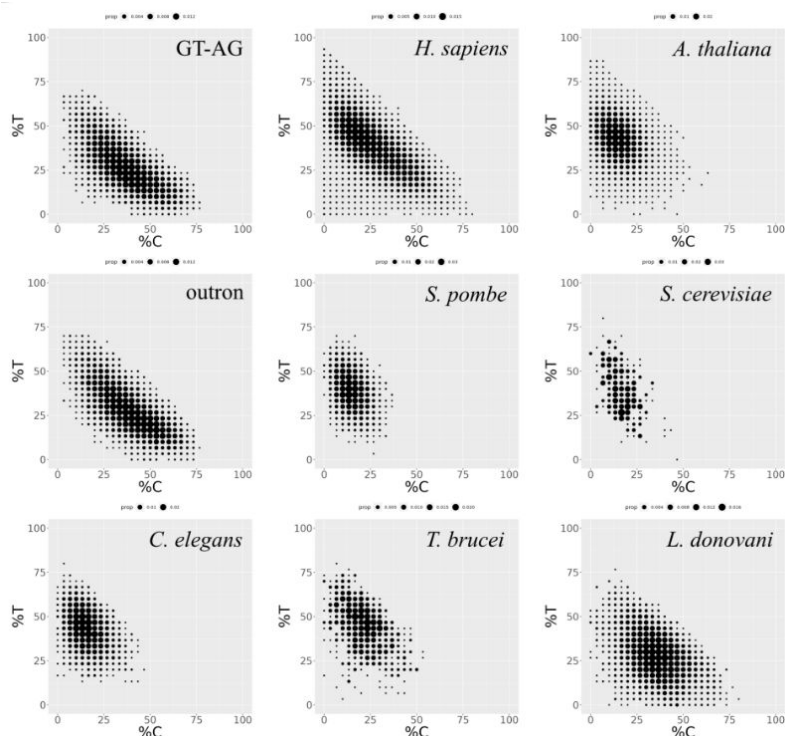


图 2-9 纤细裸藻、人类、拟南芥、裂殖酵母、酿酒酵母、秀丽隐杆线虫、布氏锥虫和杜氏利什曼原虫的内含子和 outtron 序列 3'ss 区 C/T 比率

Fig.2-9 Ratios of C/T in the 3'ss regions of introns and/or outtrons in *E. gracilis*, human, *A. thaliana*, *S. pombe*, *S. cerevisiae*, *C. elegans*, *T. brucei*, and *L. donovani*

注:所有序列都以 0-2 位的垂直比例绘制。

Note: All sequence logos are plotted on a vertical scale of 0-2 bits of information.

3.5 细胞核 pre-mRNA 非常规内含子末端碱基配对

如上所述, ss-invariant 和 ss-variant 这两种类型的内含子被认为是非常规内含子,且具有相似的长度分布。首先,我们研究了 ss-invariant 类型内含子的 5'ss 长度 10 bp 区域。我们在这类内含子中发现了一个横跨+3 到+7 位置的 CCAGG 序列(图 2-10A)。此外,我们从 ss-invariant 型内含子的 5'端提取含有常规 GT 边界的序列,发现在相同位置仍然具有类似的碱基(图 2-10A)。接下来继续研究 3'ss 长

度 10 bp 区域, 我们发现了跨越-9 到-6 位置的 CCTG 序列(图 2-10B)。这一序列仍然存在于 ss-invariant 型内含子的 3'端含有常规 AG 边界区域中。

最后, 我们研究了 ss-variant 类型内含子(图 2-10C)。在 5'ss 区域, 该类型内含子的序列保守性与上述 ss-invariant 型内含子相似。然而, 在 3'ss 区域的序列保守性上, 两者有细微的差异。值得注意的是, 由于 ss-variant 型内含子的末端核苷酸的模糊性, 这组内含子的末端序列保守性无法准确判定。

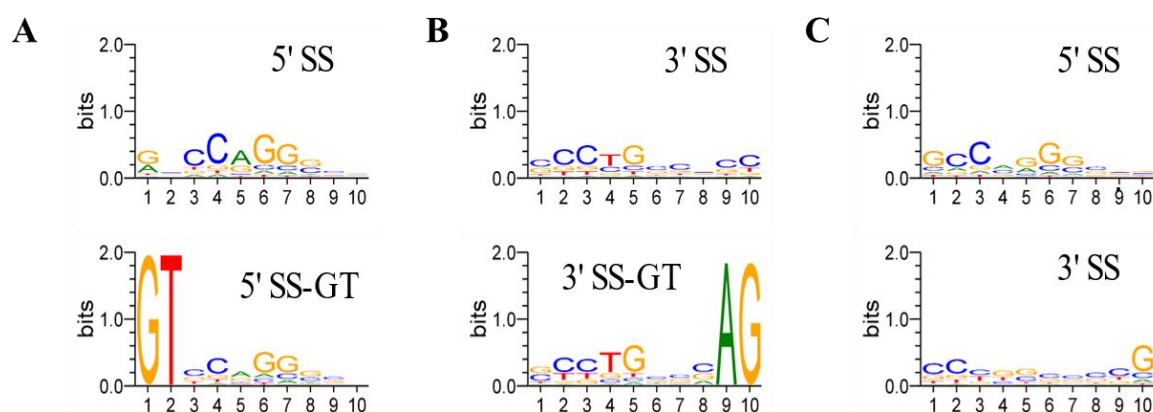


图 2-10 纤细裸藻非常规内含子 ss 区的序列保守性与碱基配对

Fig.2-10 Sequence conservation and base-pairing of ss region of the *E. gracilis* nonconventional introns

注: A 为 ss-invariant 非常规内含子的 5'ss 区序列保守性; B 为 ss-invariant 非常规内含子的 3'ss 区序列保守性; C 为 ss-variant 非常规内含子的 5'ss 和 3'ss 区域的序列保守性。所有序列都以 0-2 位的垂直比例绘制。

Note: A is sequence conservation of the 5'ss region of nonconventional ss-invariant introns; B is sequence conservation of the 3'ss region of nonconventional ss-invariant introns; C is sequence conservation of the 5' and 3'ss regions of nonconventional ss-variant introns. All sequence logos are plotted on a vertical scale of 0-2 bits of information.

非常规内含子 5'以及 3'末端的序列表明, 两端之间可能存在碱基配对。为了研究两种非常规内含子末端碱基配对的潜在差异, 并将两者与常规 GT-AG 型内含子进行比较, 我们使用 RNAfold 检测了在 5'与 3'末端之间形成的 RNA 二级结构。如图 2-11A 所示, 两末端 10 bp 长度之间是人为添加 7 个尿苷(U)的环状结构, 得到长度为 27 bp 的嵌合序列。该嵌合序列之间的碱基匹配程度由热值反映(图 2-11A)。通过对这三类内含子的热值和频率的分析, 我们发现两种非常规内含子表

现出相似的模式，大多数内含子在 $-10 \sim -2$ kcal/mol 的能量范围内表现出末端碱基配对。而 GT-AG 型内含子则大致呈线性分布，没有明显的峰(图 2-11B)。

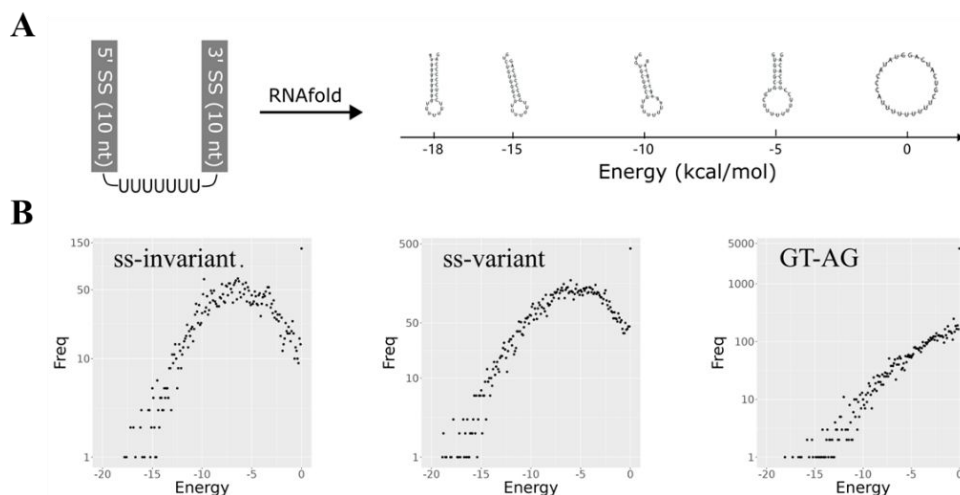


图 2-11 5'ss 与 3'ss 末端之间形成的 RNA 二级结构

Fig.2-11 The RNA secondary structure formed between the 5' and 3'ss terminal regions

注：A 为 RNAfold 产生的末端碱基配对示意图以及不同碱基配对级别所对应的热值；B 是纤细裸藻中 ss-invariant、ss-variant 和 GT-AG 内含子末端碱基配对的热值和频率图。所有序列都以 0-2 位的垂直比例绘制。

Note: A is illustration of terminal base-pairing generated by RNAfold and the energies associated with different levels of base-pairing; B is plots of energy and frequency of terminal base-pairing of ss-invariant, ss-variant and GT-AG introns in *E. gracilis*. All sequence logos are plotted on a vertical scale of 0-2 bits of information.

3.6 纤细裸藻细胞核 pre-mRNA 内含子的选择性剪接

基于上述鉴定的纤细裸藻细胞核 pre-mRNA 内含子，我们对纤细裸藻细胞核基因的选择性剪接 (Alternative Splicing, AS) 进行了探讨。为此，我们对鉴定得到的 28337 个内含子与 18643 个 mRNA 序列进行 BLAST 比对。我们总共发现了 9 个可变 5'ss (A5SS)，9 个可变 3'ss (A3SS)，11 个外显子跳跃 (SE)，以及 161 个内含子滞留 (RI) (图 2-12)。值得注意的是，在 11 个 SE 序列中，我们只发现了一个典型的 SE 序列（即具有相同的末端剪接位点），其余 10 个 SE 序列，我们称之为非典型 SE 序列，它们具有不同的末端 ss 位置(图 2-12)。在这 10 个非典型 SE 序列中，只有 3 个具有相同的 5'ss 位置，且所有序列均不具有相同的 3'ss 位置。除

了上述类型的 AS，我们观察到某些转录本具有重叠的内含子区域，但这些内含子的末端 ss 位置并不完全不同。我们将这种类型的 AS 称为重叠(图 2-12)。我们确定了总共 59 个重叠事件（见附录二）。

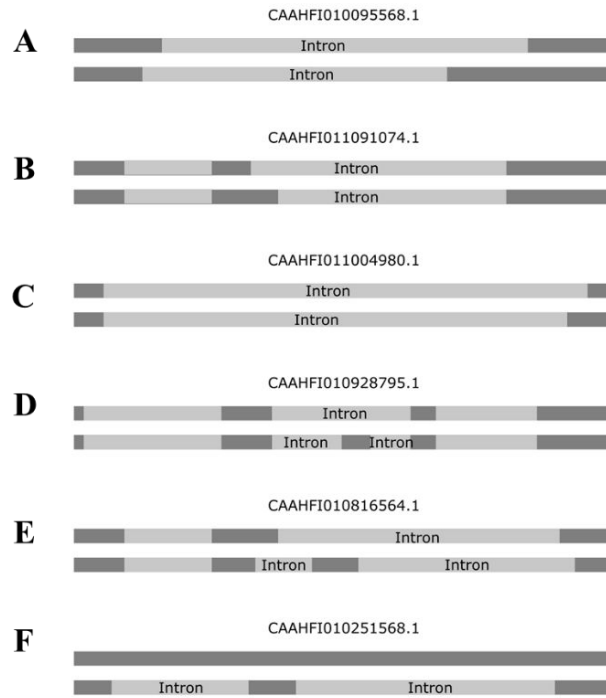


图 2-12 纤细裸藻细胞核前体 mRNA 内含子的选择性剪接形式

Fig.2-12 Alternative Splicing patterns of the *E. gracilis* nuclear pre-mRNA introns

注：示意图为重叠形式(A)、A5SS形式(B)、A3SS形式(C)、SE形式(D)、非典型SE形式(E)和内含子滞留形式(F)。外显子用深灰色表示，内含子用浅灰色表示。每种形式的基因组编号显示在转录组上方，内含子被标记在转录组上。

Note: Schematic examples of overlapping(A), A5SS(B), A3SS(C), SE(D), atypical SE(E), and RI(F) are shown. Exons are depicted in dark gray and introns in light gray. The genomic accessions for each pattern are indicated above the transcripts and introns are labelled on the transcripts.

3.7 纤细裸藻细胞核 pre-mRNA 内含子串联重复序列

据报道，纤细裸藻基因组含有大量重复序列，一些纤细裸藻基因的内含子也含有串联重复序列(TRs)。例如，纤细裸藻的 γ -tubulin 基因在第 11 个内含子上有一个 88 bp 的 GT/GC 串联重复序列以及在第 13 个内含子上有一个 482 bp 的 AT/AC 串联重复序列。在研究所鉴定出的内含子序列时，我们也观察到了这样的重复序列。

为了对纤细裸藻内含子中的 TRs 进行全面的分析，我们将 GT-AG、ss-invariant 和 ss-variant 三种类型内含子以及人类内含子通过 TRF 程序进行检测。随后，我们分析了这些 TR 的数量和长度之间的关系。结果表明，这些内含子都表现出相似的模式，大部分都为短重复序列，很少具有长重复(图 2-13)。

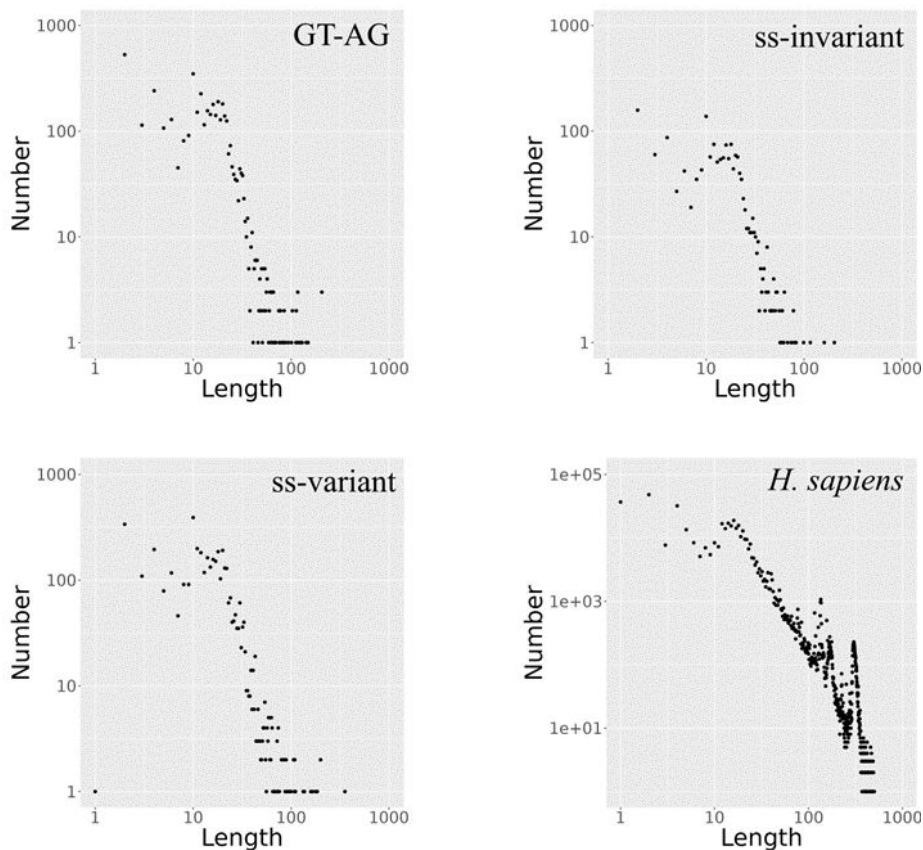


图 2-13 研究了纤细裸藻的 GT-AG、ss-invariant 和 ss-variant 中串联重复序列长度和数目的分布。包括人类内含子中串联重复序列的分布以供比较

Fig.2-13 The distribution of TRs lengths and numbers in GT-AG, ss-invariant and ss-variant introns of *E. gracilis*. The distribution of TRs in human introns is included for comparison

此外，我们计算了 TRs 内含子数与碱基含量的比率(图 2-14)。我们发现人类内含子中串联重复序列的比例要高于纤细裸藻，这可能归因于人类内含子比纤细裸藻要长。关于 TRs 碱基含量的比较，我们发现在纤细裸藻中这一值为 3.22%，略高于人类的 2.26%(图 2-14)。总的来说，TRs 的结果表明纤细裸藻内含子与人类内含子在 TR 分布上具有相似性。

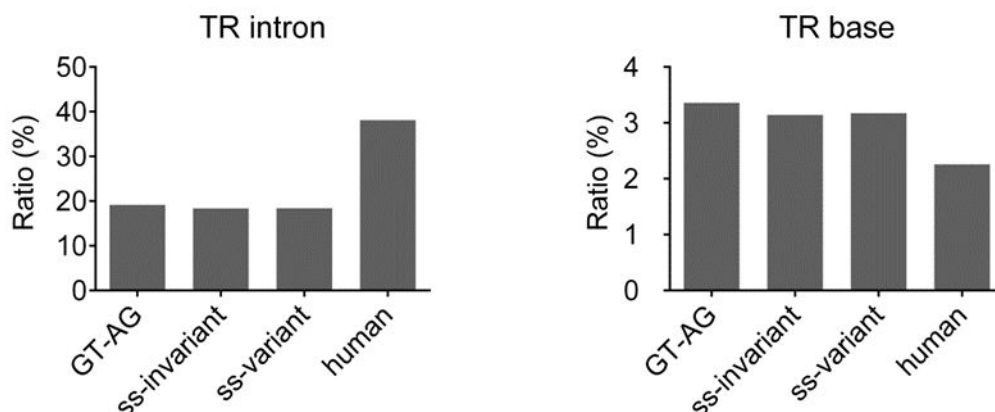


图 2-14 含串联重复序列的内含子与碱基含量的比率

Fig.2-14 Ratios of TRs-containing introns and TRs-containing bases

4 结论

我们成功地鉴定了 28337 个常规和非常规内含子，并将其分为 GT-AG、GC-AG、ss-invariant 和 ss-variant 四种类型。我们发现纤细裸藻 SL-RNA 的 5'ss 区域中存在与人类相似的 extended U6/5'ss 螺旋，以及在纤细裸藻的 3'ss 区域富含 C，这与人类、拟南芥等富含 T 不同。本研究揭示了纤细裸藻剪接的独特性，有助于揭示新的剪接机制，理解裸藻的内含子进化和其基因组学。

5 讨论

在本项研究中，我们利用转录组和基因组数据对纤细裸藻细胞核 pre-mRNA 进行了大规模的生物信息学分析。纤细裸藻细胞核 pre-mRNA 内含子以两种形式存在：位于 pre-mRNA 分子内部的顺式内含子和占据分子 5'端的 outtron。我们共鉴定出 28337 个顺式内含子和 11921 个 outtron 序列。但是，由于缺乏转录起始位点，我们鉴定的 outtron 是不完整的，只含有 3'末端区域。此外，我们鉴定到的内含子数量可能也是不完整的，因为我们研究使用了不完整的基因组^[15]。

我们对纤细裸藻含有的内含子的序列保守性分析表明，传统的 GT/GC-AG 顺式内含子具有 5'端 GTGTG 和 3'端 CAG 这两个一致性序列，以及独特的富含 C 的聚胞苷序列。传统的 5'ss 的序列与 SL-RNA 不同，它们确保了顺式和反式剪接的有效进行。传统的顺式内含子和 outtron 在 3'ss 区域上的高度相似性表明它们的剪

接反应都依赖于剪接体。值得注意的是，纤细裸藻常规内含子富含 C 的多聚胞苷序列的存在，表明结合多聚嘧啶序列的剪接因子 U2AF2 在纤细裸藻中可能具有新的识别模式。此外，我们对非常规内含子的分析表明，尽管在剪接位点末端没有一致的保守序列，但这些内含子依然会进行互补配对，通常能够形成较强的末端碱基配对。这一发现与之前对少量的核基因的研究是一致的^[16, 35]。

在纤细裸藻和其他物种中发现了 SL-RNA 和 U6 之间的 extended U6/5'ss 螺旋，这表明 U6 对 SL-RNA 的识别在剪接中至关重要。这个 extended U6/5'ss 螺旋结构首先在人类剪接体 B 复合体中被发现，但酵母剪接体结构中却不存在，因此有研究提出它是后生动物所特有的^[48]。在具有 SL 介导的反式剪接的物种中这种螺旋结构的存在，表明它可能是 SL 介导的反式剪接所必需的。此外，对人类、拟南芥、以及芽殖酵母的研究表明，这些物种中也有少数内含子可以形成这种螺旋结构。因此，我们认为只要内含子有能力与 U6 形成这一结构，这种螺旋就可能出现于相应的物种中。在剪接体 B 复合体中，这个螺旋周围存在三个蛋白^[50]。在这些蛋白中，Prp38 和 ZMAT2/Snu23 在纤细裸藻中有相应的同源蛋白^[47]。我们认为这些蛋白可能与纤细裸藻中的这一螺旋相互作用，类似于它们在人类剪接体中的作用。

纤细裸藻内含子的另一个独特特征是它的选择性剪接(AS)。纤细裸藻内含子表现出一种称为重叠的特殊类型，其中两个内含子共享相同的基因组位置，但其末端序列的位置不同。即使是非典型的外显子跳跃序列(SE)，在一定程度上也可以看作是重叠形式的一种组合。由于目前纤细裸藻全基因组尚不完整，因此还不清楚纤细裸藻中选择性剪接的范围和数量。此外，尚不清楚这种选择性剪接的重叠类型是否存在于其他物种中。

第三章 纤细裸藻 Dbr1 蛋白抑制剂的筛选及功能验证

1 目的

在真核细胞中，mRNA 转录后需要经过进一步加工成熟才能进行蛋白质翻译。在加工过程中最重要的一环就是 pre-mRNA 剪接。在 pre-mRNA 内含子的剪接过程中，切除下来的内含子 5'剪接位点与分支位点腺苷会形成一个 2'-5'磷酸二酯键的套索状结构。套索 RNA 为剪接的副产物，而真核细胞内存在特定的 RNA 脱支酶 Dbr1，能够催化套索 RNA 线性化而使其降解^[51]。Dbr1 蛋白在结构上包括一个 N 端的金属磷酸酯酶(MPE)和一个独特的 C 端结构域(CTD)，其 MPE 具有一段独有的 LRL (lariat recognition loop) 元件。LRL 元件作为识别模块可与含有分支点腺苷的单链 RNA 的 3'端碱基结合。LRL 和 CTD 的接触则进一步稳定了套索 RNA 与 LRL 之间的相互作用^[52]，从而将套索 RNA 转变成线性，随后进行降解。

Dbr1 的突变或活性降低都能够使套索 RNA 在体内大量积累，进而干扰细胞生长。研究发现，在小鼠和拟南芥中，Dbr1 突变体都是胚胎致死性的^[53,54]。此外，Dbr1 与 HIV 和 HSV 病毒侵染、肌萎缩侧索硬化症及癌症的发生也有关联^[53]。这些研究表明 Dbr1 与疾病关系密切。目前为止，Dbr1 蛋白的小分子抑制剂还未有报道。近年虚拟筛选技术蓬勃发展，利用 AlphaFold 或 Robetta 等人工智能程序预测蛋白质三维结构，并进行基于三维结构的大规模小分子虚拟筛选，从而鉴定出能够与目标蛋白进行结合的小分子化合物。因此，我们计划通过虚拟筛选技术获得能够抑制纤细裸藻 RNA 脱支酶 Dbr1 活性的小分子抑制剂，为实验鉴定纤细裸藻内含子种类及数量奠定基础。

2 材料与方法

2.1 细胞株

实验所用纤细裸藻细胞株与第一章相同

2.2 主要试剂

表 3-1 试剂名称及来源

Table 3-1 Reagent name and source

试剂名称	生产公司
麦芽糖	生工, 中国
Amylose Resin	NEB, 美国
2x ChamQ Universal SYBR qPCR Master Mix	诺唯赞, 中国
HiScript-TS 5'/3'RACE Kit	诺唯赞, 中国
Trizol	生工, 中国
DEPC	生工, 中国
氯仿	科隆, 中国
异丙醇	科隆, 中国
75%乙醇	科隆, 中国
FastKing cDNA 第一链合成试剂盒(去基因组)	天根, 中国
Phanta Super-Fidelity DNA Polymerase	诺唯赞, 中国
质粒小提试剂盒	诺唯赞, 中国
凝胶回收试剂盒	诺唯赞, 中国
Trelief™ 5α Chemically Competent Cell	擎科, 中国
琼脂粉末	擎科, 中国
异丙基-β-D-硫代半乳糖苷(IPTG)	生工, 中国
考马斯亮蓝	生工, 中国
核酸染料	生工, 中国
氨苄青霉素钠	生工, 中国
苯甲基磺酰氟(PMSF)	源叶生物, 中国
限制性内切酶	NEB, 美国
DNA 分子量标准 Marker D(100~2000 bp)	生工, 中国
非预染蛋白 Marker	雅酶, 中国
ClonExpress II One Step cloning Ki	诺唯赞, 中国

2.3 主要仪器及耗材

表 3-2 仪器设备名称及来源

Table 3-2 Reagent name and source of instruments and equipment

仪器名称	生产厂家
实时荧光定量 PCR 仪	Bio-Rad, 美国
漩涡混合器 XW-80A	其林贝尔, 中国
高速强阴离子交换层析	博奥龙, 中国
AKTAprius plus	GE, 美国
移液器	Eppendorf, 德国
落地式低速冷冻离心机	Thermo, 美国
落地式高速冷冻离心机	Thermo, 美国

续表 3-2

仪器名称	生产厂家
ND2000 微量分光光度计	Thermo, 美国
纯水仪	优普, 中国
微波炉	Midea, 中国
恒温培养箱	上海智城, 中国
电泳仪	Bio-Rad, 美国
PCR 仪	Bio-Rad, 美国
凝胶成像仪	Bio-Rad, 美国
层析柜 SL-II	松源华兴, 中国
超净工作台	苏州净化, 中国
超声波细胞粉碎机	宁波新芝, 中国
恒温金属浴	杭州博日, 中国
恒温培养振荡箱	上海智城, 中国

2.4 常规实验试剂配置

表 3-3 LB 培养基

Table 3-3 LB Culture medium

试剂名称	用量
酵母提取物	5 g
胰蛋白胨	10 g
氯化钠	10 g
去离子水定容至	1 L

表 3-4 Buffer A

Table 3-4 Buffer A

试剂名称	用量
1 M Tris-HCl(pH8.0)	10 mL
5 M NaCl	2.5 mL
去离子水定容至	500 mL

表 3-5 分离胶(10%)

Table 3-5 10% separating gel

试剂名称	用量
分离液	3.25 mL
30%丙烯酰胺	1.66 mL
10%APS	0.04 mL
TEMED	0.005 mL
去离子水定容至	5 mL

表 3-6 分离液

Table 3-6 Separating solution

试剂名称	用量
1 M Tris-HCl(pH8.8)	289 mL
10%SDS	8 mL
去离子水定容至	500 mL

表 3-7 浓缩胶 (5%)

Table 3-7 5% stacking gel

试剂名称	用量
浓缩液	2.05 mL
30%丙烯酰胺	0.42 mL
10%APS	0.02 mL
TEMED	0.005 mL
去离子水定容至	2.50 mL

表 3-8 浓缩液

Table 3-8 Stacking solution

试剂名称	用量
1 M Tris-HCl(pH6.8)	38 mL
10%SDS	4 mL
去离子水定容至	250 mL

表 3-9 2×蛋白上样缓冲液

Table 3-9 2×Loading Buffer

试剂名称	用量
SDS	0.4 g
2-Mercaptoethanol	1 mL
Glycerol	2 mL
Bromophenol blue	0.0004 g
1 M Tris-HCl (pH6.8)	1.25 mL
去离子水定容至	10 mL

表 3-10 电泳缓冲液

Table 3-10 Tris-glycine SDS running buffer

试剂名称	用量
Tris	15.1 g
Glycine	94 g
SDS	5 g
去离子水定容至	1 L

表 3-11 过硫酸铵(10%)

Table 3-11 10% Ammonium persulphate

试剂名称	用量
APS	1 g
去离子水定容至	10 mL

表 3-12 十二烷基硫酸钠溶液(10%)

Table 3-12 10% Sodium dodecyl sulfate

试剂名称	用量
SDS	10 g
去离子水定容至	100 mL

表 3-13 丙烯酰胺溶液(30%)

Table 3-13 30% Acrylamide

试剂名称	用量
Acrylamide	290 g
Bis-Acrylamide	10 g
去离子水定容至	1 L

表 3-14 1 M Tris-HCl (pH6.8)

Table 3-14 1 M Tris-HCl(pH6.8)

试剂名称	用量
Tris	121.2 g
浓盐酸调 pH 至	6.8
去离子水定容至	1 L

表 3-15 1 M Tris-HCl (pH8.0)

Table 3-15 1 M Tris-HCl(pH8.0)

试剂名称	用量
Tris	121.2 g
浓盐酸调 pH 至	8.0
去离子水定容至	1 L

表 3-16 1 M Tris-HCl (pH8.8)

Table 3-16 1 M Tris-HCl(pH8.8)

试剂名称	用量
Tris	121.2 g
浓盐酸调 pH 至	8.8
去离子水定容至	1 L

表 3-17 考马斯亮蓝染液

Table 3-17 Coomassie blue staining

试剂名称	用量
考马斯亮蓝 R-250	0.01 g
甲醇	5 mL
冰乙酸	1 mL
去离子水定容至	10 mL

表 3-18 脱色液

Table 3-18 Destaining solution

试剂名称	用量
甲醇	400 mL
冰乙酸	100 mL
去离子水定容至	1 L

表 3-19 Inoue 转化缓冲液

Table 3-19 Inoue transformation buffer

试剂名称	用量
MnCl ₂ ·4H ₂ O	10.88 g
CaCl ₂ ·2H ₂ O	2.2 g
KCl	18.65 g
PIPES (0.5 M pH6.7)	20 mL
去离子水定容至	1 L
过滤除菌	

表 3-20 0.5 M PIPES (pH6.7)

Table 3-20 0.5 M PIPES (pH6.7)

试剂名称	用量
PIPES	15.1 g
氢氧化钾调 pH 至	6.7
去离子水定容至	100 mL
过滤除菌	

表 3-21 5 M NaCl

Table 3-21 5 M NaCl

试剂名称	用量
NaCl	146.25 g
去离子水定容至	500 mL

表 3-22 0.1 M 苯甲基磺酰氟

Table 3-22 0.1 M Phenylmethylsulfonyl fluoride

试剂名称	用量
PMSF	17.42 g
异丙醇定容至	1 L

表 3-23 0.5 M 异丙基-β-D-硫代半乳糖苷

Table 3-23 0.5 M Isopropyl β-D-Thiogalactoside

试剂名称	用量
IPTG	11.92 g
去离子水定容至	100 mL

表 3-24 50×TAE

Table 3-24 50×TAE

试剂名称	用量
Tris	2.42 g
Na ₂ EDTA·2H ₂ O	37.2 g
冰乙酸	57.1 mL
去离子水定容至	2.50 mL

表 3-25 氨苄抗生素(100 mg/mL)

Table 3-25 Ampicillin(100 mg/mL)

试剂名称	用量
氨苄青霉素	1 g
去离子水定容至	10 mL

表 3-26 1 M 麦芽糖

Table 3-26 1 M Maltose

试剂名称	用量
麦芽糖	3.6 g
去离子水定容至	10 mL
过滤除菌	

表 3-27 50 mM NaOH

Table 3-27 50 mM NaOH

试剂名称	用量
NaOH	2 g
去离子水定容至	1 L

表 3-28 Buffer A6

Table 3-28 Buffer A6

试剂名称	用量
1 M Tris-HCl(pH8.0)	20 mL
5 M NaCl	5 mL
去离子水定容至	1 L
浓盐酸调 pH 至	6.0

表 3-29 Buffer B6

Table 3-29 Buffer B6

试剂名称	用量
1 M Tris-HCl(pH8.0)	20 mL
5 M NaCl	200 mL
去离子水定容至	1 L
浓盐酸调 pH 至	6.0

2.5 本实验相关引物序列

实验中所用到的 PCR 引物均由 Primer6 软件设计, 由擎科生物科技有限公司合成, 见表 3-30。

表 3-30 PCR 引物序列

Table 3-30 PCR primer sequences

基因	序列(5'-3')
Eg-Dbr1_F	GATGTACGTAGCGGTGGAGG
Eg-Dbr1_R	CAGCCTTTGCACACGCAGG
Eg-Dbr1_R2	CCCAGAAAGGCGAGGAATTC
Dbr1-M_F	CCACCATGTCAAGTTCCTCAG
Dbr1-M_R	CTGGGAACTTGACATGGTGG
Dbr1-3RACE_F	CGGCAGCCCTGCCTCGCGGCAG
mbp-dbr1co_F	GTATTTTCAGGGATCCGTGTATGTGGCAGTTG
mbp-dbr1co_R	GGCCAGTGCCAAGCTTAACCCAGAAATGCCAGAAATTCAC
Dbr1_oF	CTGGTGCCCCATGCCCCCGCCGCCACCCGCTTCCTC
Dbr1_oR	GAGGAAGCGGGTGGCGGCGGGGGCATGGGGCACCAG
Prp8ia_qF	CTTGGGCTAATGTCTCCTGCTCTTG
Prp8ia_qR	CCTGGCTGAAATGATGATTGATGGTTG
Prp8ib_qF	AAAGGGCACTCTGGATGGCAATT
Prp8ib_qR	CGACCAACCAAATGACGACTTATGAAC
SF3B5_qF	TTTGGTGGGCACATATTTGGCAGG
SF3B5_qR	TTAGGCAATGAACTTGGCGATGGT
SF3B6_qF	ACAGTGGTCTATGGAAGGCTCAGT
SF3B6_qR	GGCTGGGAGTTGCTGGAAAGAAA

2.6 实验方法

2.6.1 纤细裸藻 Dbr1 蛋白基因的生化鉴定

我们通过第一章的生物信息学方法获得了纤细裸藻 Dbr1 全长基因，根据 Dbr1 基因序列设计引物（见表 3-30），进行 PCR 克隆。首先，提取纤细裸藻总 RNA，提取方法见第一章 2.6.5。利用基因特异性引物 Eg-Dbr1_R 按照 FastKing cDNA 第一链合成试剂盒进行逆转录，将 Eg-Dbr1_R 替换表 1-10 中的 FQ-RT Primer Mix，其余步骤参照第一章 2.6.6。以逆转录获得的 cDNA 为模板，用引物 Eg-Dbr1_F 和 Dbr1-M_R 扩增出 Dbr1 基因的 5'端片段。其次，按照 HiScript-TS 5'/3'RACE Kit 的说明，设计 Tm 为 60-70℃的引物 Dbr1-3RACE_F，以提取的纤细裸藻总 RNA 为模板，逆转录获得 Dbr1 基因的 3'末端 cDNA 片段。详细步骤参照 HiScript-TS 5'/3'RACE Kit 说明书。

再以上述 3'RACE 扩增产物为模板, 利用引物 Dbr1-M_F 和 Eg-Dbr1_R 扩增出 Dbr1 3'端片段。最后以扩增出的 Dbr1 基因 5'端与 3'端片段作为模板加入到 PCR 反应体系, 暂时不添加引物进行 5 个循环反应, 用于两段模板的连接, 随后加入引物 Eg-Dbr1_F 和 Eg-Dbr1_R 进行常规 PCR 反应 30 个循环, 扩增出 Dbr1 全长片段。对所得序列进行 Sanger 测序分析, 确定获得正确的纤细裸藻 Dbr1 基因序列。

2.6.2 体外纯化纤细裸藻 Dbr1 蛋白

(1) 片段扩增及同源重组

构建带有目的基因 Dbr1 的表达质粒。将去除柔性序列的 Dbr1 序列送往九天基因公司进行密码子优化和基因合成。将优化后的 Dbr1co 通过引物 mbp-dbr1co_F 和 mbp-dbr1co_R 进行扩增。根据试剂盒 (ClonExpress II One Step cloning Kit, Vazyme) 说明将扩增出的 PCR 片段与 BamHI/HindIII 双酶切后的 pMALt 线性化载体在加入酶及缓冲液后于 37℃ 恒温培养箱中共同孵育半小时进行同源重组, 构建含有 TEV 酶切位点的质粒 pMALt-Dbr1。

(2) 细菌转化及菌落鉴定

将商品化的 DH5 α 超级感受态放置冰上解冻, 将上述重组体加到感受态细胞中, 充分混匀。冰上放置半小时后使用恒温金属浴 42℃ 热击 90 sec, 热击后立即置于冰上 2 min。再加入 1 mL 无抗生素的 LB 培养基, 37℃ 恒温振荡箱培养 1 h, 12000 rpm 室温离心 1 min, 弃去上清, 细胞用 100 μ l 的 LB 培养基悬浮并滴加至含 Amp 的 LB 固体培养基上, 平板涂匀后放置 37℃ 恒温培养箱孵育 12-24 h。观察平板上的菌落生长情况, 挑取单克隆至 3 mL 的 LB (Amp) 培养基中, 37℃ 恒温振荡箱培养 8 h 至菌液浑浊。以菌液为模板, 进行鉴定 PCR。从鉴定结果为阳性克隆的菌液中提取质粒, 送往公司进行 Sanger 测序, 以此来验证质粒是否构建成功。

(3) 体外诱导蛋白及淀粉直链层析纯化

①BL21 (DE3) 感受态的制备。将冻存的大肠杆菌表达菌株 BL21 (DE3) 接种至 3 mL 的 LB 培养基中, 37℃ 过夜活化。吸取 1 mL 菌液到 100 mL 的 LB 培养基中,

放置 18℃ 恒温振荡箱中低温培养，至 OD 600 值为 0.5 左右。4℃ 2500 g 离心 10 min 收集菌体，弃上清，沉淀用 80 mL 预冷的 Inoue 转化缓冲液重悬，4℃ 2500 g 离心 10 min，弃上清，沉淀用 16 mL 预冷的 Inoue 转化缓冲液重悬。重悬液中加入 1 mL 二甲亚砜（DMSO），轻轻混匀，冰上放置 10 min。随后分装至无菌预冷的 1.5 mL EP 管中并迅速在液氮中速冻，-80℃ 保存。

②目的蛋白表达与纯化。将质粒转化到 BL21 (DE3) 表达菌株中，转化方法参照上述 DH5 α 感受态。12 h 培养后挑取单克隆到 3 mL 含有 Amp 的 LB 培养基中，37℃ 培养至菌液浑浊，吸取 200 μ l 菌液到 20 mL 的 LB (Amp) 培养基中，37℃ 培养至 OD 600 值为 0.5 左右，然后将 10 mL 的菌液接到无菌的离心管中，在 MBP-Dbr1 中加入终浓度为 0.25 mM 的 IPTG 进行诱导，剩余 10 mL 作为对照组，在 18℃ 下共同培养 12 h。4000 rpm 10 min 收集菌体，加入无菌水将菌体清洗一遍，再次离心。加入预冷的缓冲液 buffer A 以及百分之一的 0.1 M PMSF 重悬菌体，超声后进行 4℃ 低温 12000 rpm 离心 30 min，吸取上清与 2 $\times$ Loading Buffer 混匀，煮沸 10 min，制备样品完成，利用 10% SDS-聚丙烯酰胺凝胶电泳和考马斯亮蓝染色观察目的蛋白是否表达。

将单克隆接种到 10 mL 的 LB (Amp) 培养基中培养至菌液浑浊后，全部菌液转入 1 L 的 LB (Amp) 培养基，OD 600 值为 0.5 左右，在菌液中加入终浓度为 0.25 mM 的 IPTG，18℃ 下诱导 12 h。收集菌体，加入 Buffer A 缓冲液及 PMSF 重悬，超声后进行 4℃ 12000 rpm 离心 30 min。MBP 标签融合蛋白用直链淀粉树脂亲和纯化，首先用低盐浓度的 Buffer A 将 Amylose Resin 平衡三次，将离心后的菌体上清液倒入淀粉树脂填料中，用 Buffer A 清洗树脂，用 12 mM 麦芽糖洗脱目的蛋白质。将纯化后的蛋白利用 10% SDS-PAGE 和考马斯亮蓝进行分析鉴定。

③高速强阴离子交换层析，首先设定 AKTA Concentration %B 为 50%，以 2 mL/min 流速依次使用无菌 ddH₂O、50 mM NaOH、无菌 ddH₂O 清洗系统 20 min。随后分别将探头 A 放入低盐浓度 Buffer A6 中，探头 B 放入高盐浓度 Buffer B6 中。将 Concentration %B 设置为 100%用高盐浓度 Buffer B6 进行洗柱，去除层析柱上的杂蛋白，清洗完毕后将 Concentration %B 设置为 0%利用低盐浓度 Buffer A6 进

行平衡。将 Inject Value Pos 设置为 Load，利用注射器开始进样，样品完全进入后，设置 Inject Value Pos 为 Inject。多次进样后，待紫外数值稳定，将 Concentration %B 设置为 0-100% 梯度进行洗脱，收集紫外峰处的蛋白溶液，通过 10% SDS-PAGE 和考马斯亮蓝染色检查纯化结果。

④ 蛋白酶切。将 TEV 酶与 MBP-Dbr1 蛋白进行浓度比为 1:2 混合，4℃ 或 33℃ 下作用 10 h。用低盐浓度的 buffer A 将 Amylose Resin 平衡三次，随后将混合液倒入平衡后的 Amylose Resin 填料中，4℃ 孵育一段时间后，分别收集流出的溶液以及填料。

2.6.3 纤细裸藻 Dbr1 蛋白三维结构分析

首先，通过 PDB 蛋白质数据库得到溶组织内阿米巴虫 Dbr1 结构 (PDB:4PEH)。随后，通过 Alphafold 或 Robetta 程序对纤细裸藻 Dbr1 蛋白质结构进行预测和分析。使用软件 UCSF Chimera 将纤细裸藻 Dbr1 与溶组织内阿米巴虫 Dbr1 蛋白质的三维空间结构进行叠合并比较，得到二者的 RMSD 值。

2.6.4 结合纤细裸藻 Dbr1 的小分子的虚拟筛选

根据纤细裸藻 Dbr1 与溶组织内阿米巴虫 Dbr1 结构进行对比，确定结构中结合套索 RNA 所在的部位。通过溶组织内阿米巴虫的 RNA 结合位置去推测出小分子化合物在纤细裸藻 Dbr1 蛋白中的结合区域。

基于上述推测出的小分子在纤细裸藻 Dbr1 中的结合位置，利用 AutoDock Vina 软件从 TargetMol 天然产物库中的天然产物单体进行虚拟筛选，并进行结合能力评价。最终从大量的化合物分子中挑选潜在结合活性最高的化合物用于后续实验。

2.6.5 体外验证小分子抑制剂与纤细裸藻 Dbr1 的结合

通过核磁共振化学位移扰动分析 (NMR-CSP) 实验观察蛋白与小分子化合物结合前后的谱峰变化，以此来判断两者的结合，并定量结合的强弱。此部分正在进行中。

2.6.6 体内检测小分子的生物活性

将筛选到的小分子抑制剂直接作用于纤细裸藻细胞，通过观察小分子对纤细

裸藻细胞生长及形态的影响。在纤细裸藻细胞中加入筛选到的活性小分子, 浓度为 50 μmol , 25 $^{\circ}\text{C}$ 恒温培养三天。分别提取纤细裸藻细胞加药组与对照组的总 RNA。实验步骤参照上述提取纤细裸藻总 RNA。对于加药组与对照组 cDNA 的合成, 用 FQ-RT Primer Mix 替换上述逆转录步骤中的 Eg-Dbr1_R, 其余步骤均相同。以纤细裸藻 cDNA 为模板, 通过染料法进行实时荧光定量 PCR 检测 Prp8-ia、Prp8-ib、SF3B5、SF3B6 等基因的内含子, 引物设计如见表 3-30。检测小分子化合物处理后细胞内含子去除是否受到了影响。实验反应程序设置见表 3-31。

表 3-31 qPCR 反应程序

Table 3-31 qPCR procedure

Stage1	预变性	Reps:1	95 $^{\circ}\text{C}$	30 sec
Stage 2	循环反应	Reps:40	95 $^{\circ}\text{C}$	3-10 sec
			60 $^{\circ}\text{C}$	10-30 sec
Stage3	融解曲线	Reps:1	95 $^{\circ}\text{C}$	15 sec
			60 $^{\circ}\text{C}$	60 sec
			95 $^{\circ}\text{C}$	15 sec

3 结果与分析

3.1 纤细裸藻 Dbr1 基因的确定

我们通过转录组测序以及结合 SRA 数据库分析预测获得了纤细裸藻 Dbr1 基因的完整序列。我们尝试通过 PCR 设计一对引物扩增全长 Dbr1 编码序列, 但没有成功。为了获得 Dbr1 基因全长序列, 我们采用了分段克隆及重叠 PCR 拼接的策略。首先我们利用 Eg-Dbr1_R 作为逆转录引物, 扩增出纤细裸藻 cDNA, 用引物 Eg-Dbr1_F 和 Dbr1-M_R 进行 PCR 得到纤细裸藻 Dbr1 基因的 5'端序列的片段。其次我们利用 HiScript-TS 5'/3'RACE Kit, 设计特异性引物 Dbr1-3RACE_F 进行 Dbr1 基因的 3'RACE 扩增。以 3'RACE 扩增得到的 cDNA 作为模板, 利用引物 Dbr1-M_F 和 Eg-Dbr1_R 得到 Dbr1 基因的 3'端序列的片段。最后通过重叠 PCR, 得到完整的 Dbr1 序列 (图 3-1)。测序表明扩增序列与拼接序列是相同的。这进一步提示我们所鉴定出来的纤细裸藻 Dbr1 基因存在且序列正确。

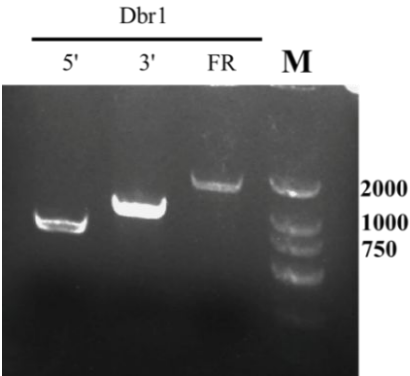


图 3-1 Dbr1 基因的 PCR 鉴定；FR:全长序列

Fig.3-1 PCR verification of Dbr1gene. FR: the full-length sequence

3.2 纤细裸藻 Dbr1 蛋白的纯化

我们将纤细裸藻 Dbr1 基因进行体外纯化，首先去除 Dbr1 蛋白表面柔性序列、密码子优化及基因合成，然后装入带有 TEV 酶切位点的细菌表达质粒 pMALt，获得质粒 pMAL-Dbr1（图 3-2）。接下来我们将质粒转化细菌 BL21 (DE3)，通过 IPTG 诱导、Amylose Resin 亲和层析柱、TEV 酶切以及离子交换纯化获得 Dbr1 蛋白（图 3-3 和 3-4）。



图 3-2 质粒 pMAL-Dbr1 结构示意图

Fig.3-2 Schematic depiction of a plasmid, pMAL-Dbr1

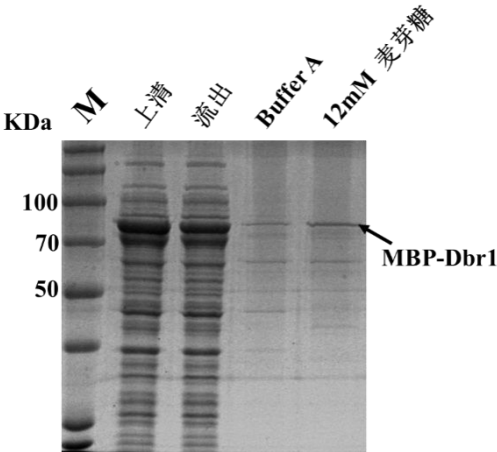


图 3-3 10%SDS-PAGE 电泳分析纯化蛋白

Fig.3-3 10% SDS-PAGE analysis of purified proteins

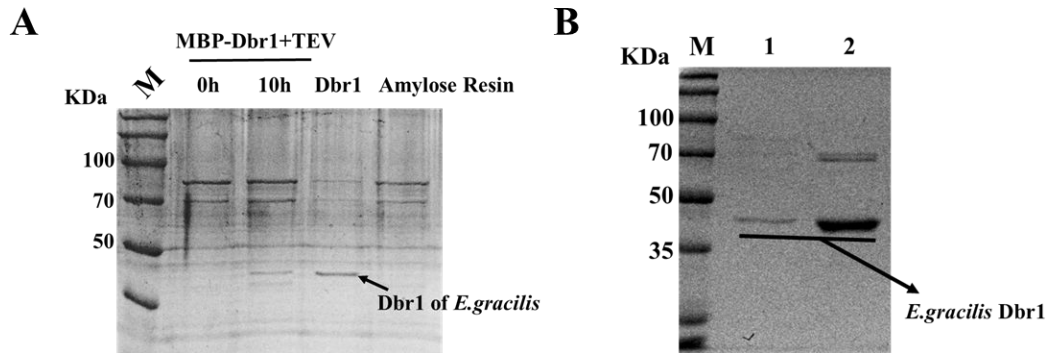


图 3-4 纤细裸藻 Dbr1 蛋白的酶切与纯化

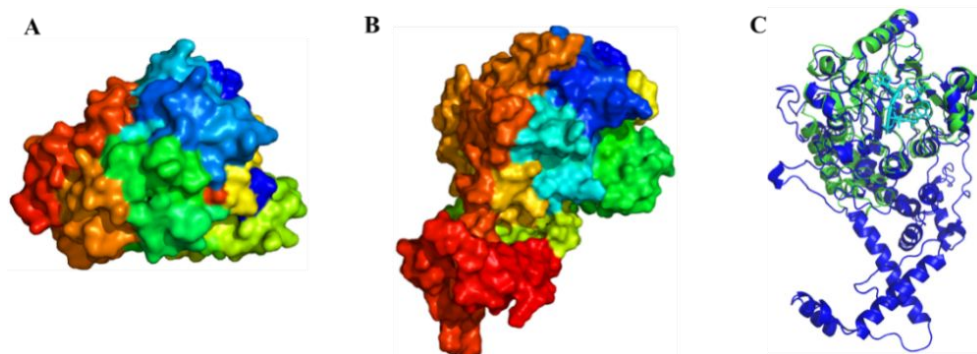
Fig.3-4 Enzymatic digestion and purification of *E. gracilis* Dbr1 protein

注：A 为 10%SDS-PAGE 电泳分析经 TEV 酶酶切后的纤细裸藻 Dbr1 蛋白；B 为酶切后超滤，保存的纤细裸藻 Dbr1 蛋白。

Note: A is 10% SDS-PAGE analysis of *E. gracilis* Dbr1 protein after digestion of TEV enzyme; B is preserved Dbr1 protein of *E. gracilis* after enzymatic digestion and ultrafiltration.

3.3 纤细裸藻 Dbr1 蛋白结构预测

使用 Alphafold 和 Robetta 程序对纤细裸藻 Dbr1 蛋白质序列进行建模，预测出纤细裸藻 Dbr1 蛋白晶体三维结构图（图 3-5B）。并用软件 UCSF Chimera 将纤细裸藻 Dbr1 与溶组织内阿米巴虫 Dbr1 晶体结构(PDB:4PEH)（图 3-5A）进行叠合比对分析得到 RMSD 值为 1.024(图 3-5C)，说明建立的模型可靠性较高。



D

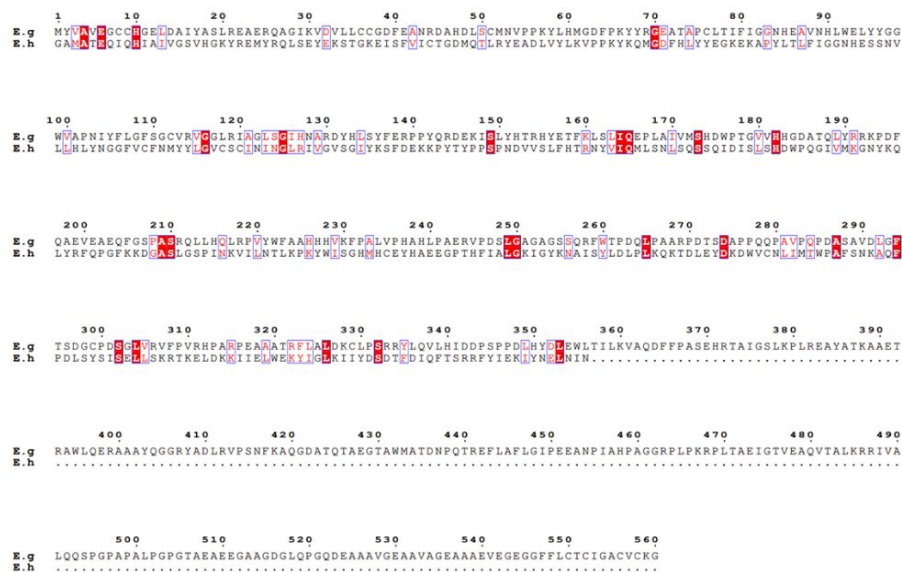


图 3-5 溶组织内阿米巴虫与纤毛裸藻 Dbr1 蛋白

Fig.3-5 *E. histolytica* and *E. gracilis* Dbr1 protein

注：A 为溶组织内阿米巴虫 Dbr1 蛋白结构图（PDB:4PEH）；B 为 AlphaFold 和 Robetta 等 AI 程序预测出的纤毛裸藻 Dbr1 蛋白结构图；C 为溶组织内阿米巴虫 Dbr1（绿色）与纤毛裸藻 Dbr1（蓝色）叠合后的蛋白质三维空间结构比对；D 为纤毛裸藻与溶组织内阿米巴虫 Dbr1 蛋白序列对比图。

Note: A is *E. histolytica* Dbr1 protein structure (PDB:4PEH); B is the structure of *E. gracilis* Dbr1 is predicted by AI programs including AlphaFold and Robetta; C is superposition of *E. histolytica* (green) and *E. gracilis* Dbr1 structures (blue); D is comparison of Dbr1 protein sequences in *E. gracilis* and *E. histolytica*.

3.4 小分子化合物的筛选

由阿米巴虫 Dbr1 晶体结构中 RNA (5'-CUAXCAA-3') 分子的位置 (图 3-6A)，推测纤毛裸藻 Dbr1 蛋白质与抑制剂小分子结合的空间坐标位置：坐标中心为 X: 13.71, Y: 51.294, Z: -44.973，范围大小为 X: 24, Y: 26, Z: 42 (图 3-6B)。通过 AutoDock Vina 程序进行虚拟筛选，从中挑选出结合能最低、对接结果最稳定的数据，即鉴定出 1 个具有潜在结合活性的小分子 (图 3-6C)。

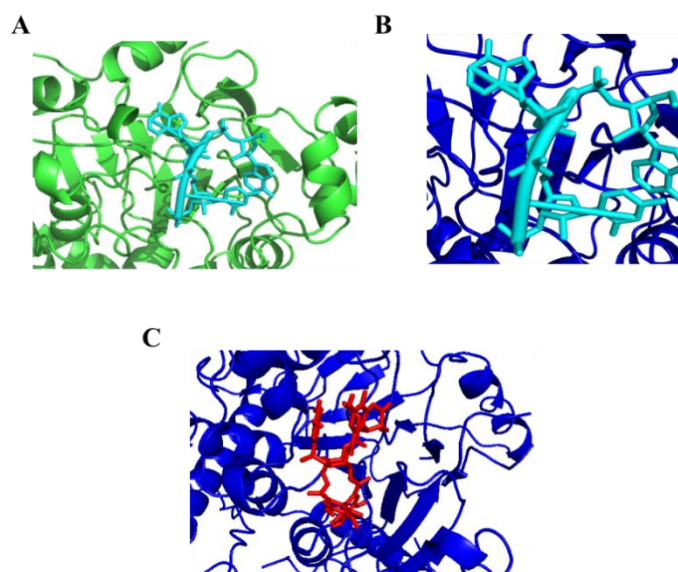


图 3-6 溶组织内阿米巴虫与纤细裸藻 Dbr1 蛋白与小分子作用三维结构图

Fig.3-6 The three-dimensional structures of *E. histolytica* and *E. gracilis* Dbr1 and small molecules interaction

注：A 为溶组织内阿米巴虫 Dbr1 蛋白（绿色）与 RNA 分子（浅蓝色）空间结合位置；B 为纤细裸藻 Dbr1 蛋白（蓝色）与预测的抑制剂小分子（浅蓝色）结合区域；C 为小分子 Ca（红色）与纤细裸藻 Dbr1 结合。

Note: A is the spatial coordinate of small molecule RNA (light blue) in the *E. histolytica* Dbr1 structure (green); B is the binding region of the Dbr1 protein (blue) and the predicted small molecule (light blue) of *E. gracilis* (blue); C is Ca (red) is bound to *E. gracilis* Dbr1(blue).

3.5 小分子化合物的体内活性检测

将虚拟筛选出的小分子 Ca 加入到纤细裸藻细胞中，终浓度为 50 μmol 。纤细裸藻细胞在 25 $^{\circ}$ 恒温摇床中培养 3 天，我们发现其形态以及生长并未受到影响。随后我们提取小分子处理后的细胞总 RNA，对剪接基因内含子 Prp8-ia、Prp8-ib、SF3B5、SF3B6 进行 RT-qPCR 检测。qPCR 结果显示经过 Ca 处理后的纤细裸藻其多种内含子表达量要显著高于对照细胞（图 3-7）。

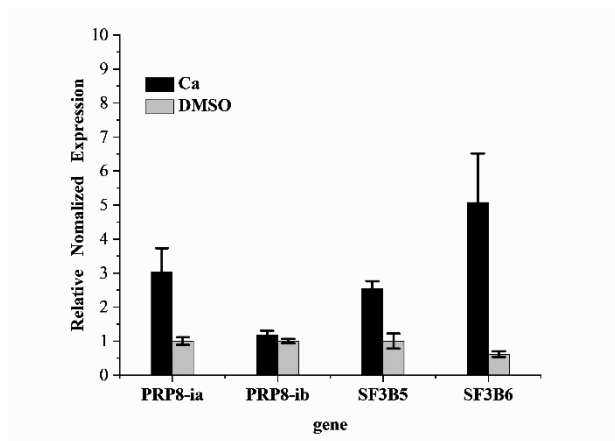


图 3-7 小分子化合物对纤细裸藻 Dbr1 的体内作用

Fig.3-7 In vivo effects of small molecular compounds on *E. gracilis* Dbr1

4 结论

通过虚拟筛选技术，结合体外及体内实验获得了一种抑制 RNA 脱支酶 Dbr1 蛋白的小分子化合物 Ca，为将来对纤细裸藻内含子种类及数量的研究奠定基础。

5 讨论

在进行 Dbr1 克隆时发现，我们使用常规 PCR 方法无法扩增到全长 Dbr1。有文献报道过在植物细胞中 Dbr1 属于低拷贝基因^[55]，我们推测纤细裸藻内 Dbr1 基因的表达量同样不高，故无法用常规的 RT-PCR 方法扩增出全长片段。

在进行小分子抑制剂 Ca 的体内功能验证实验中，Ca 作用纤细裸藻细胞三天后通过实时荧光定量 PCR 我们发现细胞体内内含子的累积量明显高于正常细胞，但是同时 Ca 处理后的细胞在显微镜下观察并没有发现形态变化。与正常细胞相比在细胞数量上变化也不明显。我们猜测随着培养时间的延长，在 25℃ 下小分子的结构稳定性会逐渐变弱，导致其对 Dbr1 的抑制效果逐渐变差，体内一些基因仍然在正确的进行剪接成熟。因此还不能导致细胞的形态以及数量的变化。

目前，未见关于 Dbr1 小分子抑制剂的报道。我们以内含子类型独特的纤细裸藻作为研究对象，通过生信及实验确定纤细裸藻 Dbr1 基因，虚拟筛选出针对 Dbr1 活性位点的小分子抑制剂。通过体外检测小分子抑制剂与蛋白的结合活性及体内鉴定小分子抑制剂对蛋白催化活性的抑制能力。Dbr1 小分子抑制剂的鉴定将为下一步转录组鉴定纤细裸藻的内含子种类及数量及研究 Dbr1 催化活性奠定基础。

第四章 纤细裸藻生物学转化方法的构建

1 目的

纤细裸藻是一种单细胞真核生物，其含有能进行光合作用的叶绿体，并且尾端有鞭毛起到帮助细胞运动的作用。因此纤细裸藻具有植物和动物的共同特征。它能够合成 β -胡萝卜素、副淀粉、维生素和脂肪酸等高附加值产物，具有重要的科研和经济价值^[56]。目前，将外源物质转入纤细裸藻细胞的方法主要有农杆菌介导的核转化、电穿孔、生物轰击和单细胞显微注射四种^[57]。虽然这些方法对纤细裸藻细胞的转化是可行的，而且效率很高，但它们都需要复杂的设备和优化的参数，或者需要相对较长的时间来筛选阳性克隆。在裸藻基因工程中，转化是实现高效遗传改造的必要步骤。因此，我们针对这些问题发展了两种全新的转化方法。第一种是跨膜肽(CPPs)，本研究基于细菌表达质粒pET21-EGFP，成功构建含有跨膜肽标签Pep-1序列的相关质粒。通过观察细胞内荧光信号来检测跨膜肽赋予结合的外源物质进入细胞的能力。与跨膜肽相似，第二种方法所用试剂二甲亚砜也能够影响膜渗透性，并且已经用于酵母转化^[58]。我们的研究目的是为了丰富纤细裸藻转化的方法选择，促进未来对裸藻生物的分子调控。建立微藻物种的基因工程工具对微藻产业化至关重要。

2 材料与方法

2.1 细胞株

实验所用纤细裸藻细胞株与第一章相同

2.2 主要试剂

表 4-1 试剂名称及来源

Table 4-1 Reagent name and source

试剂名称	生产公司
二甲亚砜 (DMSO)	MP, 美国
DAPI 染色液	生工, 中国
双色预染蛋白 Marker	生工, 中国
Imidazole (咪唑)	Biosharp, 中国
4%多聚甲醛通用型组织固定液	Biosharp, 中国

续表 4-1

试剂名称	生产公司
镍 NTA 琼脂糖凝胶 6FF	源叶生物, 中国
TritonX-100	Biofroxx, 德国

其余实验试剂与第三章相同, 详见表 3-1

2.3 主要仪器及耗材

表 4-2 仪器设备名称及来源

Table 4-2 Reagent name and source of instruments and equipment

仪器名称	生产厂家
BX63 正置荧光显微镜	奥林巴斯, 日本

其余仪器设备与第三章相同, 详见表 3-2

2.4 常规实验试剂配置

表 4-3 10×磷酸缓冲盐溶液

Table 4-3 10×PBS

试剂名称	用量
NaCl	40 g
KCl	1 g
Na ₂ HPO ₄ · 12H ₂ O	7.70 g
KH ₂ PO ₄	0.90 g
浓盐酸调节 pH 值	7.20~7.40
去离子水定容至	500 mL

表 4-4 1 M 咪唑

Table 4-4 1 M Imidazole

试剂名称	用量
咪唑	79.11 g
去离子水定容至	1 L
过滤除菌	

其余试剂配置与第三章相同, 详见第三章 2.4

2.5 本实验相关引物序列

实验中所用到的 PCR 引物均由 Primer6 软件设计，由擎科生物科技有限公司合成，见表 4-5。

表 4-5 PCR 引物序列

Table 4-5 PCR primer sequences

基因	序列 (5'-3')
21G-pep-1sc_F	AGGAGATATACATATGAAAGAAACCTGGTGGG
21G-pep-1sc_R	CCTTGCTCACGGATCCCACCTTTACGTTTTTTTTTTC
GP_oF	CGGCATGGACGAGCTGTACAAGGGATCCAAAGAAACCTGGTGGGA AACC
P21sc_R	GGTGGTGGTGGCTCGAGCACTTTACGTTTTTTTTTTCGGC
21HGPsc_F	AATGGGTCGCGGATCCCACCACCACCACCACCTGAGCAAGGGC GAGGA
21HGPsc_R	GGTGGTGGTGGCTCGAGCTACACTTTACGTTTTTTTTTTCGGCTG
21HPGsc_F	AGGAGATATACATATGCACCACCACCACCACAAAGAAACCTGG TGGGAAACC
21HPGsc_R	GGTGGTGGTGGCTCGAGCTAGTACAGCTC
egP8_F	CAACAACCTGACGGACATCTG
Cy5-egP8_F	Cy5-CAACAACCTGACGGACATCTG
egP8_R	GCTTCTGTACTTCATGCTCTCC
211-L5sc_F	ATCCTCTAGAGTCGACAGGGAAACGTCTCTTTTGGC
mGFP6seq_R	GGGTTGGCCATGGAACAGG

2.6 实验方法

2.6.1 质粒构建

(1) 片段的扩增及同源重组

本研究采用不依赖连接的克隆方法 (ClonExpress II One Step cloning Kit, Vazyme)，以 pET21-EGFP 质粒为基础构建而成（图 4-1），该质粒含有 EGFP 编码序列，上游为 T7 标签，下游为 6xHis 残基。pET21-PGH 是用引物 21G-pep-1sc_F 和 21G-pep-1sc_R 将 Pep-1 (LETWWETWWTEWSQPKKKRKV) 序列扩增，取代 pET21-EGFP 上的 T7 标签序列构建的。再用引物 GP_oF 和 P21sc_R 扩增 Pep-1 片段，将其扩增出的片段插入 pET21-EGFP 中构建 pET21-GPH。为了构建 pET21-HGP，从 pET21-GPH 中扩增出 GFP-Pep-1 片段，正向引物 21HGPsc_F 包含 6xHis 序列，反向引物 21HGPsc_R 包含 Pep-1 序列后的终止密码子。随后，将

扩增的 GFP-Pep-1 片段插入到由 BamHI/XhoI 双酶切的 pET21-EGFP 中，生成 pET21-HGP。与 pET21-HGP 类似，pET21a-HPG 是将 pET21-PGH 中的 Pep-1-GFP 片段用含有 6×His 序列的正向引物 21HPGsc_F 和 EGFP 序列后含有终止密码子的反向引物 21HPGsc_R 扩增，然后插入 NdeI/XhoI 双酶切的 pET21-EGFP 中构建的。

(2) 细菌转化及菌落鉴定见第三章 2.6.2(2)。

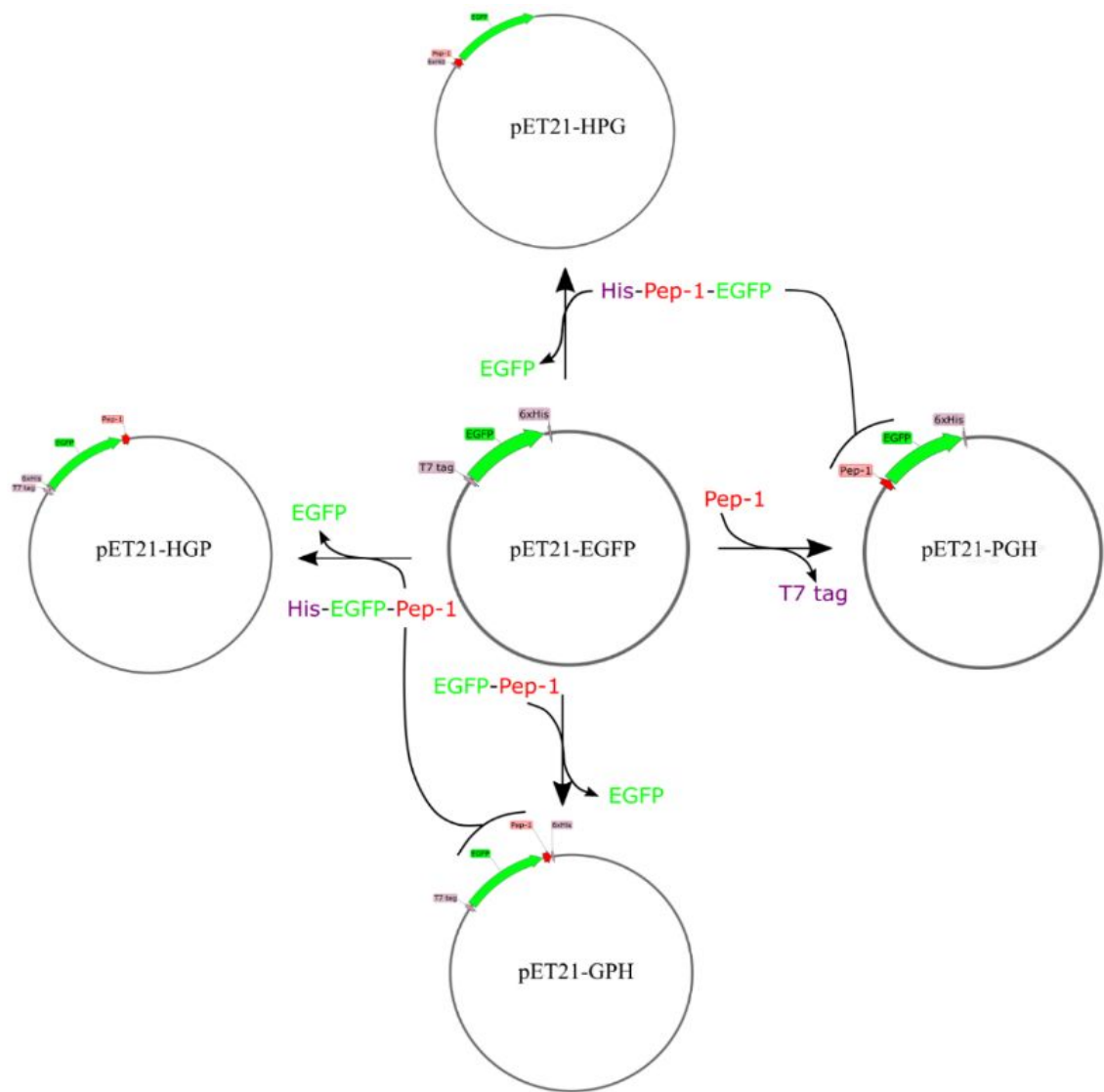


图 4-1 质粒构建流程示意图

Fig.4-1 Schematic diagram of the procedure for plasmid construction

2.6.2 体外诱导蛋白及纯化

(1) BL21 (DE3) 感受态的制备见第三章 2.6.2(3)。

(2) 目的蛋白表达

质粒转化及诱导操作步骤参照第三章 2.6.2(3)。

(3) 目的蛋白纯化

通过小剂量诱导的结果观察到目的蛋白可以成功表达后，我们立即进行大量纯化。单克隆接种到 10 mL 的 LB (Amp) 培养基中培养至浑浊后，再次将 10 mL 菌液转入到 1 L 的 LB (Amp) 培养基，培养到 OD 600 值为 0.5 左右，加入终浓度为 0.5 mM 的 IPTG 在 37℃ 诱导 3 h。收集菌液，加入 buffer A 缓冲液及 PMSF 重悬，超声后进行 4℃ 低温 12000 rpm 离心 30 min。可溶性 His（组氨酸）标签融合蛋白用 Ni-NTA 树脂亲和纯化，首先用低盐浓度的 buffer A 将镍填料平衡三次，将离心后的上清液倒入填料中挂柱，用 30 mM 咪唑清洗树脂，用 300 mM 咪唑洗脱蛋白质。洗脱后的蛋白通过超滤进一步浓缩，并在 PBS 上透析。纯化后的蛋白利用 10% SDS-PAGE 和考马斯亮蓝染色进行观察。

2.6.3 纤细裸藻细胞内转化

(1) 跨膜肽（CPPs）介导的细胞内转化

收集纤细裸藻细胞，室温下 1000 g 离心 1 min，用 HUT 培养基洗涤 1 次。随后，将制备好的不同浓度的蛋白加入细胞中，细胞 OD 600 值为 1 (0.86×10^6 个细胞/mL)。然后将细胞在 37℃ 下孵育不同时间。

(2) 二甲亚砜（DMSO）介导的细胞内转化

收集纤细裸藻细胞后，细胞在含有不同浓度 DMSO 的 HUT 培养基中 25℃ 孵育 1 小时，PBS 洗涤 3 次后加入含有外源蛋白或核酸的 HUT 培养基，25℃ 下孵育 12 小时。

2.6.4 显微观察

蛋白或核酸孵育后，室温下 1000 g 离心 1 min 收集纤细裸藻细胞，PBS 冲洗多次后，使用 DAPI 染色: 4% PFA 固定细胞 20 分钟，0.5% tritonX-100 处理细胞 15 分钟，1 μ g/mL DAPI 染色 5 分钟，再次使用 PBS 冲洗几次。最后 BX63 正置荧光显微镜观察染色结果。

2.6.5 统计学分析

每个实验至少重复三次，测试的浓度平均为三份。采用 Prism7 (GraphPad) 软件进行统计分析。使用显著性水平设置为 95%。基于差异有统计学意义，采用 Tukey 多重事后检验 (PGH 转化) 和单因素方差分析检验 (DMSO 转化) 对各组数据进行分析。对于 PGH 和 DMSO 转化，当 $P < 0.0001$ 和 $P < 0.05$ 时，表示具有统计学意义。

3 结果与分析

3.1 Pep-1-GFP 融合蛋白及相关蛋白的纯化

为了研究跨膜肽 (CPPs) 介导的外源物质在纤细裸藻中的转化作用，我们选择了两亲性 CPP 肽 Pep-1 进行研究^[59]。根据相关文献合成了 Pep-1 的核酸序列^[60]。以含有 EGFP 编码序列的质粒 pET21-EGFP 为模板构建 Pep-1 相关质粒。将 Pep-1 插入 pET21-EGFP 中，在 T7 标签和 EGFP 之间，获得 pET21-PGH 质粒 (图 4-2)。随后，将 pET21-PGH 转化到 BL21 (DE3) 菌株中进行蛋白表达。经 IPTG 诱导后，SDS-PAGE 证实 PGH 蛋白的表达。由于 C-末端 6xHis 标签的存在，该蛋白的分子量为 31 kD，通过 Ni-NTA 亲和层析进一步纯化，并通过 SDS-PAGE 和考马斯亮蓝染色进行观察 (图 4-3)。

除了 pET21-PGH 质粒外，我们还构建了另外三个与 Pep-1 相关的质粒。这三种质粒 pET21-HPG、pET21-HGP 和 pET21-PGH 仅在 His 标签和 Pep-1 的位置上有所不同 (图 4-2)。在与 pET21-PGH 相同的实验条件下，成功表达了这些质粒以及原始质粒 pET21-EGFP，并纯化了相应的蛋白 (图 4-3)。此外，我们还构建了不含 T7 标签序列的 pET21-GPH 质粒。然而，我们在 IPTG 诱导后未检测到该质粒的蛋白表达，提示 GPH 表达需要 T7 标记序列。

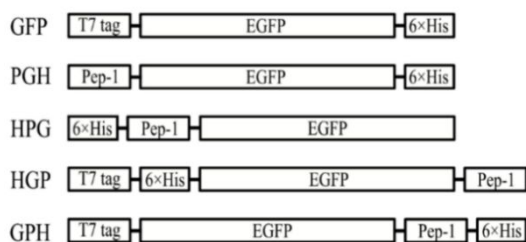


图 4-2 四种 Pep-1 融合蛋白与对照 GFP 蛋白结构示意图

Fig.4-2 Schematic depiction of four Pep-1 fusion protein and control GFP protein

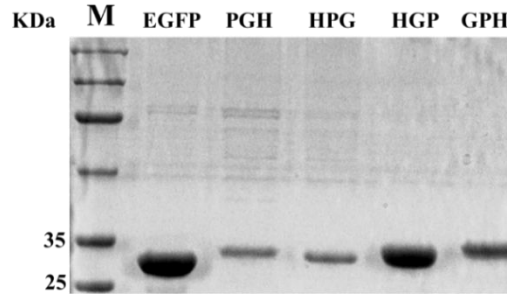


图 4-3 10%SDS-PAGE 分析纯化后的四种融合蛋白

Fig.4-3 10% SDS-PAGE analysis of four purified proteins

3.2 PGH 蛋白的细胞内转化

先前有研究发现，Pep-1 融合蛋白可以在 0.25 μM 、15 分钟的时间内进入细胞^[59]。然而，当我们将 5 μM 的 PGH 加入到纤细裸藻细胞中，在 25 $^{\circ}\text{C}$ 孵育 12 小时后，没有观察到荧光信号。只有将 PGH 浓度增加到 50 μM 后，我们才能观察到很少的荧光信号。由于人类细胞通常在 37 $^{\circ}\text{C}$ 下培养，我们推测在这个更高的温度下，纤细裸藻细胞中的 PGH 转化效率可能会增加。事实上，我们在 37 $^{\circ}\text{C}$ 时发现的荧光信号比 25 $^{\circ}\text{C}$ 时多(图 4-4)。因此，我们在这个更高的温度下进行了 Pep-1 融合蛋白的转化。

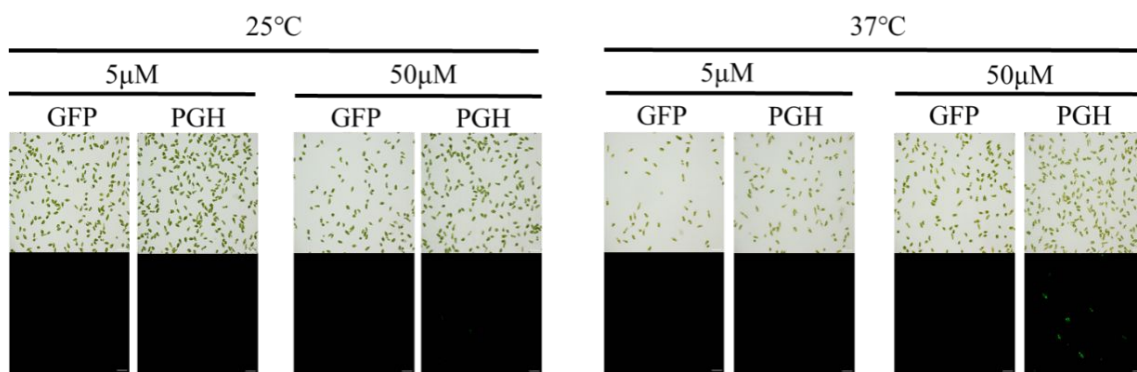


图 4-4 比较 25 $^{\circ}\text{C}$ 与 37 $^{\circ}\text{C}$ 下 PGH 的细胞内转化。加入 5 或 50 μM PGH 蛋白，于 25 $^{\circ}\text{C}$ 和 37 $^{\circ}\text{C}$ 孵育 12 小时。标尺：50 μm

Fig.4-4 Intracellular transformation of PGH at 25 $^{\circ}\text{C}$ and 37 $^{\circ}\text{C}$ was compared. 5 or 50 μM PGH protein was added to *E. gracilis* cells and incubated at 25 $^{\circ}\text{C}$ and 37 $^{\circ}\text{C}$ for 12 hours. Scale bar: 50 μm

我们测试了三种浓度的 PGH 对纤细裸藻转化的影响。将 $5\ \mu\text{M}$ 的 PGH 加入到纤细裸藻细胞中，孵育 2 小时和 7 小时后均未观察到荧光信号。即使经过 12 小时的孵育，也只有极少数细胞(约 1%)表现出荧光信号。然后将 PGH 浓度分别提高到 $50\ \mu\text{M}$ 和 $100\ \mu\text{M}$ 。在 $50\ \mu\text{M}$ 和 $100\ \mu\text{M}$ PGH 孵育 7 小时后，我们可以分别观察到超过 25% 和 60% 的转化效率(图 4-5)。孵育时间延长至 12 小时，仅略微提高转化效率($100\ \mu\text{M}$ 时为 70%)。虽然 PGH 在纤细裸藻细胞中的浓度远高于人体细胞，并且对纤细裸藻的孵育温度 (37°C) 远高于其生理温度 (25°C)，但我们发现 PGH 转化的纤细裸藻细胞仍然具有运动能力。除了 Pep-1，我们还用 TAT(一种富含精氨酸的阳离子跨膜肽)进行了测试，并发现 TAT-EGFP 的转化效率与 PGH 相当。在所有这些转化条件下，我们都没有发现不含有 Pep-1 的 GFP 蛋白的荧光信号，表明细胞内荧光信号确实需要跨膜肽的存在。

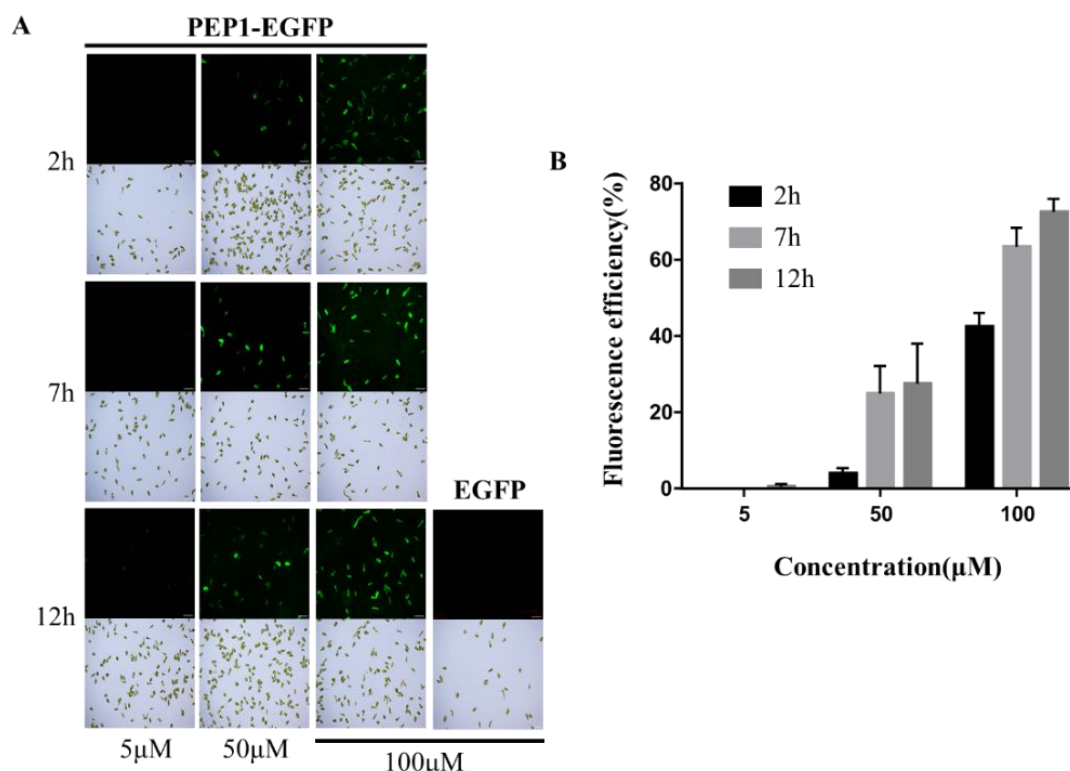


图 4-5 Pep-1 融合蛋白 PGH 的纤细裸藻细胞内转化

Fig.4-5 Intracellular transformation of the Pep-1 fusion protein PGH in *E. gracilis* cells

注：A 为纤细裸藻细胞中 PGH 的时间和浓度依赖性细胞内转化。细胞分别用不同浓度的 PGH 蛋白处理 2、7 或 12 小时，然后在显微镜下观察。标尺： $50\ \mu\text{m}$ ；B 为纤细裸藻细胞

PGH 转化效率的定量。误差条作为三个独立实验的标准差。。

Note: A is time- and concentration-dependent intracellular transformation of PGH in *E. gracilis* cells. Cell were treated with different concentrations of PGH protein as indicated for 2, 7 or 12 hours before observing under microscope. Scale bar: 50 μ m; B is quantification of PGH transformation efficiency in *E. gracilis* cells. Error bars are given as the SD of three independent experiments.

3.3 Pep-1 的位置转化效率的影响

除了上述 PGH 蛋白外，我们还纯化了另外三种与 Pep-1 相关的融合蛋白。如上所述，这些蛋白仅在 Pep-1 和 His 标签的位置上有所不同。纯化这些蛋白的目的是研究 Pep-1 的位置对转化效率的影响。事实上，在将这三种蛋白与上述 PGH 蛋白分别添加到纤细裸藻细胞后，我们注意到 PGH 和 HPG 的转化效率远高于 HGP 和 GPH（图 4-6）。这一结果表明 N-端 Pep-1 更有效地促进了细胞内转化。

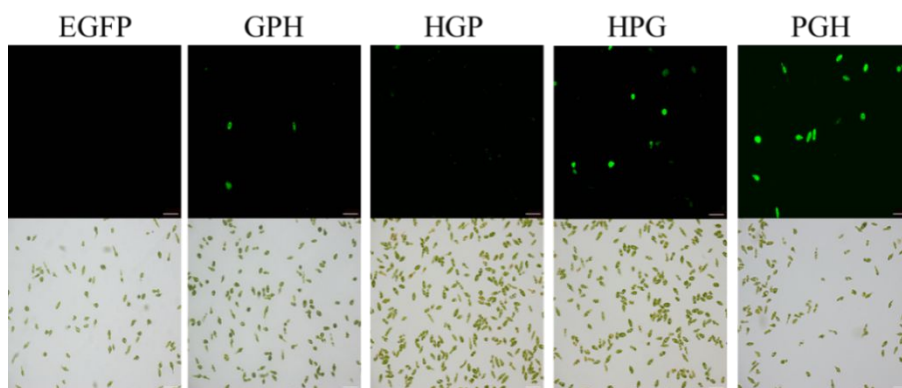


图 4-6 四种 Pep-1 融合蛋白的细胞内转化，其中 Pep-1 标签位于各自蛋白的 N 端或 C 端。标尺：50 μ m

Fig.4-6 Intracellular transformation of four Pep-1 fusion proteins with Pep-1 tag located at either the N- or C-terminus of respective protein. Scale bar: 50 μ m

3.4 二甲亚砜 (DMSO) 介导的细胞内传导

如上所述，较高的温度 (37 $^{\circ}$ C 比 25 $^{\circ}$ C) 促进了 PGH 的转化。由于较高的温度增加了细胞膜的流动性，我们推测影响膜流动性的试剂也可能会影响外源物质的转化。因此，我们利用 DMSO 来验证这个想法，DMSO 是一种常用的实验室试剂，可以增强膜的渗透性。用不同浓度的 DMSO 处理纤细裸藻细胞 1 小时。PBS 洗涤去除 DMSO 后，加入不含 Pep-1 标记的 GFP 蛋白，孵育 12 小时。如图 4-7 所示，GFP 在纤细裸藻细胞中的转化表现出 DMSO 浓度依赖的方式，表明其作用与

Pep-1 相似，当 DMSO 浓度为 10% 时，我们观察到约 80% 的转化效率。

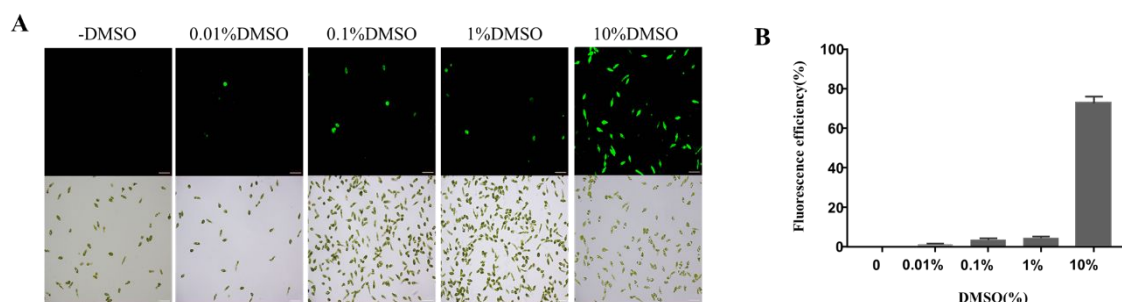


图 4-7 由 DMSO 介导的外源蛋白在纤细裸藻细胞内的转化

Fig.4-7 Intracellular transformation of exogenous protein in *E. gracilis* cells with DMSO

注：A 为不同浓度 DMSO 处理纤细裸藻细胞后进行无 Pep-1 标记的 GFP 蛋白的转化。标尺：50 μ m；B 为 GFP 转化效率的定量化。误差条作为三个独立实验的标准差。

Note: A is transformation efficiency of Pep-1 tag-free GFP in *E. gracilis* cells treated with different concentrations of DMSO. Scale bar: 50 μ m; B is quantification of GFP transformation efficiency. Error bars are given as the SD of three independent experiments.

除了上述用蛋白质转化外，我们还用核酸进行了检测。设计一对引物（见表 4-5），扩增纤细裸藻的 Prp8 基因，长度为 1.7 kb。为了像上述荧光蛋白一样直接观察该 DNA 片段在细胞内的定位，我们将荧光染料 Cy5 偶联到正向引物的 5' 端处。通过与 GFP 相同的转化过程，我们在显微镜下观察到了纤细裸藻细胞中该 DNA 片段的红色荧光信号（图 4-8A）。值得注意的是，该 DNA 片段的细胞内转化效率甚至高于上述 GFP 蛋白。最后，为了检验质粒在纤细裸藻细胞中 DMSO 转化的可行性，我们使用了一个 12 kb 的 pPZP211 质粒。DMSO 处理 1 小时后，用 PBS 洗涤去除残留的 DMSO 溶液，将终浓度 10 ng/ μ l 的质粒加入纤细裸藻细胞中。培养 3 天后，清洗细胞，使用质粒小提试剂盒提取细胞 DNA。我们从 DMSO 处理的细胞中检测到了质粒特有的大小为 736 bp 的片段。相比之下，未经 DMSO 处理的细胞中没有扩增 DNA（图 4-8B）。结果表明，12 kb 的质粒可以进入纤细裸藻细胞。

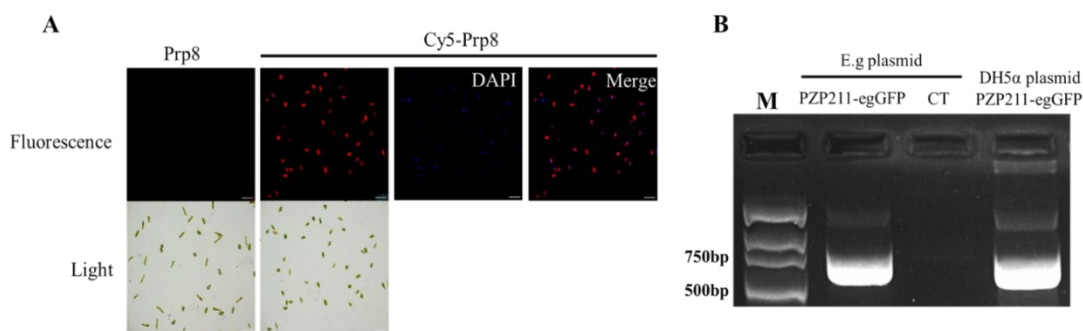


图 4-8 由 DMSO 介导的外源核酸在纤细裸藻细胞内的转化

Fig.4-8 Intracellular transformation of exogenous nucleic acids in *E. gracilis* cells with DMSO

注：A 为纤细裸藻细胞经 10% DMSO 处理后 DNA 片段的细胞内转化，将带或不带 Cy5 标记的 1.7 kb Prp8 片段以 40 ng/μl 的终浓度添加到纤细裸藻细胞中。对于 Cy5- prp8 转化的细胞，DAPI 染色的细胞核位置也显示出来，并与 Cy5 荧光信号叠加。标尺：50 μm；B 为用 PCR 检测 pPZP211 质粒的胞内转化。仅在 10% DMSO 处理后纤细裸藻细胞中检测到了 736 bp 大小的目的条带。大肠杆菌的质粒 PCR 作为阳性对照。

Note: A is intracellular transformation of a DNA fragment. The 1.7 kb Prp8 fragment with or without a Cy5 label was added at a final concentration of 40 ng/μl to the *E. gracilis* cells after 10% DMSO treatment. For cells with the Cy5-Prp8 transformation, locations of nuclei stained with DAPI were also shown and superimposed with the Cy5 fluorescent signals. Scale bar: 50 μm; B is examination of intracellular transformation of a pPZP211 plasmid by PCR. A band of 736 bp in size was recovered only from cells treated with 10%DMSO. PCR from the plasmid was served as a positive control.

4 结论

我们利用跨膜肽 Pep-1 和 DMSO 开发了两种全新的纤细裸藻转化方法。与现有的方法相比，本方法能够将外源物质更加高效、方便、快速的转化到纤细裸藻细胞中。

5 讨论

纤细裸藻是一种进化独特的生物，它作为模式生物，其基因组序列目前还尚未完全了解。迄今为止，还没有成熟的生化和遗传方法对纤细裸藻的基因进行深入研究。目前现有的遗传操作农杆菌介导的插入，电穿孔等方法对于阳性克隆筛选花费时间太长，而且仪器设施要求较高。为此，我们建立了另外两种快速的外

源物质导入方法。在跨膜肽 Pep-1 介导的转化中我们发现，对于人类细胞，Pep-1 融合蛋白浓度为 0.25 μM 就会产生较强的荧光信号^[60]，而对于纤细裸藻细胞，蛋白浓度要高达 100 μM 才会有明显效果。至今尚不清楚为什么纤细裸藻细胞需要如此高的浓度来进行有效的蛋白质转化。与人类细胞相比，纤细裸藻细胞的细胞膜内侧有一种独特的致密蛋白结构，称为 pellicle^[61]。我们推测 pellicle 可能会阻碍 Pep-1 的进入，只有较高浓度的 Pep-1 才能打破这一屏障。若如此，则 pellicle 的作用被抑制后，低浓度的 PGH 蛋白的转化效率可能会提高到与人类相似的水平。由于 pellicle 的分子组成还不清楚，所以目前为止还不能检验这个假设。综上所述我们得出影响膜流动性或膜间隙的因素，都能够影响外源物质进入纤细裸藻细胞，而 DMSO 试剂常被用作药物递送的载体，具有较好的膜穿孔作用。因此我们尝试使用不同浓度的 DMSO 作用于纤细裸藻细胞，在 10% DMSO 作用细胞 1 小时后，细胞处于应激状态，暂时丧失运动能力，将含有目的蛋白的 HUT 培养基培养细胞 12h 后，观察到荧光，再加入更多的 HUT 培养基继续进行培养，2-3 天细胞恢复正常运动，说明 10% DMSO 只会对细胞膜造成打孔作用，不会造成细胞死亡，而且纤细裸藻细胞膜具有自我修复能力。在研究 Pep-1 在蛋白不同位置对转化效率的影响时，我们发现 N 端 Pep-1 更有效地促进了细胞内转化，这与之前的报道是一致的^[62]。

与现有的农杆菌介导核转化、电穿孔和显微注射等方法相比，我们目前开发的 Pep-1 和 DMSO 介导的转化方法具有快速、简便、高效的特点。目前这些方法我们只测试了纤细裸藻，对于在其他藻类中进行大分子的细胞内转化，我们认为应该也是有效的，这有待进一步的验证。

全文总结

1. 我们鉴定出 166 个纤细裸藻剪接蛋白基因，确定了纤细裸藻中 U2 和 U6 snRNA 的存在。并且我们发现古虫界生物之间在剪接体组成上具有明显差异，以及纤细裸藻对目前人类剪接抑制剂均不敏感。

2. 我们鉴定出 28337 个常规和非常规内含子，并将其分为四种类型：GT-AG、GC-AG、ss-invariant 和 ss-variant。我们发现纤细裸藻 SL-RNA 的 5'ss 区域中存在与人类相似的 extended U6/5'ss 螺旋，以及在 3'ss 多嘧啶区域富含 C，这与其它物种富含 T 完全不同。

3. 我们通过虚拟筛选技术、体外蛋白纯化技术以及小分子结合实验获得了一种抑制 RNA 脱支酶 Dbr1 蛋白的小分子化合物 Ca。

4. 我们利用跨膜肽 Pep-1 和 DMSO 开发了两种全新的纤细裸藻转化方法。

参考文献

- [1] Luz Ambrósio D, Lee J H, Panigrahi A K, et al. Spliceosomal proteomics in *Trypanosoma brucei* reveal new RNA splicing factors [J]. *Eukaryot Cell*, 2009, 8(7): 990-1000.
- [2] Tessier LH, Keller M, Chan RL, et al. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena* [J]. *EMBO J*, 1991, 10(9): 2621-2625.
- [3] Ebel C, Frantz C, Paulus F, et al. Trans-splicing and cis-splicing in the colourless Euglenoid, *Entosiphon sulcatum* [J]. *Curr Genet*, 1999, 35(5): 542-550.
- [4] Russell A G, Charette J M, Spencer D F, et al. An early evolutionary origin for the minor spliceosome [J]. *Nature*, 2006, 443(7113): 863-866.
- [5] Will C L, Lührmann R. Spliceosome structure and function [J]. *Cold Spring Harb Perspect Biol*, 2011, 3(7): a003707.
- [6] Fica S M. Cryo-EM snapshots of the human spliceosome reveal structural adaptations for splicing regulation [J]. *Curr Opin Struct Biol*, 2020, 65: 139-148.
- [7] Townsend C, Leelaram M N, Agafonov D E, et al. Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation [J]. *Science*, 2020, 370(6523): eabc3753.
- [8] Wilkinson M E, Fica S M, Galej W P, et al. Structural basis for conformational equilibrium of the catalytic spliceosome [J]. *Mol Cell*, 2021, 81(7): 1439-1452.
- [9] Bai R, Wan R, Wang L, et al. Structure of the activated human minor spliceosome [J]. *Science*, 2021, 371(6535): eabg0879.
- [10] Bai R, Yuan M, Zhang P, Luo T, Shi Y, Wan R. Structural basis of U12-type intron engagement by the fully assembled human minor spliceosome [J]. *Science*, 2024, 383(6688):1245-1252.
- [11] Valdés J, Nozaki T, Sato E, et al. Proteomic analysis of *Entamoeba histolytica* in vivo assembled pre-mRNA splicing complexes [J]. *J Proteomics*, 2014, 111: 30-45.
- [12] Barbosa R L, Da Cunha J P C, Menezes A T, et al. Proteomic analysis of *Trypanosoma cruzi* spliceosome complex [J]. *J Proteomics*, 2020, 223: 103822.
- [13] Doetsch N A, Favreau M R, Kuscuoglu N, et al. Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic psbK operon [J]. *Curr Genet*, 2001, 39(1): 49-60.
- [14] Schwartzbach S D, Shigeoka S. *Euglena: Biochemistry, Cell and Molecular Biology Preface* [M]//SCHWARTZBACH S D, SHIGEOKA S. *Euglena: Biochemistry, Cell and Molecular Biology*, 2017: VII-VIII.

- [15] Ebenezer T E, Zoltner M, Burrell A, et al. Transcriptome, proteome and draft genome of *Euglena gracilis* [J]. BMC Biol, 2019, 17(1): 11.
- [16] Frantz C, Ebel C, Paulus F, et al. Characterization of trans-splicing in Euglenoids [J]. Curr Genet, 2000, 37(6): 349-355.
- [17] Preußer C, Jaé N, Bindereif A. mRNA splicing in trypanosomes [J]. Int J Med Microbiol, 2012, 302(4-5): 221-224.
- [18] Mcwatters D C, Russell A G. *Euglena* Transcript Processing [J]. Adv Exp Med Biol, 2017, 979: 141-158.
- [19] Keller M, Tessier L H, Chan R L, et al. In *Euglena*, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated [J]. Nucleic Acids Res, 1992, 20(7): 1711-1715.
- [20] Breckenridge D G, Watanabe Y, Greenwood S J, et al. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis* [J]. Proc Natl Acad Sci U S A, 1999, 96(3): 852-856.
- [21] Charette J M, Gray M W. U3 snoRNA genes are multi-copy and frequently linked to U5 snRNA genes in *Euglena gracilis* [J]. BMC Genomics, 2009, 10: 528.
- [22] Yoshida Y, Tomiyama T, Maruta T, et al. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions [J]. BMC Genomics, 2016, 17: 182.
- [23] Hegele A, Kamburov A, Grossmann A, et al. Dynamic protein-protein interaction wiring of the human spliceosome [J]. Mol Cell, 2012, 45(4): 567-580.
- [24] Wahl M C, Lührmann R. SnapShot: Spliceosome Dynamics I [J]. Cell, 2015, 161(6): 1474-e1.
- [25] Kastner B, Will C L, Stark H, et al. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes [J]. Cold Spring Harb Perspect Biol, 2019, 11(11):a032417.
- [26] Zhan X, Lu Y, Zhang X, et al. Mechanism of exon ligation by human spliceosome [J]. Mol Cell, 2022, 82(15): 2769-2778.
- [27] Tschudi C, Krainer A R, Ullu E. The U6 small nuclear RNA from *Trypanosoma brucei* [J]. Nucleic Acids Res, 1988, 16(23): 11375.
- [28] Cretu C, Agrawal A A, Cook A, et al. Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds [J]. Mol Cell, 2018, 70(2): 265-273.
- [29] Finci L I, Zhang X, Huang X, et al. The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action [J]. Genes Dev, 2018, 32(3-4): 309-320.
- [30] Kelso S, O'Brien S, Kurinov I, et al. Crystal structure of the CDK11 kinase domain bound to the small-molecule inhibitor OTS964 [J]. Structure, 2022, 30(12): 1615-1625.

- [31] Lin A, Giuliano C J, Palladino A, et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials [J]. *Sci Transl Med*, 2019, 11(509):8412.
- [32] Hluchý M, Gajdušková P, Ruiz De Los Mozos I, et al. CDK11 regulates pre-mRNA splicing by phosphorylation of SF3B1 [J]. *Nature*, 2022, 609(7928): 829-834.
- [33] Wan R, Bai R, Zhan X, et al. How Is Precursor Messenger RNA Spliced by the Spliceosome? [J]. *Annu Rev Biochem*, 2020, 89: 333-358.
- [34] Lasda E L, Blumenthal T. Trans-splicing [J]. *Wiley Interdiscip Rev RNA*, 2011, 2(3): 417-434.
- [35] Milanowski R, Gumińska N, Karnkowska A, et al. Intermediate introns in nuclear genes of euglenids - are they a distinct type? [J]. *BMC Evol Biol*, 2016, 16: 49.
- [36] Gumińska N, Zakryś B, Milanowski R. A New Type of Circular RNA derived from Nonconventional Introns in Nuclear Genes of Euglenids [J]. *J Mol Biol*, 2021, 433(3): 166758.
- [37] Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications [J]. *BMC Bioinformatics*, 2009, 10: 421.
- [38] Hallick R B, Hong L, Drager R G, et al. Complete sequence of *Euglena gracilis* chloroplast DNA [J]. *Nucleic Acids Res*, 1993, 21(15): 3537-3544.
- [39] Dobáková E, Flegontov P, Skalický T, et al. Unexpectedly Streamlined Mitochondrial Genome of the Euglenozoan *Euglena gracilis* [J]. *Genome Biol Evol*, 2015, 7(12): 3358-3367.
- [40] Schnare M N, Cook J R, Gray M W. Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis* [J]. *J Mol Biol*, 1990, 215(1): 85-91.
- [41] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences [J]. *Bioinformatics*, 2006, 22(13): 1658-1659.
- [42] Slater G S, Birney E. Automated generation of heuristics for biological sequence comparison [J]. *BMC Bioinformatics*, 2005, 6: 31.
- [43] Shen W, Le S, Li Y, et al. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation [J]. *PLoS One*, 2016, 11(10): e0163962.
- [44] Crooks G E, Hon G, Chandonia J M, et al. WebLogo: a sequence logo generator [J]. *Genome Res*, 2004, 14(6): 1188-1190.
- [45] Lorenz R, Bernhart S H, Höner Zu Siederdisen C, et al. ViennaRNA Package 2.0 [J]. *Algorithms Mol Biol*, 2011, 6: 26.
- [46] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. *Nucleic Acids Res*, 1999, 27(2): 573-580.
- [47] Gao P, Zhong Y, Sun C. Transcriptomic and genomic identification of spliceosomal genes

- p>from
- Euglena gracilis*
- [J].
- Acta Biochim Biophys Sin (Shanghai)*
- , 2023, 55(11): 1740-1748.
- [48] Bertram K, Agafonov D E, Dybkov O, et al. Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation [J]. *Cell*, 2017, 170(4): 701-713.
- [49] Graber J H, Salisbury J, Hutchins L N, et al. *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing [J]. *RNA*, 2007, 13(9): 1409-1426.
- [50] Zhang Z, Kumar V, Dybkov O, et al. Cryo-EM analyses of dimerized spliceosomes provide new insights into the functions of B complex proteins [J]. *EMBO J*, 2024.
- [51] Mohanta A, Chakrabarti K. Dbr1 functions in mRNA processing, intron turnover and human diseases [J]. *Biochimie*, 2021, 180: 134-142.
- [52] Montemayor E J, Katolik A, Clark N E, et al. Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1 [J]. *Nucleic Acids Res*, 2014, 42(16): 10845-10855.
- [53] Zheng S, Vuong B Q, Vaidyanathan B, et al. Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA [J]. *Cell*, 2015, 161(4): 762-773.
- [54] Wang H, Hill K, Perry S E. An Arabidopsis RNA lariat debranching enzyme is essential for embryogenesis [J]. *J Biol Chem*, 2004, 279(2): 1468-1473.
- [55] Zhou S, Ni S, Dai J, et al. Natural hybridization between *Phyllagathis* and *Sporoxeia* species produces a hybrid without reproductive organs [J]. *PLOS One*, 2020, 15(1): e0227625.
- [56] Grimm P, Risse J M, Cholewa D, et al. Applicability of *Euglena gracilis* for biorefineries demonstrated by the production of α -tocopherol and paramylon followed by anaerobic digestion [J]. *J Biotechnol*, 2015, 215: 72-79.
- [57] Chen Z, Zhu J, Du M, et al. A Synthetic Biology Perspective on the Bioengineering Tools for an Industrial Microalga: *Euglena gracilis* [J]. *Front Bioeng Biotechnol*, 2022, 10: 882391.
- [58] Hill J, Donald K A, Griffiths D E. DMSO-enhanced whole cell yeast transformation [J]. *Nucleic Acids Res*, 1991, 19(20): 5791.
- [59] Morris M C, Depollier J, Mery J, et al. A peptide carrier for the delivery of biologically active proteins into mammalian cells [J]. *Nat Biotechnol*, 2001, 19(12): 1173-1176.
- [60] Choi S H, Kim S Y, An J J, et al. Human PEP-1-ribosomal protein S3 protects against UV-induced skin cell death [J]. *FEBS Lett*, 2006, 580(30): 6755-6762.
- [61] Leander B S, Esson H J, Breglia S A. Macroevolution of complex cytoskeletal systems in euglenids [J]. *Bioessays*, 2007, 29(10): 987-1000.
- [62] Ul Ain Q, Lee J H, Woo Y S, et al. Effects of protein transduction domain (PTD) selection and position for improved intracellular delivery of PTD-Hsp27 fusion protein formulations [J]. *Arch Pharm Res*, 2016, 39(9): 1266-1274.

文献综述

反式剪接研究进展

高萍蔚 1, 孙成副 2

(1.学校: 成都医学院, 地址: 四川省成都市新都区新都大道 783 号, 邮编: 610500; 2.通讯作者: 孙成副)

摘要: 反式剪接是将两个不同的基因连接到一起的剪接类型。反式剪接包括两种类型, 一种是两个前体 mRNA(pre-mRNA)间的连接, 另一种是由剪接前导序列(spliced leader, SL)介导。在 SL 介导的反式剪接中, SL RNA 外显子上存在 5'剪接位点, pre-mRNA 上存在分支点腺苷(branch point adenosine, BP-A)和 3'剪接位点。在剪接体的作用下, SL RNA 的 5'端外显子替换掉了 pre-mRNA 的 5'端 outtron 序列, 形成成熟 mRNA。SL 介导的反式剪接主要存在于布氏锥虫、秀丽隐杆线虫等病原生物中, 而在人类和酵母中是不存在的。随着在更多的物种中检测到了 SL 反式剪接, 不仅加深了我们对其分子机制的理解, 还为未来对病原生物的研究奠定了重要基础。

关键词: 反式剪接; SL-RNA; 剪接机制; 病原生物; 纤细裸藻

Abstract: *Trans*-splicing is the joining of two different genes together, one between two precursor mRNAs and the other mediated by a spliced leader (SL) sequence, with a 5' splice site on the exon of the SL RNA and a branch point adenosine (BP-A) and a 3' splice site on the pre-mRNA. Under the action of the spliceosome, the exon at the 5' end of the SL RNA replaces the outtron sequence at the 5' end of the pre-mRNA to form the mature mRNA. Most research on *trans*-splicing by SL has found that it is mainly present in pathogenic organisms such as *T. brucei* and *C. elegans* and is absent in humans and yeast. With the detection of SL *trans*-splicing in more species, it not only deepens our understanding of its molecular mechanism but also lays an important foundation for future studies on pathogenic organisms.

Key words: *Trans*-splicing; SL-RNA; Splicing mechanism; Pathogenic organisms; *Euglena gracilis*

1 前言

真核生物中大多数编码蛋白质的基因均表现为外显子和内含子交替的中断结构。为了产生功能性的 mRNA，编码基因转录后，首先必须从初级转录物 (pre-mRNA) 中去除内含子，随后再进行 5'端加帽、3'端加聚腺苷尾，最终生成可翻译的成熟 mRNA。在整个过程中内含子的精确切除和相邻外显子的连接是一个复杂而精密的步骤，这一过程通常被称为剪接。一个 pre-mRNA 分子中所有外显子间的内含子的去除，称为顺式剪接。这个过程发生在两步机制中，每一步都包括一个酯交换反应。但在早期也发现了一种新型的 RNA 剪接形式，称为反式剪接。体外研究表明，在复杂真核生物的一些 pre-mRNA 中，两个不同分子在剪接体依赖的形式下通过反式剪接将外显子相互精确连接^[1,2]。在不同种属的物种中尤其是低等真核生物，已经报道了几种不同形式的反式剪接^[3,4,5]。反式剪接主要分为两类：基因间的反式剪接和剪接前导序列 (spliced leader, SL) 介导的反式剪接。基因间的反式剪接的外显子可能来自同一基因的不同 pre-mRNA，也可能来自不同基因或基因间区域的转录本，甚至包括来自不同染色体的转录本^[6]。而在 SL 介导的反式剪接中，一个短外显子来自一个特殊的较长的 RNA 分子的 5'端，由剪接体包括 snRNA (U2、U4、U5 和 U6，但不包括 U1) 介导^[7]，将其连接到 pre-mRNA 的 5'端，从而成为该 mRNA 的第一个外显子。另外，这个短外显子会被剪接到不同基因的 mRNA 的 5'端，使大多数 mRNA 具有共同的前导序列^[8]。

在低等真核生物中，SL 介导的反式剪接在成熟 mRNA 加工过程中起关键作用^[9]，尤其是多顺反子转录单位^[10]。在锥虫的 RNA 剪接过程中，一个短前导序列被转移到表面糖蛋白的 pre-mRNA 的 5'端^[11,12]。另外，在秀丽隐杆线虫的肌动蛋白 mRNA 的 5'端也发现了一个 22 nt 的前导序列^[13]。此外，SL 反式剪接还参与秀丽隐杆线虫的生长恢复^[14]和海洋脊索动物的营养依赖性翻译控制^[15]。然而，SL 介导的反式剪接的功能、进化和潜在机制仍然未知。在这篇综述中，我们讨论了 SL 介导的反式剪接发生的分子机制，探讨了 SL 反式剪接在秀丽隐杆线虫等病原生物以及纤细裸藻中的研究进展。

2 SL 反式剪接的分子机制

已知 SL 介导的反式剪接广泛存在于许多后生动物中，所有 SL 反式剪接都有一段称为 SL RNA 的序列，其外显子连接到许多不同的 pre-mRNA 的 5'端。SL RNA 分子类似于参与常规剪接的 U1、U2、U4 和 U5 snRNAs，它们都是以 snRNPs(小核核糖核蛋白复合物)形式存在。与其它 snRNA 一样，SL RNA 5'端形成具有高甲基化的 m^{-2,2,7}GpppN 三甲基鸟苷(TMG)帽子结构，也有一定可能性形成分子内碱基配对的茎环结构^[16]。SL RNA 的全长大小在不同的生物体中有所不同，从 45 到 140 nt 不等，SL RNA 外显子的大小也不相同，从被囊动物玻璃海鞘的 16 nt^[17]到扁形虫的 51 nt^[18]。

常规的 pre-mRNA 的 5'末端含有一个外显子的，而在发生 SL 反式剪接的基因中其 pre-mRNA 的 5'端有一个类似内含子特征的 RNA 区域（例如类似的 GC 含量），该序列被称之为“outtron”^[19]。Outtron 序列位于基因剪接位点上游，含有与 3'反式剪接位点相关的分支点腺苷和多聚嘧啶序列。虽然目前 SL 反式剪接基因的启动子或 outtron 起始位点已被确定，但由于 SL 反式剪接发生的效率相对较高，很难分离出含有 outtron 的 pre-mRNA，因此很少有 outtron 序列被分析出来^[20,21]。

在 SL 介导的反式剪接过程中，进行 SL 反式剪接的 pre-mRNA 的 outtron 上只有 3'剪接位点，上游没有 5'剪接位点，这个 3'剪接位点与位于 SL RNA 分子上的 5'剪接位点进行配对。研究表明，在某些体系中，通过添加或去除上游的 5'剪接位点，SL 介导的反式和常规顺式的 3'剪接位点可以相互转换^[22,23]。由于该基因上缺乏 5'剪接位点，因此是 SL RNA 上的 5'剪接位点结合到基因 5'端 outtron 上的分支点腺苷（A）上，形成中间体。该过程中 pre-mRNA 不能与 U1 snRNP 结合，但有 SL snRNP 参与，它与 U2 snRNP 在基因上的 3'剪接位点相互作用，SL 外显子被剪接后连接到 mRNA 上成为第一个外显子。SL snRNP 是如何在 pre-mRNA 上游没有 5'剪接位点的情况下被吸引到 pre-mRNA 的 3'剪接位点上的，这个问题现在仍然是未知的。目前出现两种猜测，第一种是上游没 U1 snRNP 时，U2 snRNP 与 3'剪接位点相互作用并吸引 SL snRNP；第二种可能是，SL snRNP 被吸引到反式剪接体上是由于对 pre-mRNA 的 outtron 序列 5'末端的核帽结合复合物具有亲和力，或者是由于对 RNA 聚合酶 II C 端结构域具有亲和力^[24]。此外，有些生物会

使用同一剪接位点交替进行 SL 介导的反式和常规顺式剪接。在锥虫^[25]和秀丽隐杆线虫^[26]中就有这种情况。

3 SL 反式剪接在病原生物中的研究

许多原生生物的 SL 反式剪接依赖于剪接体 snRNA 和相关蛋白对反式剪接位点序列的准确识别,进而将 SL RNA 的外显子序列取代 mRNA 的 5'端。不同生物的 SL 序列具有很强的保守性^[27],剪接位点的识别在不同生物中也普遍保守,但是仍存在一定的差异。

3.1 秀丽隐杆线虫

在秀丽隐杆线虫中大约 70%的基因的转录本是经过 SL 反式剪接的。在该生物中存在两种 SL RNA,即 SL1 或 SL2。虽然 SL1 与 SL2 RNA 均存在于生物体内,但并不是同时被剪接到一条编码链上的。

其中大多数基因进行的是由 SL1 介导的反式剪接,有研究报道,62%的 mRNA 上存在一个特定的 SL1,SL1 取代了 mRNA 初始转录序列的 5'非翻译区(5'UTR)。由于 5'UTR 在调节翻译效率中的关键作用,并且 SL1 的剪接位置非常靠近起始的甲硫氨酸密码子^[28],因此 SL1 似乎在翻译起始中发挥作用。此外,有研究也阐明了 SL1 介导的反式剪接能有效的提高线虫的翻译效率。该研究发现 SL1 介导的反式剪接可能是通过缩短 5'UTR,从而减少上游起始密码子(uAUG)的存在并削弱 mRNA 的二级结构^[29],证明了 SL1 反式剪接在提高必需基因的翻译效率方面起着关键作用。

在秀丽隐杆线虫中,SL2 占有所有 SL 反式剪接 RNA 的 7%^[30]。SL1 外显子主要与带有 outtron 的基因相互连接,相比之下,SL2 则用于介导操纵子转录本,其几乎完全被反式剪接到由操纵子产生的多顺反子 pre-mRNA 的下游^[31]。SL2 反式剪接的发生是为了在全基因组范围内分离操纵子的基因产物。顺反子间区距离越短,SL2 的使用越多。综上所述,在秀丽隐杆线虫中,SL1 和 SL2 的反式剪接使用了不同的潜在机制。

3.2 布氏锥虫

锥虫是早期分化的原生生物寄生虫,其中布氏锥虫、克氏锥虫在人类中引起

严重的、致命的疾病。锥虫对全球健康的影响促使了人们对其的研究，为了更好地对抗这些寄生虫，它们的分子生物学一直是多年来的研究重点。在布氏锥虫体内所有的 mRNA 分子都是两个序列的嵌合体，其 5'端的前 39 个核苷酸对细胞内所有 mRNA 都是共同的^[32]。SL 反式剪接是布氏锥虫中所有核 pre-mRNA 成熟的必要步骤，因此，布氏锥虫仍然是 SL 反式剪接研究的首选生物体^[33]。对布氏锥虫的研究发现，其体内进行 SL 反式剪接也需要完整的 U2 和 U4/U6 snRNP^[34]。

在布氏锥虫 SL RNA 中，修饰的核苷酸聚集在 5'末端，包含了特殊的 cap4 结构，由 7-甲基鸟苷与四个修饰的核苷相连。在 RNA 成熟过程中，反式剪接将 SL RNA 外显子包括 cap4 结构转移到所有 mRNA 的 5'端。有研究表明，将可渗透的锥虫细胞暴露于 s-腺苷-L-同型半胱氨酸中可抑制新合成的 SL RNA 的 7-甲基鸟苷邻近核苷的甲基化，并阻止 SL RNA 在反式剪接中的利用。但是其并没有抑制细胞中先前存在的 SL RNA 的反式剪接^[35]。与对照细胞相比较，新合成的 SL RNA 似乎是反式剪接机制中唯一已知的因甲基化抑制而功能受损的成分，也进一步说明了 SL RNA 的 cap4 结构在反式剪接过程中起着至关重要的作用。SL 序列的 5'修饰似乎是通过锥虫科的进化而得到了保护，如两种锥虫科的 SL RNA，即布氏锥虫和 *Leptomonas collosoma*，带有类似的帽结构^[36]。

4 在纤细裸藻中的研究现状

虽然目前已知进行 SL 反式剪接的裸藻类物种大约有 1000 个，但我们在分子水平上的大部分知识几乎都来自于对纤细裸藻的研究。通过 cDNA 序列以及杂交捕获技术，验证了纤细裸藻细胞质 mRNA 的 5'端存在且极其相似的短序列。通过克隆纤细裸藻细胞核 DNA 发现该短序列不存在于其核基因序列的 5'端，而是位于一些小 RNA 的 5'末端^[37]。因此研究提出纤细裸藻细胞体内也广泛存在 SL 介导的反式剪接。纤细裸藻中 SL 反式剪接的存在为其进化关系与锥虫较近提供了强有力的论据^[38]。

纤细裸藻是 SL 反式与常规顺式剪接机制共存的独特生物，通过剪接体催化将 SL-RNA 转移到绝大多数 mRNA 上。研究发现在纤细裸藻中 26 个核苷酸组成的 SL 外显子，构成了 SL RNA 的 5'端。一个与其他 snRNA 类似的 Sm 结合位点

构成了 SL RNA 的 3'端。SL-RNA 长度约为 100 个核苷酸，可以折叠成一个三茎环结构^[39]。这些 SL RNA 不像锥虫那样是唯一的，纤细裸藻中 SL RNA 下游序列不完全相同，是一个多基因家族。SL-RNAs 的多样性（SLa/SLb）促使了对纤细裸藻 SL-RNAs 的编码基因的研究。

纤细裸藻的前导序列 SL 位于编码 5S rRNA 基因的重复单元中^[40]，SL-RNA 和 5S rRNA 基因在串联重复单位中的关联也在线虫中被发现。在纤细裸藻中发现大多数 SL-RNA 与细胞质 5S rRNA 基因以 0.6 kb 串联重复单位相关^[41]。这种串联重复的现象也存在于其他物种中，比如秀丽隐杆线虫 SL1-RNA 基因位于重复 100 次的 1kb 单元中，并且这个区域也含有 5S rRNA 基因^[42]。然而，布氏锥虫却不存在上述现象^[43,44]。

5 讨论与展望

随着越来越多的 SL RNA 的发现，SL 介导的反式剪接领域在传统研究上逐渐扩大。现在已经在许多不同的物种中发现 SL 反式剪接的存在，毫无疑问，由于在基因组和转录组水平上测序的物种大量增加，SL 反式剪接或许将在更多的物种中被发现。此外，基因反式剪接的例子也可能随着测序深度的增加而增加。

SL 介导的反式剪接的功能及原理上仍然有未解决的问题，例如在大多数生物中并不是每个基因都会进行 SL 反式剪接，而是在特定基因上，其中机制尚不清楚。含有 5'剪接位点的 SL RNA 是如何接近 pre-mRNA 上 3'剪接位点的，这个问题也需要进一步探讨。

参考文献

- [1] Konarska M M, Padgett R A, Sharp P A. Trans splicing of mRNA precursors in vitro [J]. Cell, 1985, 42(1): 165-171.
- [2] Solnick D. Trans splicing of mRNA precursors [J]. Cell, 1985, 42(1): 157-164.
- [3] Murphy W J, Watkins K P, Agabian N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing [J]. Cell, 1986, 47(4): 517-525.
- [4] Rajkovic A, Davis R E, Simonsen J N, et al. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni* [J]. Proc Natl Acad Sci U S A, 1990, 87(22): 8879-8883.
- [5] Sutton R E, Boothroyd J C. Evidence for trans splicing in trypanosomes [J]. Cell, 1986, 47(4): 527-535.
- [6] Gingeras T R. Implications of chimaeric non-co-linear transcripts [J]. Nature, 2009, 461(7261): 206-211.
- [7] Hannon G J, Maroney P A, Nilsen T W. U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro [J]. J Biol Chem, 1991, 266(34): 22792-22795.
- [8] Van Doren K, Hirsh D. Trans-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *Caenorhabditis elegans* [J]. Nature, 1988, 335(6190): 556-559.
- [9] Hastings K E. SL trans-splicing: easy come or easy go? [J]. Trends Genet, 2005, 21(4): 240-247.
- [10] Nilsen T W. Trans-splicing of nematode premessenger RNA [J]. Annu Rev Microbiol, 1993, 47: 413-440.
- [11] Boothroyd J C, Cross G A. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5' end [J]. Gene, 1982, 20(2): 281-289.
- [12] Van Der Ploeg L H, Liu A Y, Michels P A, et al. RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes [J]. Nucleic Acids Res, 1982, 10(12): 3591-3604.
- [13] Krause M, Hirsh D. A trans-spliced leader sequence on actin mRNA in *C. elegans* [J]. Cell, 1987, 49(6): 753-761.
- [14] Zaslaver A, Baugh L R, Sternberg P W. Metazoan operons accelerate recovery from growth-arrested states [J]. Cell, 2011, 145(6): 981-992.
- [15] Danks G B, Raasholm M, Campsteijn C, et al. Trans-splicing and operons in metazoans: translational control in maternally regulated development and recovery from growth arrest [J]. Mol Biol Evol, 2015, 32(3): 585-599.
- [16] Blumenthal T, Steward K. RNA Processing and Gene Structure [M]. *C elegans II*. Cold Spring

- Harbor (NY), Cold Spring Harbor Laboratory Press
- [17] Vandenbergh A E, Meedel T H, Hastings K E. mRNA 5'-leader trans-splicing in the chordates [J]. *Genes Dev*, 2001, 15(3): 294-303.
 - [18] Davis R E. Surprising diversity and distribution of spliced leader RNAs in flatworms [J]. *Mol Biochem Parasitol*, 1997, 87(1): 29-48.
 - [19] Conrad R, Thomas J, Spieth J, et al. Insertion of part of an intron into the 5' untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene [J]. *Mol Cell Biol*, 1991, 11(4): 1921-1926.
 - [20] Park Y S, Kramer J M. Tandemly duplicated *Caenorhabditis elegans* collagen genes differ in their modes of splicing [J]. *J Mol Biol*, 1990, 211(2): 395-406.
 - [21] Conrad R, Liou R F, Blumenthal T. Conversion of a trans-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site [J]. *EMBO J*, 1993, 12(3): 1249-1255.
 - [22] Maroney P A, Romfo C M, Nilsen T W. Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly [J]. *Mol Cell*, 2000, 6(2): 317-328.
 - [23] Boukis L A, Bruzik J P. Functional selection of splicing enhancers that stimulate trans-splicing in vitro [J]. *Rna*, 2001, 7(6): 793-805.
 - [24] Blumenthal T. Trans-splicing and operons [J]. *WormBook*, 2005: 1-9.
 - [25] Mair G, Shi H, Li H, et al. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA [J]. *Rna*, 2000, 6(2): 163-169.
 - [26] Allen M A, Hillier L W, Waterston R H, et al. A global analysis of *C. elegans* trans-splicing [J]. *Genome Res*, 2011, 21(2): 255-264.
 - [27] Ferguson K C, Heid P J, Rothman J H. The SL1 trans-spliced leader RNA performs an essential embryonic function in *Caenorhabditis elegans* that can also be supplied by SL2 RNA [J]. *Genes Dev*, 1996, 10(12): 1543-1556.
 - [28] Lall S, Friedman C C, Jankowska-Anyszka M, et al. Contribution of trans-splicing, 5' -leader length, cap-poly(A) synergism, and initiation factors to nematode translation in an *Ascaris suum* embryo cell-free system [J]. *J Biol Chem*, 2004, 279(44): 45573-45585.
 - [29] Yang Y F, Zhang X, Ma X, et al. Trans-splicing enhances translational efficiency in *C. elegans* [J]. *Genome Res*, 2017, 27(9): 1525-1535.
 - [30] Bernard F, Dargère D, Rechavi O, et al. Quantitative analysis of *C. elegans* transcripts by Nanopore direct-cDNA sequencing reveals terminal hairpins in non trans-spliced mRNAs [J]. *Nat Commun*, 2023, 14(1): 1229.
 - [31] Evans D, Blumenthal T. trans splicing of polycistronic *Caenorhabditis elegans* pre-mRNAs: analysis of the SL2 RNA [J]. *Mol Cell Biol*, 2000, 20(18): 6659-6667.

- [32] Laird P W. Trans splicing in trypanosomes--archaism or adaptation? [J]. Trends Genet, 1989, 5(7):204-208.
- [33] Günzl A. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? [J]. Eukaryot Cell, 2010, 9(8): 1159-1170.
- [34] Tschudi C, Ullu E. Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells [J]. Cell, 1990, 61(3): 459-466.
- [35] Perry K L, Watkins K P, Agabian N. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader [J]. Proc Natl Acad Sci U S A, 1987, 84(23): 8190-8194.
- [36] Miller S I, Landfear S M, Wirth D F. Cloning and characterization of a Leishmania gene encoding a RNA spliced leader sequence [J]. Nucleic acids research, 1986, 14(18): 7341-7360.
- [37] Tessier L H, Keller M, Chan R L, et al. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena [J]. EMBO J, 1991, 10(9): 2621-2625.
- [38] Kivic P A, Walne P L. An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida [J]. Origins Life Evol Biosphere, 13(3): 269-288.
- [39] Agabian N. Trans splicing of nuclear pre-mRNAs [J]. Cell, 1990, 61(7): 1157-1160.
- [40] Tessier L H, Chan R L, Keller M, et al. The Euglena gracilis rbcS gene contains introns with unusual borders [J]. FEBS Lett, 1992, 304(2-3): 252-255.
- [41] Keller M, Tessier L H, Chan R L, et al. In Euglena, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated [J]. Nucleic Acids Res, 1992, 20(7): 1711-1715.
- [42] Nelson D W, Honda B M. Genes coding for 5S ribosomal RNA of the nematode Caenorhabditis elegans [J]. Gene, 1985, 38(1-3): 245-251.
- [43] De Lange T, Liu A Y, Van Der Ploeg L H, et al. Tandem repetition of the 5' mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? [J]. Cell, 1983, 34(3): 891-900.
- [44] Milhausen M, Nelson R G, Sather S, et al. Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? [J]. Cell, 1984, 38(3): 721-729.

附录一

人类与纤细裸藻以及其它三种原生生物的剪接蛋白比较

Complex	Genes	<i>H. sapiens</i>	<i>E. gracilis</i> ^a	<i>T. brucei</i> ^b	<i>L. major</i>	<i>D. papillatum</i> ^c
Sm	SmB	NP_003082	OP185536	XP_951644	XP_003722006	OQ872397
	SmD1	NP_008869	OP185537	XP_845935	XP_001683257	OQ872398
	SmD2	NP_808210	OP185538	XP_951729	XP_001686081	OQ872399
	SmD3	NP_004166	OP185539 OP185540	XP_844239	XP_001686492	OQ872400 OQ872401
	SmE	NP_003085	OP185541	XP_845391	XP_001684716	OQ872402
	SmF	NP_003086	OP185542	XP_827316	XP_003722843	OQ872403
	SmG	NP_003087	OP185543	XP_829487	XP_001685423	OQ872404
LSm	LSm2	NP_067000	OP185544			OQ872405
	LSm3	NP_055278	OP185545			OQ872406
	LSm4	NP_036453	OP185546	XP_829452	XP_001685387	OQ872407
	LSm5	NP_036454	OP185547			OQ872408
	LSm6	NP_009011	OP185548			OQ872409
	LSm7	NP_057283	OP185549	XP_845073	XP_001684007	OQ872410
	LSm8	NP_057284	OP185550			OQ872411
U1 snRNP	U1-70K	NP_003080	OP185551	XP_847313	XP_001682242	OQ872412
	U1A [®]	NP_004587	OP185562 OP185563	XP_843940	XP_003722162	OQ872413 OQ872414
	U1C	NP_003084	OP185552			
	RBM25	NP_067062	OP185553			
	PRP39	NP_060392	No hit			
	PRP40	NP_060362 NP_0010268 68	OP185554 OP185555			OQ872415
	TIA1	NP_071505	OP185556			
	LUC7L	NP_060502	OP185557			
	SRPK1 [®]	NP_003128	OP185558 OP185559 OP185560	XP_845614 XP_845720	XP_001684968 XP_001684102	OQ872416 OQ872417 OQ872418 OQ872419
U2 snRNP	U2A'	NP_003081	OP185561	XP_827916	XP_001683928	OQ872420
	U2B ^{"&}	NP_003083	OP185562 OP185563	XP_843940	XP_003722162	OQ872413 OQ872414
	SF3A1	NP_005868	OP185564	XP_828057	XP_001682710	OQ872421
	SF3A2	NP_009096	OP185565	XP_951536	XP_003721618	OQ872422
	SF3A3	NP_006793	OP185566	XP_845437	XP_001684786	OQ872423
	SF3B1	NP_036565	OP185567	XP_829247	XP_001684544	OQ872424
	SF3B2	NP_006833	OP185568	XP_845321	XP_001684647	OQ872425
	SF3B3	NP_036558	OP185569	XP_846318	XP_001687566	OQ872426
	SF3B4	NP_005841	OP185570	XP_844118	XP_003722112	OQ872427
	SF3B5	NP_112577	OP185571	XP_827364	XP_003722805	OQ872428
	SF3B6	NP_057131	OP185572	XP_822981	XP_001686883	OQ872429
	PHF5A	NP_116147	OP185573	XP_822974	XP_001686871	OQ872430
U2 related	U2AF65	NP_0010124 96	OP185574			
	U2AF35	NP_0010203 74	OP185575 OP185576	XP_822572	XP_003721697	OQ872431

附录

			OP185577			
	PUF60	NP_510965	OQ397584 OQ397585			
	SPF30	NP_005862	OP185579			OQ872432
	SPF31	NP_055095	No hit			
	SPF45	NP_116294	OP185580			OQ872433
	CHERP	NP_006378	OP185581			
	U2SURP	NP_0010738 84	OP185582	XP_846348	XP_001687538	OQ872434 OQ872435
	DDX42	NP_031398	OP185583			OQ872436
U6 snRNP	SART3	NP_055521	No hit			
U4/U6 di-snRNP	PRP3	NP_004689	OP185584			OQ872437
	PRP4	NP_004688	OP185585	XP_822354	XP_001683064	OQ872438
	PRP31	NP_056444	OP185586	XP_823290	XP_001685738	OQ872439 OQ872440
	NHP2L1	NP_0010037 96	OP185587	XP_803820	XP_001682066	OQ872441
	PPIH	NP_006338	OP185588	XP_845254 XP_845256 XP_845262	XP_001684591	OQ872442
U5 snRNP	PRP8	NP_006436	OP185589	XP_827386	XP_003722790	OQ872443
	BRR2	NP_054733	OP185590	XP_844903	XP_001681794 XP_001682350	OQ872444
	SNU116	NP_004238	OP185591	XP_829601	XP_001685543	OQ872445
	PRP6	NP_036601	OP185592			OQ872446
	TXNL4A	NP_006692	OP185593	XP_847089	XP_001683400	OQ872447
	CD2BP2	NP_006101	OP185594 OP185595			
	SNRNP40	NP_004805	OP185596			OQ872448
	AAR2	NP_056326	OP185597			
U4/U6.U5 tri-nRNP	SART1	NP_005137	OP185598			
	USP39	NP_006581	OP185599			
	RBM42	NP_077297	OP185600			OQ872449
	SNRNP27	NP_006848	No hit			
	SRPK2 ⁶	NP_872634	OP185558 OP185559 OP185560	XP_845614 XP_845720	XP_001684968 XP_001684102	OQ872416 OQ872417 OQ872418 OQ872419
NTC	PRP19	NP_055317	OP185601	XP_951696	XP_003722060	OQ872450
	CDC5L	NP_001244	OP185602	XP_844880	XP_001682333	OQ872451
	SPF27	NP_005863	OP185603			
	PRL1	NP_002660	OP185604	XP_823238	XP_001687067	OQ872452
	AD002	NP_057487	OP185605			OQ872453
	HSPA8 ⁶	NP_006588	OP185606 OP185607 OQ397586 OQ397587 OQ397588 OQ397589	XP_829198	XP_001684128 XP_001684563 XP_001684568 XP_001684402	OQ872454 OQ872455 OQ872456 OQ872457
	CTNNBL1	NP_110517	OP185608			OQ872458
NTR	SKIP	NP_036377	OP185609			OQ872459
	RBM22	NP_060517	OP185610			OQ872460

	CRNKL1	NP_057736	OP185611	XP_823189	XP_001687012	OQ872461
	PPIL1	NP_057143	OP185612	XP_827807	XP_001687067	OQ872462 OQ872463
	G10	NP_003901	OP185613	XP_822907	XP_001681198	
	SYF2	NP_056299	OP185614			OQ872464
	PRCC	NP_005964	No hit			
IBC	AQR	NP_055506	OP185615			OQ872465
	XAB2	NP_064581	OP185616	XP_844808	XP_001683515	OQ872466
	ISY1	NP_065752	OP185617	XP_847026	XP_001680905	OQ872467
	ZNF830	NP_443089	No hit			
	PPIE	NP_006103	OP185618			OQ872468
RES complex	RBMX2	NP_057108	OP185619	XP_845742	XP_001684080	OQ872469
	SNIP1	NP_078976	OP185620	XP_844126	XP_003722104	OQ872470
	BUD13	NP_116114	OP185621			
EJC/TREX	eIF4AIII	NP_055555	OP185622	XP_828941	XP_001684436	OQ872471
	MAGO	NP_002361	OP185623	XP_845612	XP_001684966	OQ872472
	Y14	NP_005096	OP185624		XP_001684082	OQ872473
	MLN51	NP_031385	No hit			
	Acinus	NP_055792	No hit			
	SAP18	NP_005861	No hit			OQ872474
	THOC4	NP_005773	No hit			
	THOC1	NP_005122	OP185625			
	THOC2	NP_001075019	OP185626			
	THOC3	NP_115737	OP185627			OQ872475
	C17orf85	NP_001107590	No hit			
	Pinin	NP_002678	No hit			
	THOC7	NP_079351	OP185628			
	RNPS1	NP_006702	OP185629			
	THOC5	NP_001002877	No hit			
A specific	HTATSF1	NP_055315	OP185630			OQ872476
	SF1*	NP_004621	OP185631			
	TRIR	NP_076943	No hit			
	CCAR1*	NP_060707	OP185632 OP185633			
	CDC2L2	NP_076916	OP185634			OQ872477
	BUB3	NP_001007794	OP185635			
	SUGP1	NP_757386	OP185636			
	TLS/FUS	NP_004951	No hit			
	TCERG1	NP_006697	OP185637			
	p68/p72	NP_004387 NP_006377	OQ397590 OQ397591 OQ397592 OQ397593 OQ397594 OQ397595	XP_822905	XP_001680939 XP_001686790	OQ872478

附录

	RBM5/RB M10	NP_005769 NP_005667	No hit			
	RBM23/RB M39	NP_0010708 19 NP_004893	OP185641			OQ872479
pre-B specific	PRP4K	NP_003904	OP185642 OP185643 OP185644			OQ872480
B specific	PRP38	NP_116253 NP_060531	OP185645 OP185646			OQ872481
	ZMAT2	NP_653324	OP185647			OQ872482
	MFAP1	NP_005917	OP185648			OQ872483
	RED	NP_006074	OP185649			OQ872484
	THRAP3	NP_0013084 00	No hit			
	PQBP1	NP_0010275 53	OP185650			
	WBP11	NP_057396	No hit			
	HSP27	NP_001531	No hit			
	SMU1	NP_060695	OP185651			OQ872485
	FBP21	NP_009118	No hit			
	UBL5	NP_077268	OP185652			OQ872486
pre-Bact specific	KIN17	NP_036443	OP185653			OQ872487
Bact specific	RNF113A	NP_008909	OP185654			
	NY-CO-10	NP_005860	OP185655			OQ872488
	PPIL2	NP_055152	OQ397596 OQ397597			OQ872489
	CCDC12	NP_653317	OP185657			
	SRRM1	NP_005830	No hit			
	SRRM2	NP_057417	OP185658			OQ872490
	CWC22	NP_0013629 58	OP185659	XP_829139	XP_001687102	OQ872491
B*/C specific	GPKOW	NP_056513	OP185660			
	CCDC94	NP_060544	OP185661			
	CCDC49	NP_060218	No hit			
	PPWD1	NP_056157	OP185662			OQ872492
	PPIL4	NP_624311	OP185663	XP_845045	XP_001682143	OQ872493
	FRG1	NP_004468	No hit			
	WDR70	NP_060504	OP185664			
	WDR83	NP_115708	No hit			OQ872494
	GPATCH1	NP_060495	OP185665			OQ872495
	FAM50	NP_004690 NP_036267	OP185666			
	NOSIP	NP_057037	OP185667			
	C9orf78	NP_057604	No hit			
	CXorf56	NP_071384	No hit			
	NSRP1	NP_115517	No hit			
	CIR1/RP9	NP_004873 NP_976033	No hit			

	Abstrakt	NP_057306	OP185668			OQ872496
	PPIL3	NP_570981	OQ397598 OQ397599	XP_827430	XP_003722755	OQ872497
	PPIG	NP_004783	No hit			
pre-C*	FAM192A	NP_079222	No hit			
C*/P specific	PRP17	NP_056975	OP185671	XP_843787	XP_001683944	OQ872498
	PRP18	NP_003666	OP185672			OQ872499
	SLU7	NP_006416	OP185673	XP_844974	XP_001681147	OQ872500
	PRKRIP1	NP_078929	No hit			
	DDX35	NP_068750	OP185674			OQ872501
	FAM32A	NP_054796	OP185675			
	NKAP	NP_078804	OP185676	XP_823332	XP_001685799	OQ872502
	Cactin	NP_067054	OP185677	XP_829224	XP_001684517	OQ872503
	SDE2	NP_689821	OP185678			
C/C* misc	DDX57	NP_945314	OQ397600 OQ397601	XP_829273	XP_001681170	
	RACK1	NP_006089	OP185680	XP_829201	XP_001684560	OQ872504 OQ872505
	Matrin-3	NP_0011818 83	No hit			
	DBPA	NP_003642	OP185682			
	TOE1	NP_079353	No hit			
	RBM4	NP_002887	OP185683 OQ397602	XP_951655	XP_003722019	
	JUP	NP_0013397 02	No hit			
	HSPA1A*	NP_005336	OP185606 OP185607 OQ397586 OQ397587 OQ397588 OQ397589	XP_829198	XP_001684563 XP_001684568	
	TTC14	NP_597719	OP185684			
	ZCCHC10	NP_0012877 45	OP185685			
	CDK10	NP_443714	OP185686			OQ872506
	FRA10AC1	NP_0013346 41	OP185687			
	DGCR14	NP_073210	OP185688			OQ872507
ILS specific	TFIP11	NP_0010086 97	OP185689			
	CWF19L2	NP_689647	No hit			
ATP helicase	UAP56	NP_004631	OP185690	XP_822312	XP_001683110	OQ872508
	PRP5	NP_055644	OP185691	XP_845030	XP_001681037	OQ872509
	PRP28	NP_004809	OP185692			OQ872510
	PRP2	NP_003578	OP185693	XP_822773	XP_001686659	OQ872511
	PRP16	NP_054722	OP185694	XP_828648	XP_001683625	OQ872512
	PRP22	NP_004932	OQ397603 OQ397604 OQ397605 OQ397606	XP_822963 XP_823137	XP_001686861	OQ872513 OQ872514
	PRP43	NP_001349	OQ397607 OQ397608	XP_844789 XP_844389	XP_003722513	OQ872515
pre-mRNA/intr	CBP80	NP_002477	OP185697			

附录

on binding proteins	CBP20	NP_031388	OP185698	XP_845318	XP_001684644	OQ918096
	YBX1	NP_004550	OP185682			
	SRRT/ASR2B	NP_056992	No hit			
	RBM7	NP_057174	No hit			
	ELAVL1	NP_001410	No hit	XP_847495	XP_001683802 XP_001683803	OQ872516
	PABP1	NP_002559	OQ397609 OQ397610 OQ397611 OQ397612 OQ397613	XP_843988 XP_843985 XP_827237 XP_827358	XP_003722897	OQ872517 OQ872518 OQ872519 OQ872520
	NF45	NP_004506	No hit			
	ZC3H18	NP_653205	No hit			
	DDX3	NP_001347	OQ397614 OQ397615 OQ397616 OQ397617 OQ397618	XP_827934 XP_827494	XP_001685353	
	Dbr1	NP_057300	OP185699	XP_844887	XP_001682341	OQ872521
SR and related proteins	ASF/SF2	NP_008855	No hit			
	SRSF7	NP_001026854	OQ397619	XP_827049		
	SRSF3	NP_003008	No hit			
	SRSF9	NP_003760	No hit			
	SRSF10	NP_473357	OQ397620			
	SRSF5	NP_001034554	No hit			
	SRSF6	NP_006266	No hit			
	SRSF4	NP_005617	No hit			
	SRSF2	NP_001182356	No hit			
	hTra2a	NP_037425	No hit			
	hTra2b/SF RS10	NP_004584	OQ397621			
	ARGLU1	NP_060481	No hit			
hnRNP proteins	hnRNP A0/A1/A3/A/B/A2/B1	NP_006796 NP_112420 NP_001317178 NP_004490 NP_002128	OQ397622			
	hnRNP C1/C2	NP_001070910	No hit			
	hnRNP D	NP_112738	No hit	XP_846936		
	hnRNP F	NP_001091674	No hit	XP_951606 [§]		OQ872522
	hnRNP G	NP_002130	No hit			
	hnRNP H1/2/3	NP_001244222 NP_001027565 NP_001309367	No hit	XP_951606 [§]		OQ872523 OQ872524
	hnRNP K	NP_002131	No hit			

	hnRNP M	NP_005959	OQ397623 OQ397624			OQ872525
	hnRNP Q	NP_006363	No hit			
	hnRNP R	NP_005817	OQ397625 OQ397602			
	hnRNP U	NP_114032	No hit			
	RALY	NP_057951	No hit			
	E1B-AP5	NP_008971	OQ397626 OQ397627			OQ872526
	PTBP1/2	NP_002810 NP_0012879 14	OQ397628 OQ397629 OQ397630 OQ397631 OQ397632 OQ397633	XP_827198	XP_888628	OQ872527 OQ872528
	PCBP1/2	NP_006187 NP_005007	No hit			
miscellaneous proteins	CFAP20	NP_037374	OP185700	XP_822473	XP_001682971	OQ872529
	CCAR2*	NP_0013809 26	OP185632 OP185633			
	NFAR	NP_0013817 40	No hit			
	ZNF207	NP_003448	OQ397634			OQ872530
	NRIP2	NP_113662	No hit			
	SAP30BP	NP_037392	No hit			
	DDX9	NP_001348	No hit			
	DNAJC6	NP_055602	No hit			
	PPP1CA	NP_002699	OQ397635 OQ397636	XP_844512	XP_001686170 XP_001686172 XP_001686173	OQ872531 OQ872532
	GCFC	NP_057715	No hit			
	BAG2	NP_004273	No hit			
	CIRP	NP_001271	OP185702	XP_845656	XP_001685908	
	NIPP1	NP_054829	OQ397637	XP_845871	XP_001683205	OQ872533
	SMN1	NP_000335	No hit			
	CLNS1A	NP_001284	No hit			
	MEP50	NP_077007	No hit			
	PRMT5	NP_006100	OP185704	XP_822322	XP_001683098	OQ872534
	SEC31B	NP_056305	No hit			
	RBBP6	NP_008841	OP185705	XP_827503	XP_003722699	OQ918097
	AGGF1	NP_060516	No hit			
	CELF1	NP_0013633 71	OP279271			
	RBFOX2	NP_0010760 48	No hit			
	Quaking/Sa m68/Slm-2 &	NP_006766 NP_006550 NP_006549	OP185631			

附录二

纤细裸藻内含子的选择性剪接类型

AS-type [↗]	Genomic accession [↗]	Transcriptomic accession [↗]	intron size (bp) [↗]
Overlapping [↗]	CAAHFI010095568.1 [↗]	m64269e_220926_224707/43059924/ccs [↗]	507 [↗]
		m64269e_220926_224707/47121973/ccs [↗]	422 [↗]
	CAAHFI010245853.1 [↗]	m64269e_220926_224707/39389135/ccs [↗]	244 [↗]
		m64269e_220926_224707/105775799/ccs [↗]	181 [↗]
	CAAHFI010365539.1 [↗]	m64269e_220926_224707/130025364/ccs [↗]	1275 [↗]
		m64269e_220926_224707/24773118/ccs [↗]	1214 [↗]
	CAAHFI010548366.1 [↗]	m64269e_220926_224707/38732698/ccs [↗]	1056 [↗]
		m64269e_220926_224707/146606680/ccs [↗]	1001 [↗]
	CAAHFI011060464.1 [↗]	m64269e_220926_224707/106299553/ccs [↗]	455 [↗]
		m64269e_220926_224707/77400317/ccs [↗]	428 [↗]
	CAAHFI011991134.1 [↗]	m64269e_220926_224707/59966605/ccs [↗]	632 [↗]
		m64269e_220926_224707/49941311/ccs [↗]	586 [↗]
	CAAHFI010356404.1 [↗]	m64269e_220926_224707/132121167/ccs [↗]	425 [↗]
		m64269e_220926_224707/98501514/ccs [↗]	382 [↗]
	CAAHFI010783765.1 [↗]	m64269e_220926_224707/89589654/ccs [↗]	314 [↗]
		m64269e_220926_224707/90114539/ccs [↗]	267 [↗]
	CAAHFI011214082.1 [↗]	m64269e_220926_224707/79364138/ccs [↗]	606 [↗]
		m64269e_220926_224707/151063201/ccs [↗]	538 [↗]
	CAAHFI011234586.1 [↗]	m64269e_220926_224707/5636690/ccs [↗]	626 [↗]
		m64269e_220926_224707/19858408/ccs [↗]	509 [↗]
	CAAHFI011313882.1 [↗]	m64269e_220926_224707/90245730/ccs [↗]	75 [↗]
		m64269e_220926_224707/45351257/ccs [↗]	53 [↗]
	CAAHFI011433261.1 [↗]	m64269e_220926_224707/55182940/ccs [↗]	361 [↗]
		m64269e_220926_224707/138610068/ccs [↗]	325 [↗]
	CAAHFI011569944.1 [↗]	m64269e_220926_224707/70126275/ccs [↗]	609 [↗]
		m64269e_220926_224707/86245980/ccs [↗]	443 [↗]
	CAAHFI011781967.1 [↗]	m64269e_220926_224707/127208210/ccs [↗]	257 [↗]
		m64269e_220926_224707/147130438/ccs [↗]	252 [↗]
	CAAHFI011989906.1 [↗]	m64269e_220926_224707/11797629/ccs [↗]	351 [↗]
		m64269e_220926_224707/156763211/ccs [↗]	322 [↗]
	CAAHFI012022742.1 [↗]	m64269e_220926_224707/56822410/ccs [↗]	569 [↗]
		m64269e_220926_224707/132317800/ccs [↗]	547 [↗]
	CAAHFI012050753.1 [↗]	m64269e_220926_224707/143066356/ccs [↗]	416 [↗]
		m64269e_220926_224707/139527452/ccs [↗]	313 [↗]
	CAAHFI011464222.1 [↗]	m64269e_220926_224707/79627380/ccs [↗]	625 [↗]
		m64269e_220926_224707/35258559/ccs [↗]	505 [↗]
	CAAHFI010081492.1 [↗]	m64269e_220926_224707/47974732/ccs [↗]	1189 [↗]
		m64269e_220926_224707/124387789/ccs [↗]	1173 [↗]
		m64269e_220926_224707/124387789/ccs [↗]	376 [↗]
		m64269e_220926_224707/2752992/ccs [↗]	181 [↗]
		m64269e_220926_224707/2752992/ccs [↗]	203 [↗]

	m64269e_220926_224707/124387789/ccs ⁺	197 ⁺
	m64269e_220926_224707/124387789/ccs ⁻	1173 ⁺
	m64269e_220926_224707/2752992/ccs ⁺	927 ⁺
CAAHFI011326790.1 ⁺	m64269e_220926_224707/156304278/ccs ⁻	539 ⁺
	m64269e_220926_224707/119079977/ccs ⁻	490 ⁺
CAAHFI011718962.1 ⁺	m64269e_220926_224707/100993234/ccs ⁺	2481 ⁺
	m64269e_220926_224707/1705157/ccs ⁻	616 ⁺
CAAHFI010851876.1 ⁺	m64269e_220926_224707/67504620/ccs ⁺	464 ⁺
	m64269e_220926_224707/31393641/ccs ⁻	428 ⁺
	m64269e_220926_224707/67504620/ccs ⁻	698 ⁺
	m64269e_220926_224707/31393641/ccs ⁺	691 ⁺
CAAHFI010347071.1 ⁺	m64269e_220926_224707/48498241/ccs ⁺	643 ⁺
	m64269e_220926_224707/170526546/ccs ⁺	631 ⁺
CAAHFI010347071.1 ⁺	m64269e_220926_224707/76612282/ccs ⁺	620 ⁺
	m64269e_220926_224707/170526546/ccs ⁺	631 ⁺
CAAHFI010521295.1 ⁺	m64269e_220926_224707/163187180/ccs ⁺	890 ⁺
	m64269e_220926_224707/49414239/ccs ⁺	853 ⁺
CAAHFI010536519.1 ⁺	m64269e_220926_224707/165676250/ccs ⁺	731 ⁺
	m64269e_220926_224707/30869908/ccs ⁻	708 ⁺
CAAHFI010572383.1 ⁺	m64269e_220926_224707/72223500/ccs ⁺	384 ⁺
	m64269e_220926_224707/170198887/ccs ⁺	366 ⁺
CAAHFI010909693.1 ⁺	m64269e_220926_224707/155649477/ccs ⁺	820 ⁺
	m64269e_220926_224707/65864500/ccs ⁺	764 ⁺
CAAHFI010959876.1 ⁺	m64269e_220926_224707/96536236/ccs ⁺	282 ⁺
	m64269e_220926_224707/146342703/ccs ⁺	181 ⁺
CAAHFI011183243.1 ⁺	m64269e_220926_224707/52626203/ccs ⁺	226 ⁺
	m64269e_220926_224707/85983447/ccs ⁺	182 ⁺
CAAHFI011276374.1 ⁺	m64269e_220926_224707/176882287/ccs ⁺	537 ⁺
	m64269e_220926_224707/114819535/ccs ⁺	476 ⁺
CAAHFI011317307.1 ⁺	m64269e_220926_224707/55249255/ccs ⁺	431 ⁺
	m64269e_220926_224707/2883957/ccs ⁺	405 ⁺
CAAHFI011342166.1 ⁺	m64269e_220926_224707/50397362/ccs ⁺	661 ⁺
	m64269e_220926_224707/177537392/ccs ⁻	592 ⁺
CAAHFI011552157.1 ⁺	m64269e_220926_224707/11666435/ccs ⁺	643 ⁺
	m64269e_220926_224707/136907271/ccs ⁺	633 ⁺
CAAHFI011557255.1 ⁺	m64269e_220926_224707/92473232/ccs ⁺	1125 ⁺
	m64269e_220926_224707/14024974/ccs ⁻	976 ⁺
CAAHFI011902270.1 ⁺	m64269e_220926_224707/133366989/ccs ⁺	453 ⁺
	m64269e_220926_224707/170853083/ccs ⁺	332 ⁺
CAAHFI011958562.1 ⁺	m64269e_220926_224707/122815136/ccs ⁻	186 ⁺
	m64269e_220926_224707/151192962/ccs ⁺	155 ⁺
CAAHFI011972143.1 ⁺	m64269e_220926_224707/54003200/ccs ⁻	424 ⁺

附录

		m64269e_220926_224707/110166862/ccs [↵]	176 [↵]
	CAAHFI011991608.1 [↵]	m64269e_220926_224707/92405916/ccs [↵]	197 [↵]
		m64269e_220926_224707/163184746/ccs [↵]	174 [↵]
	CAAHFI010295473.1 [↵]	m64269e_220926_224707/92602803/ccs [↵]	1534 [↵]
		m64269e_220926_224707/152897144/ccs [↵]	1528 [↵]
	CAAHFI010546828.1 [↵]	m64269e_220926_224707/28575217/ccs [↵]	797 [↵]
		m64269e_220926_224707/123995553/ccs [↵]	650 [↵]
	CAAHFI010580339.1 [↵]	m64269e_220926_224707/44630174/ccs [↵]	401 [↵]
		m64269e_220926_224707/21233997/ccs [↵]	354 [↵]
	CAAHFI010667019.1 [↵]	m64269e_220926_224707/128647318/ccs [↵]	383 [↵]
		m64269e_220926_224707/170590327/ccs [↵]	244 [↵]
	CAAHFI010874038.1 [↵]	m64269e_220926_224707/25561424/ccs [↵]	650 [↵]
		m64269e_220926_224707/25495777/ccs [↵]	571 [↵]
	CAAHFI011119898.1 [↵]	m64269e_220926_224707/61344480/ccs [↵]	529 [↵]
		m64269e_220926_224707/108266360/ccs [↵]	490 [↵]
	CAAHFI011660197.1 [↵]	m64269e_220926_224707/119932389/ccs [↵]	587 [↵]
		m64269e_220926_224707/118947893/ccs [↵]	578 [↵]
	CAAHFI012023185.1 [↵]	m64269e_220926_224707/15991524/ccs [↵]	891 [↵]
		m64269e_220926_224707/6817316/ccs [↵]	855 [↵]
	CAAHFI012051415.1 [↵]	m64269e_220926_224707/113770944/ccs [↵]	317 [↵]
		m64269e_220926_224707/123668601/ccs [↵]	296 [↵]
	CAAHFI010042953.1 [↵]	m64269e_220926_224707/25822647/ccs [↵]	168 [↵]
		m64269e_220926_224707/84149969/ccs [↵]	107 [↵]
	CAAHFI010272689.1 [↵]	m64269e_220926_224707/152371829/ccs [↵]	983 [↵]
		m64269e_220926_224707/87755449/ccs [↵]	976 [↵]
	CAAHFI010699269.1 [↵]	m64269e_220926_224707/168626511/ccs [↵]	1489 [↵]
		m64269e_220926_224707/41158690/ccs [↵]	1416 [↵]
	CAAHFI011451322.1 [↵]	m64269e_220926_224707/152764890/ccs [↵]	1338 [↵]
		m64269e_220926_224707/117245561/ccs [↵]	1302 [↵]
	CAAHFI011883343.1 [↵]	m64269e_220926_224707/85131491/ccs [↵]	611 [↵]
		m64269e_220926_224707/172753631/ccs [↵]	580 [↵]
	CAAHFI011008416.1 [↵]	m64269e_220926_224707/116327091 [↵]	607 [↵]
		m64269e_220926_224707/138085842 [↵]	595 [↵]
	CAAHFI011655251.1 [↵]	m64269e_220926_224707/64029565/ccs [↵]	723 [↵]
		m64269e_220926_224707/60490157/ccs [↵]	696 [↵]
A5SS [↵] [↵] [↵]	CAAHFI011091074.1 [↵]	m64269e_220926_224707/127140882/ccs [↵]	473 [↵]
		m64269e_220926_224707/57018893/ccs [↵]	422 [↵]
	CAAHFI010422330.1 [↵]	m64269e_220926_224707/134284305/ccs [↵]	308 [↵]
		m64269e_220926_224707/5046787/ccs [↵]	263 [↵]
	CAAHFI010306310.1 [↵]	m64269e_220926_224707/102236931/ccs [↵]	213 [↵]
		m64269e_220926_224707/88605166/ccs [↵]	189 [↵]
	CAAHFI011051963.1 [↵]	m64269e_220926_224707/155910693/ccs [↵]	1898 [↵]

		m64269e_220926_224707/85721392/ccs [↵]	1860 [↵]
	CAAHFI011442426.1 [↵]	m64269e_220926_224707/114558393/ccs [↵] m64269e_220926_224707/85002576/ccs [↵]	657 [↵] 586 [↵]
	CAAHFI010347071.1 [↵]	m64269e_220926_224707/48498241/ccs [↵] m64269e_220926_224707/76612282/ccs [↵]	643 [↵] 620 [↵]
	CAAHFI010354116.1 [↵]	m64269e_220926_224707/29820316/ccs [↵] m64269e_220926_224707/93782689/ccs [↵]	417 [↵] 409 [↵]
	CAAHFI010388402.1 [↵]	m64269e_220926_224707/7013374/ccs [↵] m64269e_220926_224707/117573555/ccs [↵]	265 [↵] 228 [↵]
	CAAHFI011223848.1 [↵]	m64269e_220926_224707/170460429/ccs [↵] m64269e_220926_224707/28117570/ccs [↵]	1010 [↵] 1005 [↵]
A3SS [↵]	CAAHFI011004980.1 [↵]	m64269e_220926_224707/149948060/ccs [↵] m64269e_220926_224707/34604746/ccs [↵]	1468 [↵] 1406 [↵]
	CAAHFI011779575.1 [↵]	m64269e_220926_224707/156239323/ccs [↵] m64269e_220926_224707/92865199/ccs [↵]	814 [↵] 745 [↵]
	CAAHFI011133615.1 [↵]	m64269e_220926_224707/56297830/ccs [↵] m64269e_220926_224707/2032067/ccs [↵]	3515 [↵] 3454 [↵]
	CAAHFI010851876.1 [↵]	m64269e_220926_224707/67504620/ccs [↵] m64269e_220926_224707/31393641/ccs [↵]	400 [↵] 376 [↵]
	CAAHFI010708712.1 [↵]	m64269e_220926_224707/23726551/ccs [↵] m64269e_220926_224707/24316759/ccs [↵]	1421 [↵] 1368 [↵]
	CAAHFI010727729.1 [↵]	m64269e_220926_224707/14092929/ccs [↵] m64269e_220926_224707/149554050/ccs [↵]	594 [↵] 591 [↵]
	CAAHFI010793419.1 [↵]	m64269e_220926_224707/167641783/ccs [↵] m64269e_220926_224707/134743117/ccs [↵]	962 [↵] 950 [↵]
	CAAHFI011631804.1 [↵]	m64269e_220926_224707/46202977/ccs [↵] m64269e_220926_224707/171377062/ccs [↵]	259 [↵] 181 [↵]
	CAAHFI011819729.1 [↵]	m64269e_220926_224707/72745495/ccs [↵] m64269e_220926_224707/111609248/ccs [↵]	544 [↵] 499 [↵]
SE [↵]	CAAHFI010928795.1 [↵]	m64269e_220926_224707/6947845/ccs [↵] m64269e_220926_224707/73729733/ccs [↵]	313 [↵] 157/90 [↵]
atypical-SE [↵]	CAAHFI010816564.1 [↵]	m64269e_220926_224707/124977878/ccs [↵] m64269e_220926_224707/91359879/ccs [↵]	740 [↵] 149/568 [↵]
	CAAHFI010746424.1 [↵]	m64269e_220926_224707/92539151/ccs [↵] m64269e_220926_224707/107611860/ccs [↵]	1591 [↵] 172/52/400/39 5/128 [↵]
	CAAHFI010851876.1 [↵]	m64269e_220926_224707/31393641/ccs [↵] m64269e_220926_224707/67504620/ccs [↵]	901 [↵] 337/310 [↵]
	CAAHFI010002354.1 [↵]	m64269e_220926_224707/103154417/ccs [↵] m64269e_220926_224707/6032004/ccs [↵]	601 [↵] 389/111 [↵]
	CAAHFI010884710.1* [↵]	m64269e_220926_224707/157615496/ccs [↵] m64269e_220926_224707/72157398/ccs [↵]	2546 [↵] 1184/1293 [↵]

	CAAHFI010971903.1* [↰]	m64269e_220926_224707/131924688/ccs [↰] m64269e_220926_224707/67635903/ccs [↰]	943 [↰] 225/563 [↰]
	CAAHFI011136820.1 [↰]	m64269e_220926_224707/3145862/ccs [↰] m64269e_220926_224707/71631551/ccs [↰]	784 [↰] 47/422 [↰]
	CAAHFI011903693.1 [↰]	m64269e_220926_224707/69338870/ccs [↰] m64269e_220926_224707/176293414/ccs [↰]	2498 [↰] 1818/510 [↰]
	CAAHFI010206979.1 [↰]	m64269e_220926_224707/165675916/ccs [↰] m64269e_220926_224707/58392993/ccs [↰]	2066 [↰] 1216/864 [↰]
	CAAHFI010129350.1* [↰]	m64269e_220926_224707/120063365/ccs [↰] m64269e_220926_224707/53609876/ccs [↰]	850 [↰] 89/728 [↰]
RI [↰]	CAAHFI010251568.1 [↰]	m64269e_220926_224707/124453184/ccs [↰] m64269e_220926_224707/52560265/ccs [↰]	301 [↰]
	CAAHFI010251568.1 [↰]	m64269e_220926_224707/124453184/ccs [↰] m64269e_220926_224707/52560265/ccs [↰]	569 [↰]
	CAAHFI010183057.1 [↰]	m64269e_220926_224707/131599195/ccs [↰] m64269e_220926_224707/46204556/ccs [↰]	206 [↰]
	CAAHFI010117114.1 [↰]	m64269e_220926_224707/178652730/ccs [↰] m64269e_220926_224707/25756321/ccs [↰]	301 [↰]
	CAAHFI010252207.1 [↰]	m64269e_220926_224707/9240933/ccs [↰] m64269e_220926_224707/127861445/ccs [↰]	107 [↰]
	CAAHFI010424696.1 [↰]	m64269e_220926_224707/119799873/ccs [↰] m64269e_220926_224707/21301410/ccs [↰]	427 [↰]
	CAAHFI010424696.1 [↰]	m64269e_220926_224707/119799873/ccs [↰] m64269e_220926_224707/21301410/ccs [↰]	64 [↰]
	CAAHFI010458001.1 [↰]	m64269e_220926_224707/170657313/ccs [↰] m64269e_220926_224707/83427732/ccs [↰]	46 [↰]
	CAAHFI011181616.1 [↰]	m64269e_220926_224707/176423776/ccs [↰] m64269e_220926_224707/163447644/ccs [↰]	43 [↰]
	CAAHFI011268395.1 [↰]	m64269e_220926_224707/44763529/ccs [↰] m64269e_220926_224707/162267615/ccs [↰]	38 [↰]
	CAAHFI011313882.1 [↰]	m64269e_220926_224707/45351257/ccs [↰] m64269e_220926_224707/76546619/ccs [↰]	53/75 [↰]
	CAAHFI011342166.1 [↰]	m64269e_220926_224707/177537392/ccs [↰] m64269e_220926_224707/50397362/ccs [↰]	75 [↰]
	CAAHFI011441990.1/CA AHFI010008997.1 [↰]	m64269e_220926_224707/75499924/ccs [↰] m64269e_220926_224707/173146546/ccs [↰]	459 [↰]
	CAAHFI011445760.1/CA AHFI011768082.1 [↰]	m64269e_220926_224707/86181388/ccs [↰] m64269e_220926_224707/17498667/ccs [↰]	323 [↰]
	CAAHFI010834601.1 [↰]	m64269e_220926_224707/9240633/ccs [↰] m64269e_220926_224707/81331924/ccs [↰] m64269e_220926_224707/49939395/ccs [↰]	137/202/256/6 39/186 [↰]
	CAAHFI010840030.1/CA	m64269e_220926_224707/105908210/ccs [↰]	257 [↰]

AHFI010871724.1 [↔]	m64269e_220926_224707/47710286/ccs [↔]	
CAAHFI010910719.1 [↔]	m64269e_220926_224707/33095776/ccs [↔] m64269e_220926_224707/99616075/ccs [↔]	943 [↔]
CAAHFI010938757.1 [↔]	m64269e_220926_224707/69143116/ccs [↔] m64269e_220926_224707/8323678/ccs [↔]	106 [↔]
CAAHFI011136523.1 [↔]	m64269e_220926_224707/57149828/ccs [↔] m64269e_220926_224707/28379574/ccs [↔]	268 [↔]
CAAHFI011136523.1 [↔]	m64269e_220926_224707/57149828/ccs [↔] m64269e_220926_224707/28379574/ccs [↔]	301 [↔]
CAAHFI011136523.1 [↔]	m64269e_220926_224707/57149828/ccs [↔] m64269e_220926_224707/28379574/ccs [↔]	317 [↔]
CAAHFI011212125.1 [↔]	m64269e_220926_224707/95552345/ccs [↔] m64269e_220926_224707/3672563/ccs [↔]	527 [↔]
CAAHFI011212125.1 [↔]	m64269e_220926_224707/95552345/ccs [↔] m64269e_220926_224707/3672563/ccs [↔]	533 [↔]
CAAHFI011248382.1 [↔]	m64269e_220926_224707/36504281/ccs [↔] m64269e_220926_224707/105253532/ccs [↔]	276 [↔]
CAAHFI011381915.1 [↔]	m64269e_220926_224707/6096316/ccs [↔] 64269e_220926_224707/59116066/ccs [↔]	245 [↔]
CAAHFI011894065.1 [↔]	m64269e_220926_224707/107874445/ccs [↔] m64269e_220926_224707/48563120/ccs [↔]	47 [↔]
CAAHFI011894065.1 [↔]	m64269e_220926_224707/107874445/ccs [↔] m64269e_220926_224707/48563120/ccs [↔]	45 [↔]
CAAHFI011956605.1/CA AHFI0111787/87.1 [↔]	m64269e_220926_224707/32704569/ccs [↔] 64269e_220926_224707/129500553/ccs [↔]	443 [↔]
CAAHFI011958562.1 [↔]	m64269e_220926_224707/151192962/ccs [↔] m64269e_220926_224707/122815136/ccs [↔] m64269e_220926_224707/97780953/ccs [↔]	299/186/155 [↔]
CAAHFI010089739.1 [↔]	m64269e_220926_224707/18088509/ccs [↔] m64269e_220926_224707/23986587/ccs [↔]	52 [↔]
CAAHFI010287594.1 [↔]	m64269e_220926_224707/19007685/ccs [↔] m64269e_220926_224707/123472344/ccs [↔]	146 [↔]
CAAHFI010289419.1 [↔]	m64269e_220926_224707/64553644/ccs [↔] m64269e_220926_224707/132317639/ccs [↔]	86 [↔]
CAAHFI010545152.1 [↔]	m64269e_220926_224707/22675755/ccs [↔] m64269e_220926_224707/99485865/ccs [↔]	657 [↔]
CAAHFI010553483.1/CA AHFI011315372.1 [↔]	m64269e_220926_224707/47580562/ccs [↔] m64269e_220926_224707/53871888/ccs [↔]	774 [↔]
CAAHFI010556605.1 [↔]	m64269e_220926_224707/88935172/ccs [↔] m64269e_220926_224707/72288817/ccs [↔]	366 [↔]
CAAHFI010621152.1 [↔]	m64269e_220926_224707/61802380/ccs [↔] m64269e_220926_224707/99879018/ccs [↔]	206/53 [↔]

CAAHFI010642553.1 [↵]	m64269e_220926_224707/29690618/ccs [↵] m64269e_220926_224707/55246883/ccs [↵]	423 [↵]
CAAHFI010739965.1 [↵]	m64269e_220926_224707/2361516/ccs [↵] 64269e_220926_224707/97518047/ccs [↵]	384 [↵]
CAAHFI010777982.1 [↵]	m64269e_220926_224707/74711382/ccs [↵] m64269e_220926_224707/34015495/ccs [↵]	492 [↵]
CAAHFI010798983.1 [↵]	m64269e_220926_224707/140641993/ccs [↵] m64269e_220926_224707/77923112/ccs [↵]	344/60 [↵]
CAAHFI010808994.1 [↵]	m64269e_220926_224707/94109738/ccs [↵] m64269e_220926_224707/142936872/ccs [↵]	86 [↵]
CAAHFI010867420.1 [↵]	m64269e_220926_224707/112984650/ccs [↵] m64269e_220926_224707/90767712/ccs [↵]	44 [↵]
CAAHFI010876927.1 [↵]	m64269e_220926_224707/56494121/ccs [↵] m64269e_220926_224707/24447304/ccs [↵]	161 [↵]
CAAHFI010997633.1 [↵]	m64269e_220926_224707/140904706/ccs [↵] m64269e_220926_224707/53674879/ccs [↵]	46 [↵]
CAAHFI011003871.1 [↵]	m64269e_220926_224707/86902153/ccs [↵] m64269e_220926_224707/37486700/ccs [↵]	46 [↵]
CAAHFI011055188.1 [↵]	m64269e_220926_224707/166724564/ccs [↵] m64269e_220926_224707/131205235/ccs [↵]	187 [↵]
CAAHFI011058993.1 [↵]	m64269e_220926_224707/109512973/ccs [↵] m64269e_220926_224707/108857003/ccs [↵]	371 [↵]
CAAHFI011205260.1 [↵]	m64269e_220926_224707/72352247/ccs [↵] m64269e_220926_224707/176424398/ccs [↵]	79 [↵]
CAAHFI011336853.1 [↵]	m64269e_220926_224707/131402195/ccs [↵] m64269e_220926_224707/82182908/ccs [↵]	802 [↵]
CAAHFI011375181.1 [↵]	m64269e_220926_224707/104333554/ccs [↵] m64269e_220926_224707/125239935/ccs [↵]	698 [↵]
CAAHFI011461811.1 [↵]	m64269e_220926_224707/87820527/ccs [↵] m64269e_220926_224707/157878403/ccs [↵]	109 [↵]
CAAHFI011519620.1 [↵]	m64269e_220926_224707/68225104/ccs [↵] m64269e_220926_224707/20056487/ccs [↵]	119 [↵]
CAAHFI011533207.1 [↵]	m64269e_220926_224707/7537261/ccs [↵] m64269e_220926_224707/45353457/ccs [↵]	488 [↵]
CAAHFI011582030.1/CA AHFI011356103.1 [↵]	m64269e_220926_224707/77007101/ccs [↵] m64269e_220926_224707/166724855/ccs [↵]	88 [↵]
CAAHFI011588633.1 [↵]	m64269e_220926_224707/138412971/ccs [↵] m64269e_220926_224707/17041535/ccs [↵]	99 [↵]
CAAHFI011591745.1 [↵]	m64269e_220926_224707/64028913/ccs [↵] m64269e_220926_224707/177997138/ccs [↵]	367 [↵]
CAAHFI011648406.1 [↵]	m64269e_220926_224707/78119856/ccs [↵] m64269e_220926_224707/60752672/ccs [↵]	146 [↵]

CAAHFI011660600.1 [↔]	m64269e_220926_224707/37358141/ccs [↔] m64269e_220926_224707/154928478/ccs [↔]	320 [↔]
CAAHFI011662811.1 [↔]	m64269e_220926_224707/34408416/ccs [↔] m64269e_220926_224707/154141428/ccs [↔]	42 [↔]
CAAHFI011864412.1/CA AHFI011960032.1 [↔]	m64269e_220926_224707/98371702/ccs [↔] m64269e_220926_224707/1705605/ccs [↔]	272 [↔]
CAAHFI011974773.1 [↔]	m64269e_220926_224707/122686065/ccs [↔] m64269e_220926_224707/144376982/ccs [↔]	53/53 [↔]
CAAHFI012040318.1 [↔]	m64269e_220926_224707/163382336/ccs [↔] m64269e_220926_224707/22022386/ccs [↔]	92 [↔]
CAAHFI010015404.1 [↔]	m64269e_220926_224707/16975318/ccs [↔] m64269e_220926_224707/34866553/ccs [↔]	42 [↔]
CAAHFI010024059.1 [↔]	m64269e_220926_224707/94308925/ccs [↔] m64269e_220926_224707/7800110/ccs [↔]	256 [↔]
CAAHFI010042953.1 [↔]	m64269e_220926_224707/84149969/ccs [↔] m64269e_220926_224707/88015560/ccs [↔]	40 [↔]
CAAHFI010056306.1 [↔]	m64269e_220926_224707/94307933/ccs [↔] m64269e_220926_224707/151587211/ccs [↔]	179 [↔]
CAAHFI010096060.1 [↔]	m64269e_220926_224707/175049399/ccs [↔] m64269e_220926_224707/92997727/ccs [↔]	49 [↔]
CAAHFI010103312.1 [↔]	m64269e_220926_224707/137953866/ccs [↔] m64269e_220926_224707/57344423/ccs [↔]	1080223/342 [↔]
CAAHFI010165720.1 [↔]	m64269e_220926_224707/8391098/ccs [↔] m64269e_220926_224707/138479954/ccs [↔]	50 [↔]
CAAHFI010175873.1 [↔]	m64269e_220926_224707/154599433/ccs [↔] m64269e_220926_224707/7078780/ccs [↔]	538/646 [↔]
CAAHFI010212341.1 [↔]	m64269e_220926_224707/140248245/ccs [↔] m64269e_220926_224707/113903349/ccs [↔]	273 [↔]
CAAHFI010221356.1/CA AHFI010806849.1 [↔]	m64269e_220926_224707/91817066/ccs [↔] m64269e_220926_224707/87295368/ccs [↔]	475 [↔]
CAAHFI010280974.1/CA AHFI010243923.1 [↔]	m64269e_220926_224707/79300653/ccs [↔] m64269e_220926_224707/105581071/ccs [↔]	138 [↔]
CAAHFI010305033.1 [↔]	m64269e_220926_224707/21102887/ccs [↔] m64269e_220926_224707/116393009/ccs [↔]	253/625 [↔]
CAAHFI010313717.1 [↔]	m64269e_220926_224707/151585502/ccs [↔] m64269e_220926_224707/90966277/ccs [↔]	120 [↔]
CAAHFI010365237.1 [↔]	m64269e_220926_224707/34537980/ccs [↔] m64269e_220926_224707/99813463/ccs [↔]	307 [↔]
CAAHFI010436661.1/CA AHFI010076854.1 [↔]	m64269e_220926_224707/16451610/ccs [↔] m64269e_220926_224707/4129539/ccs [↔]	267 [↔]
CAAHFI010477291.1 [↔]	m64269e_220926_224707/161744514/ccs [↔] m64269e_220926_224707/124650129/ccs [↔]	48 [↔]

CAAHFI010490763.1 [±]	m64269e_220926_224707/81068160/ccs [±] m64269e_220926_224707/126484816/ccs [±]	85 [±]
CAAHFI010561259.1 [±]	m64269e_220926_224707/119802601/ccs [±] m64269e_220926_224707/14876856/ccs [±]	222 [±]
CAAHFI010563531.1 [±]	m64269e_220926_224707/98371573/ccs [±] m64269e_220926_224707/78907504/ccs [±]	325 [±]
CAAHFI010580500.1 [±]	m64269e_220926_224707/14551322/ccs [±] m64269e_220926_224707/158991364/ccs [±]	207/275 [±]
CAAHFI010671002.1/CA AHFI011978759.1 [±]	m64269e_220926_224707/38142104/ccs [±] m64269e_220926_224707/35782971/ccs [±]	127 [±]
CAAHFI010740163.1 [±]	m64269e_220926_224707/59639769/ccs [±] m64269e_220926_224707/87950588/ccs [±]	69 [±]
CAAHFI010790380.1 [±]	m64269e_220926_224707/36373643/ccs [±] m64269e_220926_224707/10683252/ccs [±]	43 [±]
CAAHFI010808415.1 [±]	m64269e_220926_224707/6947578/ccs [±] m64269e_220926_224707/23399160/ccs [±]	220 [±]
CAAHFI010867865.1/CA AHFI011387466.1 [±]	m64269e_220926_224707/29231513/ccs [±] m64269e_220926_224707/55118649/ccs [±]	104 [±]
CAAHFI010979715.1 [±]	m64269e_220926_224707/72615592/ccs [±] m64269e_220926_224707/120391163/ccs [±]	260/435 [±]
CAAHFI011184642.1/CA AHFI011223267.1 [±]	m64269e_220926_224707/106235146/ccs [±] m64269e_220926_224707/90704295/ccs [±]	49 [±]
CAAHFI011251957.1 [±]	m64269e_220926_224707/16976172/ccs [±] m64269e_220926_224707/80022139/ccs [±]	242/710 [±]
CAAHFI011252142.1 [±]	m64269e_220926_224707/72091012/ccs [±] m64269e_220926_224707/61669907/ccs [±]	45 [±]
CAAHFI011339149.1 [±]	m64269e_220926_224707/51317014/ccs [±] m64269e_220926_224707/167249750/ccs [±]	355 [±]
CAAHFI011494914.1 [±]	m64269e_220926_224707/1902432/ccs [±] m64269e_220926_224707/147652615/ccs [±]	347 [±]
CAAHFI011514322.1 [±]	m64269e_220926_224707/53740971/ccs [±] m64269e_220926_224707/167708422/ccs [±]	489 [±]
CAAHFI011518668.1/CA AHFI011367957.1 [±]	m64269e_220926_224707/12715887/ccs [±] m64269e_220926_224707/141363650/ccs [±]	736 [±]
CAAHFI011548752.1/CA AHFI011985068.1 [±]	m64269e_220926_224707/115279155/ccs [±] m64269e_220926_224707/93849977/ccs [±]	301 [±]
CAAHFI011602900.1 [±]	m64269e_220926_224707/165347735/ccs [±] m64269e_220926_224707/63309526/ccs [±]	126 [±]
CAAHFI011658843.1 [±]	m64269e_220926_224707/106824634/ccs [±] m64269e_220926_224707/37291225/ccs [±]	41 [±]
CAAHFI011667060.1 [±]	m64269e_220926_224707/50005958/ccs [±] m64269e_220926_224707/84084082/ccs [±]	609 [±]

CAAHFI011682341.1/CAAHFI010052243.1 [↔]	m64269e_220926_224707/153946897/ccs [↔] m64269e_220926_224707/105253299/ccs [↔]	310 [↔]
CAAHFI011821328.1/CAAHFI010484113.1 [↔]	m64269e_220926_224707/23724442/ccs [↔] m64269e_220926_224707/19989337/ccs [↔]	666 [↔]
CAAHFI011857585.1 [↔]	m64269e_220926_224707/133498173/ccs [↔] m64269e_220926_224707/127535178/ccs [↔]	477 [↔]
CAAHFI011912623.1 [↔]	m64269e_220926_224707/29492184/ccs [↔] m64269e_220926_224707/75302521/ccs [↔]	237 [↔]
CAAHFI011917852.1/CAAHFI011586153.1 [↔]	m64269e_220926_224707/70912396/ccs [↔] m64269e_220926_224707/45812438/ccs [↔]	516 [↔]
CAAHFI011992067.1 [↔]	m64269e_220926_224707/117836412/ccs [↔] m64269e_220926_224707/114885302/ccs [↔]	46/48 [↔]
CAAHFI012034269.1/CAAHFI010098996.1 [↔]	m64269e_220926_224707/109774664/ccs [↔] m64269e_220926_224707/18614675/ccs [↔]	252 [↔]
CAAHFI010087310.1 [↔]	m64269e_220926_224707/33292576/ccs [↔] m64269e_220926_224707/85984803/ccs [↔]	44/51/51 [↔]
CAAHFI010798983.1 [↔]	m64269e_220926_224707/140641993/ccs [↔] m64269e_220926_224707/77923112/ccs [↔]	344/60 [↔]
CAAHFI010218620.1 [↔]	m64269e_220926_224707/27133920/ccs [↔] m64269e_220926_224707/123929473/ccs [↔]	102/237/75 [↔]
CAAHFI010746424.1 [↔]	m64269e_220926_224707/107611860/ccs [↔] m64269e_220926_224707/124520885/ccs [↔]	52 [↔]
CAAHFI011645292.1 [↔]	m64269e_220926_224707/42140649/ccs [↔] m64269e_220926_224707/158467092/ccs [↔]	174/318/598/95 [↔]
CAAHFI011439774.1 [↔]	m64269e_220926_224707/99812553/ccs [↔] m64269e_220926_224707/102498909/ccs [↔]	113/291/413/662 [↔]
CAAHFI010081492.1 [↔]	m64269e_220926_224707/2752992/ccs [↔] m64269e_220926_224707/47974732/ccs [↔]	927/181 [↔]
CAAHFI011947138.1 [↔]	m64269e_220926_224707/49939399/ccs [↔] m64269e_220926_224707/97321100/ccs [↔]	240/298 [↔]
CAAHFI010235868.1 [↔]	m64269e_220926_224707/158271152/ccs [↔] m64269e_220926_224707/123471428/ccs [↔]	493 [↔]
CAAHFI010248748.1 [↔]	m64269e_220926_224707/138807764/ccs [↔] m64269e_220926_224707/39977266/ccs [↔]	191 [↔]
CAAHFI010546773.1/CAAHFI011462264.1 [↔]	m64269e_220926_224707/21433056/ccs [↔] m64269e_220926_224707/141101184/ccs [↔]	47 [↔]
CAAHFI010552633.1 [↔]	m64269e_220926_224707/64883021/ccs [↔] m64269e_220926_224707/175899233/ccs [↔]	220 [↔]
CAAHFI010600833.1 [↔]	m64269e_220926_224707/89459215/ccs [↔] m64269e_220926_224707/109971974/ccs [↔]	977 [↔]
CAAHFI010617149.1 [↔]	m64269e_220926_224707/131138402/ccs [↔] m64269e_220926_224707/175901328/ccs [↔]	59/613 [↔]

CAAHFI010634071.1 [±]	m64269e_220926_224707/9635362/ccs [±] m64269e_220926_224707/12453269/ccs [±]	440/64 [±]
CAAHFI010694525.1/CA AHFI010443843.1 [±]	m64269e_220926_224707/31260827/ccs [±] m64269e_220926_224707/13434903/ccs [±]	82 [±]
CAAHFI010804878.1 [±]	m64269e_220926_224707/132187969/ccs [±] m64269e_220926_224707/139855726/ccs [±]	143 [±]
CAAHFI010959876.1 [±]	m64269e_220926_224707/96536236/ccs [±] m64269e_220926_224707/51642752/ccs [±]	282/290/303 [±]
CAAHFI010967293.1 [±]	m64269e_220926_224707/32704569/ccs [±] m64269e_220926_224707/129500553/ccs [±]	443 [±]
CAAHFI011055609.1 [±]	m64269e_220926_224707/89721314/ccs [±] m64269e_220926_224707/149292565/ccs [±]	132/66/909 [±]
CAAHFI011367070.1 [±]	m64269e_220926_224707/39519676/ccs [±] m64269e_220926_224707/51905047/ccs [±]	512/626 [±]
CAAHFI011408146.1 [±]	m64269e_220926_224707/92078858/ccs [±] m64269e_220926_224707/133890605/ccs [±]	415 [±]
CAAHFI011479801.1 [±]	m64269e_220926_224707/78578814/ccs [±] m64269e_220926_224707/153946956/ccs [±]	55 [±]
CAAHFI011510581.1 [±]	m64269e_220926_224707/57018669/ccs [±] m64269e_220926_224707/95093244/ccs [±]	418 [±]
CAAHFI011522833.1 [±]	m64269e_220926_224707/177998240/ccs [±] m64269e_220926_224707/159581740/ccs [±]	660 [±]
CAAHFI011635244.1 [±]	m64269e_220926_224707/179899056/ccs [±] m64269e_220926_224707/129959467/ccs [±]	193 [±]
CAAHFI011643931.1 [±]	m64269e_220926_224707/127076143/ccs [±] m64269e_220926_224707/79563418/ccs [±]	187/632/669 [±]
CAAHFI011734280.1 [±]	m64269e_220926_224707/122619124/ccs [±] m64269e_220926_224707/168429499/ccs [±]	251/378/442 [±]
CAAHFI011735749.1 [±]	m64269e_220926_224707/129042828/ccs [±] m64269e_220926_224707/18088090/ccs [±]	80 [±]
CAAHFI011806276.1 [±]	m64269e_220926_224707/158335562/ccs [±] m64269e_220926_224707/100665998/ccs [±]	1326/98 [±]
CAAHFI011923938.1/CA AHFI011097670.1 [±]	m64269e_220926_224707/64226920/ccs [±] m64269e_220926_224707/165743134/ccs [±]	514 [±]
CAAHFI011948404.1 [±]	m64269e_220926_224707/133562940/ccs [±] m64269e_220926_224707/96731853/ccs [±]	115/143 [±]
CAAHFI010072932.1 [±]	m64269e_220926_224707/119081761/ccs [±] m64269e_220926_224707/67701420/ccs [±]	367 [±]
CAAHFI010244548.1 [±]	m64269e_220926_224707/116591088/ccs [±] m64269e_220926_224707/142477275/ccs [±]	107 [±]
CAAHFI010645279.1 [±]	m64269e_220926_224707/157158187/ccs [±] m64269e_220926_224707/45417719/ccs [±]	197 [±]

CAAHFI010761963.1 [↵]	m64269e_220926_224707/167315204/ccs [↵] m64269e_220926_224707/7145781/ccs [↵]	46/47/47/48/53 [↵]
CAAHFI011156458.1 [↵]	m64269e_220926_224707/163055438/ccs [↵] m64269e_220926_224707/67436738/ccs [↵]	213/440/484/86 [↵]
CAAHFI011430762.1 [↵]	m64269e_220926_224707/179243393/ccs [↵] m64269e_220926_224707/164955899/ccs [↵]	228 [↵]
CAAHFI011584046.1 [↵]	m64269e_220926_224707/30737336/ccs [↵] m64269e_220926_224707/22217391/ccs [↵]	134/204 [↵]
CAAHFI011703139.1 [↵]	m64269e_220926_224707/133958119/ccs [↵] m64269e_220926_224707/74843461/ccs [↵]	111/197/803 [↵]
CAAHFI010054903.1 [↵]	m64269e_220926_224707/122489486/ccs [↵] m64269e_220926_224707/88605595/ccs [↵]	1196/185/47/75/82 [↵]
CAAHFI010078595.1 [↵]	m64269e_220926_224707/60688656/ccs [↵] m64269e_220926_224707/170852933/ccs [↵]	234 [↵]
CAAHFI010272689.1 [↵]	m64269e_220926_224707/87755449/ccs [↵] m64269e_220926_224707/152371829/ccs [↵]	121 [↵]
CAAHFI010359876.1 [↵]	m64269e_220926_224707/174721391/ccs [↵] m64269e_220926_224707/31524226/ccs [↵]	120/205/257/495 [↵]
CAAHFI010928795.1 [↵]	m64269e_220926_224707/73729733/ccs [↵] m64269e_220926_224707/6947845/ccs [↵] m64269e_220926_224707/70583704/ccs [↵]	157/311/313/90 [↵]
CAAHFI011004469.1 [↵]	m64269e_220926_224707/116263023/ccs [↵] m64269e_220926_224707/79431872/ccs [↵]	254/313 [↵]
CAAHFI011124639.1 [↵]	m64269e_220926_224707/22544738/ccs [↵] m64269e_220926_224707/22544738/ccs [↵]	86 [↵]
CAAHFI011451106.1 [↵]	m64269e_220926_224707/148635675/ccs [↵] m64269e_220926_224707/9243427/ccs [↵]	263/303/825 [↵]
CAAHFI011495331.1 [↵]	m64269e_220926_224707/90899949/ccs [↵] m64269e_220926_224707/68684069/ccs [↵]	212/308 [↵]
CAAHFI011721723.1 [↵]	m64269e_220926_224707/7932098/ccs [↵] m64269e_220926_224707/78250482/ccs [↵]	109/341 [↵]
CAAHFI011918215.1/CAAHFI011876870.1 [↵]	m64269e_220926_224707/92210343/ccs [↵] m64269e_220926_224707/40764103/ccs [↵]	165 [↵]
CAAHFI011961491.1 [↵]	m64269e_220926_224707/10422564/ccs [↵] m64269e_220926_224707/51315372/ccs [↵]	185/502/580 [↵]
CAAHFI010086765.1 [↵]	m64269e_220926_224707/48038957/ccs [↵] m64269e_220926_224707/138347703/ccs [↵]	240 [↵]
CAAHFI010463498.1 [↵]	m64269e_220926_224707/107873713/ccs [↵] m64269e_220926_224707/18940608/ccs [↵]	1013/380/59/67 [↵]
CAAHFI011860170.1 [↵]	m64269e_220926_224707/57999749/ccs [↵] m64269e_220926_224707/93651673/ccs [↵]	546 [↵]