



# 微生物小开放阅读框：小蛋白及翻译调控研究进展

杨婕<sup>1</sup>, 张安琪<sup>1</sup>, 冯家勋<sup>2</sup>, 赵帅<sup>2\*</sup>, 赵心清<sup>1\*</sup>

1 上海交通大学生命科学技术学院 微生物代谢国家重点实验室, 上海 200240

2 广西大学生命科学与技术学院 广西微生物与酶工程技术研究中心 亚热带农业生物资源保护与利用国家重点实验室, 广西 南宁 530004

杨婕, 张安琪, 冯家勋, 赵帅, 赵心清. 微生物小开放阅读框: 小蛋白及翻译调控研究进展[J]. 微生物学报, 2023, 63(1): 1-14.

YANG Jie, ZHANG Anqi, FENG Jiaxun, ZHAO Shuai, ZHAO Xinqing. Progress in small open reading frames (sORFs) in microorganisms: small proteins and translational regulation[J]. *Acta Microbiologica Sinica*, 2023, 63(1): 1-14.

**摘要:** 小开放阅读框(small open reading frame, sORF)广泛存在于不同生物基因组中, 由于其序列短, 以及编码的产物小蛋白(small protein, 或称微蛋白; microprotein 或迷你蛋白 miniprotein)检测困难等原因, 小开放阅读框长期未得到充分注释和研究。近年来, 随着高通量测序、翻译组和质谱分析等技术的不断发展, 在不同生物中发现大量新的小开放阅读框, 其编码的小蛋白及介导的翻译调控已应用于药物开发及植物抗病机理等研究。但是, 目前对微生物的小开放阅读框相关研究和应用还相对有限。本文综述了小开放阅读框编码产物小蛋白的发现和鉴定, 以及上游开放阅读框(upstream open reading frame, uORF)对 mRNA 翻译调控等最新研究进展, 重点介绍了微生物基因组中小开放阅读框的鉴定和功能研究进展, 为深入认识微生物中小开放阅读框的功能和作用机制, 以及植物和动物等高等其他生物的小蛋白和翻译调控相关研究提供参考。

**关键词:** 小开放阅读框(sORF); 小蛋白; 上游开放阅读框(uORF); 翻译调控; 微生物基因组

资助项目: 国家自然科学基金地区联合基金重点项目(U21A20178); 国家重点研发计划(2021YFC2101302)

This work was supported by the Joint Fund for Regional Innovation and Development of National Natural Science Foundation of China (U21A20178) and the National Key Research and Development Program of China (2021YFC2101302).

\*Corresponding authors. E-mail: ZHAO Shuai, shuaizhao0227@163.com; ZHAO Xinqing, xqzhao@sjtu.edu.cn

Received: 2022-04-10; Accepted: 2022-06-18; Published online: 2022-07-04

# Progress in small open reading frames (sORFs) in microorganisms: small proteins and translational regulation

YANG Jie<sup>1</sup>, ZHANG Anqi<sup>1</sup>, FENG Jiaxun<sup>2</sup>, ZHAO Shuai<sup>2\*</sup>, ZHAO Xinqing<sup>1\*</sup>

<sup>1</sup> State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup> State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi Research Center for Microbial and Enzyme Engineering Technology, College of Life Science and Technology, Guangxi University, Nanning 530004, Guangxi, China

**Abstract:** Small open reading frames (sORFs) are ubiquitous in the genomes of living organisms, which are generally not fully annotated and studied due to their short length and difficulty in detection of their encoded small proteins (also termed microproteins or miniproteins). With the advancement of high-throughput sequencing, translomics and mass spectrometry, increasing sORFs have been identified in different living organisms. At the same time, specific bioinformatic tools have also contribute to the mining of sORFs. Small proteins encoded by sORFs as well as translational regulation mediated by sORFs have been applied in medicinal development and exploration of mechanisms underlying plant disease resistance. However, the studies and applications of sORFs in microorganisms remain limited. This review summarizes the latest progress in the identification of small proteins encoded by sORFs, the annotation of upstream open reading frames (uORFs), as well as uORF-mediated regulation on mRNA translation. In particular, we reviewed the identification and functional analysis of sORFs in microbial genomes, especially budding yeast and filamentous fungi. This review provides a basis for understanding the function and mechanism of sORFs, as well as studies on microproteins and translation regulation in other organisms including plants and animals.

**Keywords:** small open reading frames (sORFs); small proteins (microproteins); upstream open reading frames (uORFs); translational regulation; Microbial genomes

小开放阅读框(small open reading frame, sORF)普遍存在于不同生物的基因组中<sup>[1]</sup>。其中一部分 sORF 可被转录和翻译,其编码产物称为小蛋白(small protein),又称为微蛋白(microprotein),或迷你蛋白(miniprotein)。但是,由于 sORF 序列短,可能不具有充分的同源序列信息,或者没有已知的结构域,因此无法进行相应的同源搜索。此外,很多 sORF 的功能未知,而且有一些功能小蛋白仅在特定条件下表达,很难在大量可能的启动子和终止子中定

义具有编码功能的 sORF。因此,长期以来很多 sORF 未被充分注释<sup>[2]</sup>。此外,相关检测技术的限制也影响了对 sORF 的充分研究,传统的基因预测为了降低假阳性,习惯性地将对编码 100 个氨基酸以下的 sORF 注释为非编码 RNA (non-coding RNA, ncRNA)<sup>[3]</sup>。近年来,随着高通量测序、翻译组和质谱分析技术的不断发展,针对小蛋白的发现开发了特定的分析方法<sup>[4-6]</sup>。针对 sORF 的生物信息分析技术也不断完善<sup>[7]</sup>,因此,近年来大量 sORF 的鉴定和功能被不断发现和报道。

越来越多的研究表明, sORF 编码的小蛋白参与了多种重要生命过程, 例如, DNA 修复、信号传导、环境胁迫耐受性、疾病和发育等, 已成为近期研究的热点, 不仅在微生物, 在斑马鱼、果蝇和哺乳动物也不断有新的发现, 在多肽类新药开发方面也取得了进展<sup>[5-6]</sup>。此外, 对 sORF 介导的翻译调控作用研究结果显示, 这些被翻译的顺式调控元件不仅具有理论研究价值, 也有潜在的应用前景。但是, 目前大量关于小蛋白和 sORF 介导的翻译调控的相关研究主要集中在植物和小鼠等高等生物中, 而在微生物中的相关研究还相对比较有限。

微生物及其产品广泛应用在食品、饲料、医药、农业和生物能源等不同领域。微生物 sORF 的发现和作用机制研究不仅有助于深入理解微生物对环境的响应和代谢重塑<sup>[8]</sup>, 也对精准设计和改造微生物菌株, 提高生产效率, 更好服务产业具有重要意义<sup>[9]</sup>。本文对目前微生物相关的 sORF 研究进行了综述, 并对未来 sORF 的研究趋势进行了展望。

## 1 sORF 和小蛋白简介

由于大量的 sORF 功能未知, sORF 被称为基因组中的“暗物质”<sup>[10]</sup>。sORF 在原核生物中指编码小于 50 个氨基酸的开放阅读框, 在真核生物中, 除植物中指编码小于 150 个氨基酸外, 一般指代编码小于 100 个氨基酸的开放阅读框。sORF 与基因组中非传统 ORF (alternative ORF, alt-ORF) 最主要的区别是对编码产物长度的限定不同。alt-ORF 指在转录本(编码或非编码)上以 ATG 起始的尚未被注释的任何编码大于 30 个氨基酸的 ORF<sup>[3]</sup>, 因此, sORF 与 alt-ORF 存在交集。尽管目前使用编码氨基酸的长度作为定义小蛋白的标准具有一定的主观性, 但值得注意的是, sORF 编码的小蛋白与常见的多

肽最主要区别为小蛋白是由独立的 ORF 所编码, 而不是通过大分子蛋白质加工或降解而成。

根据 sORF 在转录本上的位置, 分为 5'非翻译区(untranslated region, UTR)的上游 ORF (upstream ORF, uORF)、3'UTR 的下游 ORF (downstream ORF, dORF)、编码区与非编码区重叠区域的交错 sORF (interlaced-sORF)和长链非编码 RNA 上的 sORF (lncRNA-sORF)<sup>[5]</sup>(图 1)。这些 sORF 都可能翻译出具有独立功能的小蛋白。目前, 部分研究学者将能够编码翻译出小蛋白的 uORF 和 dORF 归类为 alt-ORF<sup>[3]</sup>。有的 sORF 的翻译产物具有独立的生物学功能, 这类小蛋白常发现于 lncRNA-sORF<sup>[11]</sup>; 有的 sORF 则是调控其下游或上游主 ORF 的顺式调控元件(图 1)。

## 2 sORF 和小蛋白的挖掘和鉴定

目前, 挖掘潜在 sORF 的高通量筛选方法主要包括生物信息学分析、翻译组测序策略和质谱蛋白组分析。高通量测序技术的应用对鉴定 sORF 非常重要(图 2)。虽然通过生物信息学分析可以预测出大量的 sORF, 但是真正能表达出小蛋白只是其中的小部分, 因此, 需要对翻译的 sORF 和有生物学功能的小蛋白进行高效筛选和鉴定。这里我们总结了近年来生物信息学分析技术、翻译组测序策略和质谱蛋白组分析技术在挖掘 sORF 及检测其翻译的应用。

### 2.1 生物信息学分析

生物信息学分析是预测 sORF 的第一步, 目的是将具有功能表达的元素从随机发生的噪声中区分出来, 但是, 目前实现高灵敏度和特异性的预测仍然具有挑战性。引起这种挑战性的主要原因之一是肽长度的阈值设置<sup>[12]</sup>。过去在蛋白质编码序列预测过程中, 一般设置的长度阈值为 100 个氨基酸, 忽略了更短 ORF 编码

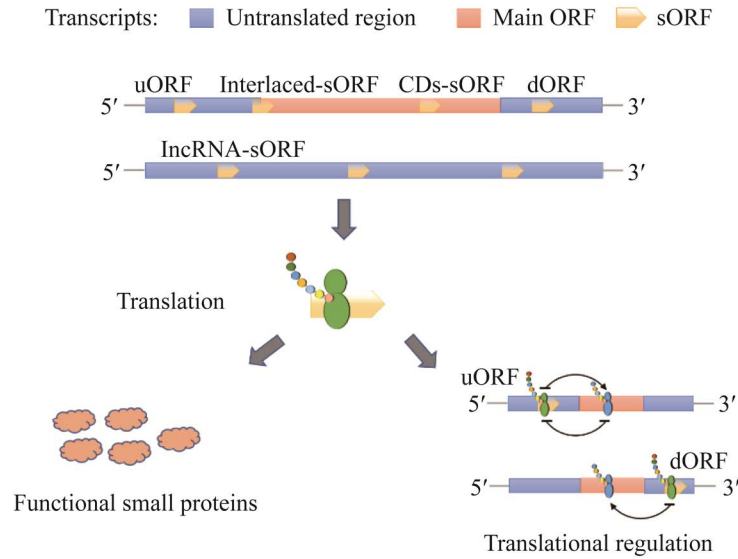


图 1 sORF 在转录本上的定位分类及其翻译后的功能分类

Figure 1 Classification of sORFs based on their localization on transcripts and classification by their post-translational functions. uORF: Upstream sORF; dORF: Downstream sORF; CD-sORF: Coding sequence-sORF.

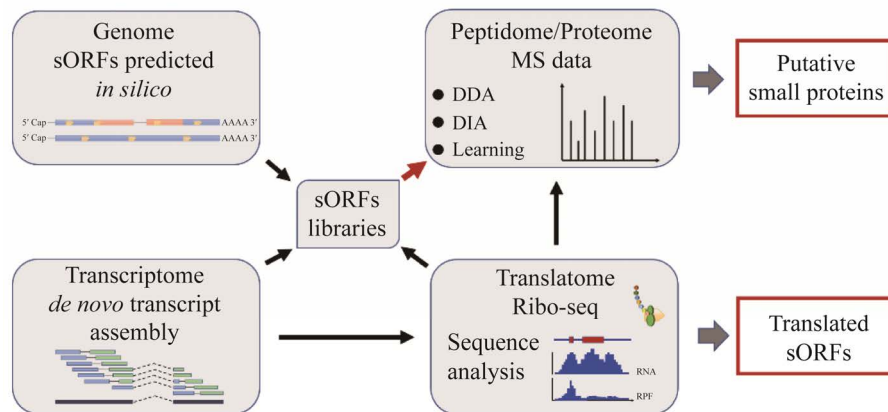


图 2 多组学技术进行 sORF 高通量筛选的方法

Figure 2 High-throughput screening methods for sORFs via multiple omics techniques.

基因的表达和功能作用；另一个原因是替代起始密码子的使用，使新 sORF 的发现更加复杂，例如，一些蛋白质以非 AUG 起始密码子起始翻译<sup>[13-14]</sup>，而这类蛋白质在基因注释中通常被忽略。

大多数情况下，生物信息学分析挖掘 sORFs 依赖于通用原则，包括 (1) 通过不同物种

之间的序列比对，筛选保守的 sORF；(2) sORFs 序列内编码区域的密码子使用等特征分析；(3) 评估与之前鉴定的蛋白质或某些功能域的序列相似性<sup>[12]</sup>。其中，序列保守性和相似性是常用的度量指标。但是，识别潜在的 sORF 受到多物种比对的质量和已知基因相似度的限制，而且短序列在定量保守方法中往往得分较低。

因此, 近期研究人员改进开发的 MicroPeptide tool (MiPepid), 可在全基因组范围内挖掘 sORF, 该工具利用小蛋白作为训练集的逻辑回归模型来预测 sORF, 提高了注释能力<sup>[15]</sup>。此外, smORFunction 通过重新注释的微阵列探针估计 sORF 的 RNA 表达, 可根据已知功能注释的相关基因预测 sORF 的功能<sup>[16]</sup>。尽管以上研究促进了 sORF 的注释和预测几率, 但是实验数据的集成仍然是鉴定真实 sORF 的关键。对于非保守型或新型的功能性 sORF, 实验方法的可靠性也是做出决定性结论所必不可少的。目前, 研究人员也开发了一些小蛋白数据库, 已收录了多种模式物种基于翻译组、蛋白组发现的小蛋白, 这些数据库为小蛋白的生物信息学分析和功能研究提供了基础<sup>[7]</sup>。

## 2.2 翻译组测序分析

传统的 mRNA 测序不能准确地显示基因的翻译情况, mRNA 转录不一定代表蛋白质的实际表达<sup>[17]</sup>。多聚体分析(polysome profiling)和翻译核糖体亲和纯化(translating ribosome affinity purification, TRAP)技术已经广泛应用于翻译调控的研究。多聚体分析是一种通过梯度离心分离多聚体相关 mRNA 的技术, 通过 Northern blotting、RT-qPCR 或深度测序来识别和量化积极翻译的转录本。其优势在于使用 RT-qPCR 靶向的分析具有成本低效率高的特点, 并且结合免疫印迹或蛋白质组学有助于监测与核糖体或翻译起始复合物相关的蛋白质<sup>[18]</sup>。TRAP 技术则是通过在生物体内表达表位标记的核糖体蛋白, 使核糖体结合的 mRNA 复合物能够免疫纯化。其优势在于能够减少试剂对 mRNA-核糖体蛋白的污染, 适用于分析植物多细胞器官或组织中特定细胞类型的翻译<sup>[19]</sup>。此外, 近来亦有研究开发利用全长翻译 mRNA 测序或称为核糖体新生链复合体结合 RNA 测序(ribosome

nascent-chain complex-bound RNA sequencing, RNC-seq), 以更好地揭示蛋白质翻译起始的全局情况<sup>[20]</sup>, 结合 mRNA-seq 和 RNC-seq 分析可以同时评估 mRNA 转录和蛋白质合成, 识别转录和转录后的基因调控。但是, 与多聚体分析和 TRAP 技术类似, 这 3 种方法都不能为转录本上的核糖体提供位置信息, 因此, 无法区分在 mRNA 上停滞不前的翻译和非翻译核糖体, 从而不能确定多顺反子的翻译, 无法识别单个转录本中和不同区域(如 sORF)的 RNA。

2009 年首次报道的核糖体印迹分析(ribosome profiling, Ribo-seq)可以直接检测全基因组 sORF 的翻译<sup>[21]</sup>。与其他技术不同的是, Ribo-seq 测定核糖体所结合的正在翻译的 mRNA, 也可进行翻译速率的检测。在不同物种中 Ribo-seq 的应用分析显示, 许多被认为缺乏编码潜能的 RNA, 包括长链非编码 RNA (long non-coding RNA, lncRNA)、5'UTR 和 3'UTR<sup>[22-24]</sup>等, 实际上是能被翻译的, 陆续有报道描述了在不同生物的 lncRNA 中检测到小蛋白的表达<sup>[5,11]</sup>。但是, Ribo-seq 所检测到的可能不是翻译本身, 也可能是共纯化的核糖体部分<sup>[25]</sup>, 因此, 提高 Ribo-seq 对于 sORF 的检测精度, 需要在实验技术上进一步改进或加强后续数据集的计算分析<sup>[26]</sup>。技术上的改进包括使用 harringtonin 或 lactimidomycin 使核糖体停留在起始密码子上, 以分类翻译起始位点<sup>[27-29]</sup>, 以及对翻译核糖体进行亲和分离, 避免超速离心时不相关核糖体的共同纯化<sup>[30]</sup>; Aspden 等提出 Poly-Ribo-Seq 方法, 该方法通过只分离出代表活性翻译的多聚体用于足迹分析<sup>[31]</sup>; Archer 等提出翻译复合体图谱测序(TCP-seq)方法, 在常规 80S 保护印迹的基础上, 利用甲醛交联分离出小核糖体亚基相关的 mRNA 片段, 从而获得翻译起始和延伸信息<sup>[32]</sup>。另一方面, 越来越多的基于 Ribo-seq

数据的检测分析工具逐渐发展起来, 包括 RiboTaper、ORFRATER 和 ORF finder<sup>[33]</sup>。此外, 建立了完善的指标来评估这些预测的编码潜力, 如 FLOSS、ORFscore 和 RRS<sup>[4,22,34]</sup>。尽管翻译组分析对确定 sORF 是否有翻译活性非常重要, 但翻译的结果并不一定会产生稳定存在于细胞中的功能性小蛋白。因此, 以质谱为基础的翻译产物识别对于小蛋白的鉴定也至关重要。

### 2.3 质谱分析鉴定小蛋白

质谱分析是一种直接检测多肽的方法。该技术应用液相色谱 (liquid chromatography, LC), 串联质谱 (tandem mass spectrometry, MS/MS) 进行分析。通过获得的 MS/MS 光谱中的多肽与参考蛋白序列数据库中所有候选多肽的理论光谱进行匹配。最常用的数据库包括 Ensembl、RefSeq 和 UniProtKB。目前存在的问题和挑战包括: 一是由于 sORF 的多态位点、可变剪接或缺乏注释, 使得许多小蛋白尚未在参考数据库中得到注释; 二是长度较短, 可能没有或很少包含合适的胰蛋白酶酶切位点, 从而使它们相对于较大和更稳定的蛋白质产生偏差; 三是丰度较低。经典自下而上的蛋白质组学使用数据依赖模式 (data dependent analysis, DDA) 来识别酶切蛋白, 导致了对高丰度肽的偏倚, 负面影响了对微量蛋白的检测。针对以上这些问题, 研究人员一方面提出了针对小蛋白鉴定的方法, 包括利用不同的蛋白酶或自上向下分析天然小蛋白, 优化提取缓冲液、大小选择或分馏方案, 以提高检测灵敏度<sup>[5,35-36]</sup>; 另一方面是建立定制的数据库, 如 Guruceaga 等提出了蛋白质基因组 (proteogenomic) 工作流程, 整合了基因组、转录组和蛋白质组数据, 根据测序信息建立自定义搜索数据库, 可以提高识别新肽的能力<sup>[37]</sup>。但是, 由于蛋白质组学方法还不够灵敏, 且考虑到 Ribo-seq 与

最终蛋白产物的密切关系, 利用 Ribo-seq 数据替代 RNA-seq 构建自定义数据库是另一种扩展 MS 搜索空间的选择<sup>[38-39]</sup>, 例如 Martinez 等将从头进行转录组组装和核糖体分析集成改进, 在人类细胞系中注释了数千个 sORF<sup>[29,40]</sup>。

另外, 数据处理也是影响检测结果的重要因素。新工具的开发促进了质谱技术的数据分析, 如数据独立获取 (data-independent acquisition, DIA) 质谱<sup>[41]</sup>, 系统地选择了整个质量范围进行进一步的碎片化, 增加了检测单个肽的概率, 为低丰富的蛋白质检测打开新的机会。

## 3 微生物小蛋白功能的多样性和作用机制

目前, 对原核微生物的新小蛋白研究表明, 小蛋白的鉴定具有广阔的挖掘空间。对肠道沙门氏菌的一项研究鉴定了 130 个未注释的 sORF, 并证实了 25 个新小蛋白的合成, 其中 Mia-28 (24 aa)、Mia-31 (13 aa)、Mia-63 (45 aa) 和 STM14\_1499 (35 aa) 4 个小蛋白在低镁胁迫下可诱导该菌被巨噬细胞吞噬<sup>[42]</sup>。在致病性大肠杆菌 O157:H7 Saka 中, Ribo-seq 分析为 14 个 sORF 以及几个稍长的 ORF 的翻译提供了证据, 其中一个潜在的小蛋白 (X049, 38 aa) 的合成被质谱证实<sup>[43]</sup>。2018 年, van Orsdel 等通过在染色体上一个 sORF 的 3' 端添加一个表位标签, 通过免疫印迹试验鉴定出 36 个新的大肠杆菌小蛋白, 提示大肠杆菌中仍存在大量尚未鉴定的小蛋白<sup>[44]</sup>。

在真核模式真菌中, 对酿酒酵母 (*Saccharomyces cerevisiae*) ATCC 204508/S288c 中小蛋白研究相对较多。酿酒酵母基因组中共有 16 条染色体, 6 275 个基因, 其中 5 885 个被鉴定为蛋白编码基因<sup>[45]</sup>。根据 UniProt 蛋白数据库的收录结果, 酿酒酵母 ATCC 204508/S288c

基因组中共有 515 个编码小蛋白(少于 100 aa)的 sORF, 占其可编码基因的 8.75%, 其中 377 个 sORF 编码的小蛋白的功能是未知的, 占了小蛋白编码基因的 73.2%, 说明酿酒酵母基因组中存在许多有待被功能注释的小蛋白, 也提示对酿酒酵母小蛋白的研究还存在着非常大的空间。酿酒酵母 S288c 基因组中小蛋白的长度分析显示, UniProt 数据库收录的酿酒酵母 515 个 sORF 中, 320 个 sORF 编码的小蛋白的长度在 61 至 100 个氨基酸之间, 136 个 sORF 编码少于 50 个氨基酸的小蛋白, 且大部分功能未知, 其中, 11 个功能已知的小蛋白(25–50 aa)与核糖体蛋白、线粒体、氨基酸合成调控功能相关。除了鉴定技术更倾向鉴定分子量较大的小蛋白外, 较短的小蛋白的疏水倾向特性较强以及小蛋白只在特定条件下表达, 这可能是造成小蛋白的鉴定偏向较长蛋白的原因<sup>[46]</sup>。

早在 30 年前, 有研究报道酿酒酵母基因组中 sORF 可能具有编码蛋白功能<sup>[47]</sup>, 继而陆续发现具有不同生物学功能的小蛋白。2006 年, Kastenmayer 等将酵母基因组中 sORF 编码的小蛋白进行同源性分析, 发现裂殖酵母(*Schizosaccharomyces pombe*)、蚯蚓(*Pheretima*)、拟南芥(*Arabidopsis thaliana*)、果蝇(*Drosophila melanogaster*)和人类等代表性真核生物中, 与酿酒酵母相似的小蛋白约占 60%, 提示了这类相对保守的 sORF 编码的小蛋白在真核生物中可能参与必要的生物功能, 从而也启示了利用同源序列挖掘及研究小蛋白功能的有效性<sup>[48]</sup>。Brar 等在 2012 年利用 Ribo-seq 发现酿酒酵母减数分裂过程中大量未注释 sORF 的增强翻译, 从而提出小蛋白可能与减数分裂调控相关, 提示了减数分裂中普遍存在翻译控制<sup>[49]</sup>。进一步在检测 2 500 个新预测的 sORF 过程中, Hollerer 等发现在酿酒酵母减数分裂过程中

sORF 的表达约占细胞翻译能力的 10%–20%, 这一发现有助于阐明减数分裂细胞广泛重组的分子基础<sup>[50]</sup>。He 等通过蛋白质富集方法成功鉴定了酿酒酵母 S288c 中 117 个小蛋白, 发现其中 31 个可能参与能量转换途径、细胞壁合成与组装、端粒保护、转录调控、翻译过程、组氨酸合成、蛋白质胞内转运和分泌过程及蛋白折叠等过程; 通过比较 4 种不同的富集策略(Urea\_Tricine、HCl\_Tricine、Urea\_MWCO 和 HCl\_MWCO)并鉴定了 3 个因小蛋白不稳定原因而漏注释的 sORF (YKL104W-A、YHR052C-B 和 YHR054C-B), 但该研究并没有进一步解析该 3 个小蛋白的具体功能<sup>[51]</sup>。

环境胁迫条件下基因表达重编程是细胞应对逆境胁迫的关键机制<sup>[52]</sup>, 值得注意的是“隐匿”的 sORF 所编码的小蛋白在胁迫逆境中发挥着重要作用。Kessler 等在 2003 年通过酿酒酵母全基因组预测蛋白序列的比较, 分析出 117 个 sORF, 其中 84 个可发生转录, 选取在多种生物中具有同源性的 smORF2 验证发现该 sORF 可被翻译成蛋白。酿酒酵母基因 *smORF2* 缺失突变体在 37 °C 下无法生长, 在 30 °C 下生长缓慢, 推测该基因与温度敏感相关<sup>[53]</sup>。Kastenmayer 等曾以构建基因缺失突变体的策略来验证酿酒酵母在高温胁迫、非发酵碳源及 DNA 受损物质存在的情况下生长所需的 sORF, 发现 YJL062W-A、YPL189C-A 和 YDR524W-C 编码的小蛋白的缺失引起了相应的表型缺陷<sup>[48]</sup>。López-Martínez 等实验证明小蛋白 Stf2p 及 Sip8p 有助于提高酿酒酵母脱水后的存活率<sup>[54–55]</sup>。此外, Vargas-Maya 等鉴定了酿酒酵母一个编码未知功能蛋白的 sORF YNR034W-A, 基因表达图谱显示 YNR034W-A 在高乙醇浓度及高糖胁迫条件下表达量显著增加, 且在酿酒酵母过表达菌株中有助于提高龙舌兰酒的发酵效率, 推

测该 sORF 在高酒精浓度等环境胁迫的耐受性方面具有一定的作用<sup>[56]</sup>。以上这些发现说明, 环境胁迫响应的 sORF 的挖掘对构建抗逆性增强的生产菌株具有很好的应用前景。表 1 总结了报道的酿酒酵母中与抗逆性和非发酵碳源等利用有关的小蛋白。

近年来也报道了一些丝状真菌小蛋白的分析和功能。例如, 在植物病原真菌灰霉菌 (*Botrytis cinerea*) 感染期的分泌物中发现了一个植物毒性小蛋白 BcSSP2 (90 aa), 该小蛋白富含半胱氨酸且没有任何已知的结构域, 进一步的研究表明该小蛋白受到类受体激酶 BAK1 和 SOBIR1 的负调控, 可诱导植物产生免疫抗性<sup>[66]</sup>。丝状真菌的 G 蛋白  $\gamma$  亚基与其生长发育、致病性等有关, 最近, 国内研究者发现草酸青霉 (*Penicillium oxalicum*) 的 G 蛋白  $\gamma$  亚基可调节植物生物物质降解酶(如纤维素酶、木聚糖酶和生淀粉酶)的生物合成<sup>[67]</sup>。对 31 个真菌基因组进行了保守小蛋白的分析发现, 1 986 个相对保守的小蛋白, 其中三分之一没有功能注释, 包括丝状真菌中的黑曲霉 (*Aspergillus niger*)、米曲

霉 (*A. oryzae*)、构巢曲霉 (*A. nidulans*) 和烟曲霉 (*A. fumigatus*) 等<sup>[68]</sup>。这些研究表明, 小蛋白对丝状真菌的发育、致病真菌防治和工业应用具有潜在的重要作用, 加强对微生物小蛋白的功能研究非常必要。

## 4 微生物 sORF 的翻译调节功能

近期, 我国学者利用生物信息学方法对多个真核生物进行 sORF 预测, 发现木霉 (*Trichoderma*) 和青霉 (*Penicillium*) 基因组存在 1 万到 5 万多个 uORF, 在产黄青霉 (*P. chrysogenum*) 中有 1.44 万个菌种独特的 uORF, 在 *P. copropilum* 中这个数字多达 5.8 万个<sup>[68]</sup>。这些 uORF 具有十分广阔的研究前景<sup>[8]</sup>, 将这些调控元件有效地与相应的生物工程技术结合, 具有广泛的应用价值, 例如, 在提高菌株耐受性和减少改造菌株的多效性以提高目标产物产量等方面。基于现有的对酿酒酵母的 5'UTR 的研究, 研究人员已构建出预测蛋白质丰度-yUTR Calculator<sup>[69]</sup>, 提示 5'UTR 调控元件(包括 uORF、Kozak 序列、上游起始密码子等)仍有待发掘<sup>[70]</sup>。虽然 sORF

表 1 酿酒酵母 sORF 编码的功能小蛋白举例

Table 1 Examples of small proteins encoded by sORFs in *Saccharomyces cerevisiae*

Function	Gene name	Systematic name	Length (amino acids)	References
Aiding in chromosome segregation	<i>DAD3</i>	YBR233W-A	94	[48]
Thermos-sensitivity	<i>smORF2</i>	YBL071W-A	82	[53]
Resistance to dehydration	<i>STF2</i>	YGR008C	84	[54]
Resistance to high sugar and alcohol concentrations	<i>EGO4</i>	YNR034W-A	98	[56]
ER to Golgi transport	<i>YOS1</i>	YERO74W	87	[57]
Pseudouridylation and processing of pre-18S rRNA	<i>NOP10</i>	YHRO72W	58	[58]
Mitochondrial protein required for outer membrane protein import	<i>MIM2</i>	YLR099W-A	87	[59]
Potentially function in the secretory pathway	<i>KSH1</i>	YNL024C-A	72	[60]
Splicing and for assembly of SF3b	<i>SF3</i>	YNL138W-A	85	[61]
DNA damage and growth	<i>TSC3</i>	YBR058C-A	80	[62]
Subunit of the mitochondrial inner membrane peptidase	<i>SOM1</i>	YEL059C-A	74	[63]
Cytochrome oxidase assembly factor	<i>COA2</i>	YPL189C-A	68	[64]
Phospholipid-binding hydrophilin	<i>SIP18</i>	YMR175W	79	[65]



作为调控元件在微生物代谢工程改造领域的应用尚未见报道,但是在植物中利用 sORF 所介导的翻译调控,尤其是 uORF 的翻译调控进行育种已有一些成功的先例<sup>[71-73]</sup>,为提高植物的耐受性、抗病性提供了精准的调控。例如,将病原菌的 TBF1 的 uORF 所介导的翻译调控导入拟南芥和水稻中,以调控自激活免疫受体 Snc1-1 和 AtNPR1 的产生,成功提高所调控相应植株的抗病性<sup>[74]</sup>。将加速细胞的 ACD11 的 uORF 所介导的翻译调控导入拟南芥中,植株提高表达凝集素受体激酶 AtLecRKVI.2,增强了植株对辣椒疫霉的抗性<sup>[75]</sup>。这些新的发现和应用为利用 sORF 进行微生物菌种选育提供了可行的参考<sup>[3,22]</sup>。

## 5 sORF 功能和机理的研究策略

深入探究 sORF 编码的小蛋白,包括其功能作用、互作网络等,需要建立可靠的研究策略来实现。基于 CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-associated protein 9)基因组编辑技术极大促进了对 sORF 编码的小蛋白的高通量功能鉴定。利用 CRISPR-Cas9 在酿酒酵母中构建 DNA 序列突变体库,从 315 个 sORF 中分析出 68 个可能与抗真菌药物抗性以及 DNA 损伤抗性功能相关的功能性小蛋白<sup>[76]</sup>;在人类细胞系中,利用基于 CRISPR-Cas9 筛选策略识别数百个 alt-ORF,并鉴定了 sORF 编码小蛋白的功能特征及白细胞抗原系统的新元件<sup>[77]</sup>。对于小蛋白互作网络的探究可利用传统的免疫沉淀,但该方法往往会有小蛋白的非特异性相互作用的富集,为此,Chu 等应用了依赖于抗坏血酸过氧化物酶 2 (APEX)的原位邻近标记方法,来探究小蛋白-蛋白相互作用<sup>[78]</sup>。Koh 等将光交联的非典型氨基酸 AbK 整合到 sORF 编码的小蛋白中,

利用基于亲和性的方法有效地识别了与之相互作用伙伴,成功地发现了一个由 sORF 编码的组蛋白结合蛋白(SEHBP),并且发现该小蛋白是一种保守的转录因子,在人类细胞中过表达时诱导一个强大的转录程序<sup>[79]</sup>。考虑到荧光标签所增加的分子量和结构可能会显著影响小蛋白在体内的生理生化特性,Lafranchi 等开发了单残基末端标记(STELLA)标签,实现在小蛋白的氨基端或羧基端引入一个非典型氨基酸,用于后续的特异性荧光标记,这种方法提供了一种普遍适用且易于扩展的策略,避免了小蛋白核心序列的改变<sup>[80]</sup>。另一方面,目前通过报告基因探究 sORF 的翻译调控作用是可行有效的研究策略。此外,Jiménez-Bremont 等将拟南芥多胺氧化酶 2 (AtPAO2) 5'UTR 区域的 uORF 及其突变体分别与 GUS 报告基因进行融合,探究了 AtPAO2 uORF 的翻译对主 ORF 翻译抑制的必要性<sup>[81]</sup>。Ai 等通过开发的荧光报告系统检测了加速细胞的 ACD11 的 uORF (uORF<sub>ACD11</sub>) 的翻译调控作用<sup>[75]</sup>。但这种传统的 uORF 功能分析限制了 uORF 的研究规模,近来,Mcmanus 等结合流式细胞仪同时检测评估了酵母的数千个 uORF 的功能及其影响,为探究 uORF 的翻译调节功能提供高通量的策略<sup>[82-83]</sup>。以上这些研究为更广泛探索微生物中 sORF 及其编码的小蛋白提供了方法学上的借鉴。

## 6 结语与展望

生物信息学的进步有助于促进 sORF 编码小蛋白的挖掘,随着转录组学、蛋白组学、核糖体分析、质谱分析、基因组编辑和机器学习<sup>[84]</sup>等技术的发展,使小蛋白的鉴定变得高效和可靠。未来针对小蛋白注释和分析的生物信息技术将不断完善。目前小蛋白数据库主要收录的仍然是模式物种,未来也将进一步扩展到更多

的物种。同时，鉴定小蛋白的质谱检测技术也将进一步成熟，并借助生物信息分析技术发现更多的小蛋白。高通量的遗传操作技术，尤其是基于 CRISPR-Cas 系统的基因编辑技术将为验证小蛋白和 sORFs 的功能提供实验证据。可以预见，未来生物信息学、质谱分析技术以及基因功能实验等多学科多领域的更充分结合，将为小蛋白的鉴定提供更有力的支撑。目前越来越多被遗漏的 sORF 和小蛋白陆续在人类细胞系和植物中被发现，预计未来会有更多相关成果报道。而微生物底盘细胞生长快、应用广和操作简易，亦有作为未来聚焦的研究对象的优势。

由于 sORF 长度短，易合成，在作为合成生物学元件方面具有独到的优势，预计能在人工调节复合体功能和活性方面做出突出的贡献。小蛋白的分子结构及代谢机制的揭示有助于促进小蛋白的开发应用，例如，对小蛋白在耐受性方面的功能研究有利于选育高抗逆的工业菌株，克服产物和底物中毒性物质的抑制。小蛋白分子量小，具渗透性及灵活性，适合应用于药物设计，尤其是小分子药物设计。但是，目前亟需克服的瓶颈依然是 sORF 及小蛋白的识别与功能鉴定。一方面，如何用多组学技术挖掘 sORF 及其编码的小蛋白，决定了我们打开“黑匣子”的程度，另一方面，目前以经典的功能缺失的遗传分析方法鉴定 sORF 功能，其最大缺陷是容易忽略对表型改变不明显的 sORF 突变体的发现，其实这些 sORF 是有功能的，因此需要考虑增强相关的表型，比如可以通过不同强度过表达 sORF 进行功能鉴定，以及开发其他不同的功能鉴定策略。

另外，目前 sORF 编码的小蛋白研究受到了普遍的关注，已经证明很多小蛋白能提高微生物的抗逆性，但是 uORF 作为有效的顺式调

控元件和翻译调控盒在微生物工程领域中的应用尚属空白。这种精细调控可减少传统过表达所产生的多效反应<sup>[75]</sup>，目前在植物抗病性方面得到良好的应用。未来相关研究应用于微生物的合成生物学和代谢工程改造有可能实现对产物的精细动态调控。

综上所述，sORF 及其编码的小蛋白的鉴定、功能和作用机理研究，以及其在生物工程领域的应用存在着巨大的潜力，值得进一步深入挖掘与探索。此外，微生物 sORF 的研究也将为研究植物和人类疾病治疗等高等生物相关的研究提供借鉴。

## 参考文献

- [1] KUTE PM, SOUKARIEH O, TJELDNE H, TRÉGOUËT DA, VALEN E. Small open reading frames, how to find them and determine their function[J]. *Frontiers in Genetics*, 2022, 12: 796060.
- [2] FUCHS S, KUCKLICK M, LEHMANN E, BECKMANN A, WILKENS M, KOLTE B, MUSTAFAYEVA A, LUDWIG T, DIWO M, WISSING J, JÄNSCH L, AHRENS CH, IGNATOVA Z, ENGELMANN S. Towards the characterization of the hidden world of small proteins in *Staphylococcus aureus*, a proteogenomics approach[J]. *PLoS Genetics*, 2021, 17(6): e1009585.
- [3] BRUNET MA, LEBLANC S, ROUCOU X. Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs[J]. *Experimental Cell Research*, 2020, 393(1): 112057.
- [4] CHUGUNOVA A, NAVALAYEU T, DONTSOVA O, SERGIEV P. Mining for small translated ORFs[J]. *Journal of Proteome Research*, 2018, 17(1): 1-11.
- [5] SCHLESINGER D, ELSÄSSER SJ. Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins[J]. *The FEBS Journal*, 2022, 289(1): 53-74.
- [6] 陈相颖, 李梦玮, 王颖, 陈权, 徐寒梅. 小开放阅读框编码微肽的研究进展[J]. *遗传*, 2021, 43(8): 737-746.  
CHEN XY, LI MW, WANG Y, CHEN Q, XU HM. Progress on sORF-encoded micropeptides[J]. *Hereditas: Beijing*, 2021, 43(8): 737-746 (in Chinese).

- [7] LEONG AZX, LEE PY, MOHTAR MA, SYAFRUDDIN SE, PUNG YF, LOW TY. Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures[J]. *Journal of Biomedical Science*, 2022, 29(1): 19.
- [8] ZHANG H, WANG YR, WU XK, TANG XL, WU CC, LU J. Determinants of genome-wide distribution and evolution of uORFs in eukaryotes[J]. *Nature Communications*, 2021, 12: 1076.
- [9] PLAZA S, MENSCHAERT G, PAYRE F. In search of lost small peptides[J]. *Annual Review of Cell and Developmental Biology*, 2017, 33: 391-416.
- [10] ORR MW, MAO YH, STORZ G, QIAN SB. Alternative ORFs and small ORFs: shedding light on the dark proteome[J]. *Nucleic Acids Research*, 2020, 48(3): 1029-1042.
- [11] 黎秋慧, 陈瑜丽, 刘向华. LncRNA 编码小肽的功能及研究现状[J]. *中国生物化学与分子生物学报*, 2021, 37(12): 1577-1583.
- LI QH, CHEN YL, LIU XH. Function and research status of small peptides encoded by LncRNA[J]. *Chinese Journal of Biochemistry and Molecular Biology*, 2021, 37(12): 1577-1583 (in Chinese).
- [12] YIN XQ, JING YY, XU HM. Mining for missed sORF-encoded peptides[J]. *Expert Review of Proteomics*, 2019, 16(3): 257-266.
- [13] IVANOV IP, FIRTH AE, MICHEL AM, ATKINS JF, BARANOV PV. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences[J]. *Nucleic Acids Research*, 2011, 39(10): 4220-4234.
- [14] IVANOV IP, WEI JJ, CASTER SZ, SMITH KM, MICHEL AM, ZHANG Y, FIRTH AE, FREITAG M, DUNLAP JC, BELL-PEDERSEN D, ATKINS JF, SACHS MS. Translation initiation from conserved non-AUG codons provides additional layers of regulation and coding capacity[J]. *mBio*, 2017, 8(3): e00844-e00817.
- [15] ZHU MM, GRIBSKOV M. MiPepid: MicroPeptide identification tool using machine learning[J]. *BMC Bioinformatics*, 2019, 20(1): 559.
- [16] JI XW, CUI CM, CUI QH. smORFunction: a tool for predicting functions of small open reading frames and microproteins[J]. *BMC Bioinformatics*, 2020, 21(1): 455.
- [17] WANG T, CUI YZ, JIN JJ, GUO JH, WANG GB, YIN XF, HE QY, ZHANG G. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific[J]. *Nucleic Acids Research*, 2013, 41(9): 4743-4754.
- [18] CHASSÉ H, BOULBEN S, COSTACHE V, CORMIER P, MORALES J. Analysis of translation using polysome profiling[J]. *Nucleic Acids Research*, 2017, 45(3): e15.
- [19] WANG Y, JIAO YL. Translating ribosome affinity purification (TRAP) for cell-specific translation profiling in developing flowers[A]//*Methods in Molecular Biology*[M]. New York, NY: Springer New York, 2013: 323-328.
- [20] ZHAO J, QIN B, NIKOLAY R, SPAHN CMT, ZHANG G. Translatomics: the global view of translation[J]. *International Journal of Molecular Sciences*, 2019, 20(1): 212.
- [21] INGOLIA NT, GHAEMMAGHAMI S, NEWMAN JRS, WEISSMAN JS. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling[J]. *Science*, 2009, 324(5924): 218-223.
- [22] CHOI SW, KIM HW, NAM JW. The small peptide world in long noncoding RNAs[J]. *Briefings in Bioinformatics*, 2019, 20(5): 1853-1864.
- [23] FESENKO I, KIROV I, KNIAZEV A, KHAZIGALEEVA R, LAZAREV V, KHARLAMPIEVA D, GRAFSKAIA E, ZGODA V, BUTENKO I, ARAPIDI G, MAMAEVA A, IVANOV V, GOVORUN V. Distinct types of short open reading frames are translated in plant cells[J]. *Genome Research*, 2019, 29(9): 1464-1477.
- [24] WU QS, WRIGHT M, GOGOL MM, BRADFORD WD, ZHANG N, BAZZINI AA. Translation of small downstream ORFs enhances translation of canonical main open reading frames[J]. *The EMBO Journal*, 2020, 39(17): e104763.
- [25] WILSON BA, MASEL J. Putatively noncoding transcripts show extensive association with ribosomes[J]. *Genome Biology and Evolution*, 2011, 3: 1245-1252.
- [26] INGOLIA NT, HUSSMANN JA, WEISSMAN JS. Ribosome profiling: global views of translation[J]. *Cold Spring Harbor Perspectives in Biology*, 2019, 11(5): a032698.
- [27] LEE S, LIU BT, LEE S, HUANG SX, SHEN B, QIAN SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(37): E2424-E2432.
- [28] GAO XW, WAN J, LIU BT, MA M, SHEN B, QIAN SB. Quantitative profiling of initiating ribosomes *in*

- vivo*[J]. *Nature Methods*, 2015, 12(2): 147-153.
- [29] MARTINEZ TF, CHU Q, DONALDSON C, TAN D, SHOKHIREV MN, SAGHATELIAN A. Accurate annotation of human protein-coding small open reading frames[J]. *Nature Chemical Biology*, 2020, 16(4): 458-468.
- [30] INGOLIA NT, BRAR GA, STERN-GINOSSAR N, HARRIS MS, TALHOUARNE GJS, JACKSON SE, WILLS MR, WEISSMAN JS. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes[J]. *Cell Reports*, 2014, 8(5): 1365-1379.
- [31] ASPDEN JL, EYRE-WALKER YC, PHILLIPS RJ, AMIN U, MUMTAZ MAS, BROCARD M, COUSO JP. Extensive translation of small open reading frames revealed by Poly-Ribo-Seq[J]. *eLife*, 2014, 3: e03528.
- [32] ARCHER SK, SHIROKIKH NE, BEILHARZ TH, PREISS T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling[J]. *Nature*, 2016, 535(7613): 570-574.
- [33] DURRANT MG, BHATT AS. Automated prediction and annotation of small open reading frames in microbial genomes[J]. *Cell Host & Microbe*, 2021, 29(1): 121-131.e4.
- [34] CALVIELLO L, OHLER U. Beyond read-counts: ribo-seq data analysis to understand the functions of the transcriptome[J]. *Trends in Genetics*, 2017, 33(10): 728-744.
- [35] CASSIDY L, HELBIG AO, KAULICH PT, WEIDENBACH K, SCHMITZ RA, THOLEY A. Multidimensional separation schemes enhance the identification and molecular characterization of low molecular weight proteomes and short open reading frame-encoded peptides in top-down proteomics[J]. *Journal of Proteomics*, 2021, 230: 103988.
- [36] KAULICH PT, CASSIDY L, WEIDENBACH K, SCHMITZ RA, THOLEY A. Complementarity of different SDS-PAGE gel staining methods for the identification of short open reading frame-encoded peptides[J]. *Proteomics*, 2020, 20(19/20): 2000084.
- [37] GURUCEAGA E, GARIN-MUGA A, SEGURA V. MiTPeptideDB: a proteogenomic resource for the discovery of novel peptides[J]. *Bioinformatics*, 2020, 36(1): 205-211.
- [38] BUDAMGUNTA H, OLEXIOUK V, LUYTEN W, SCHILDERMANS K, MAES E, BOONEN K, MENSCHAERT G, BAGGERMAN G. Comprehensive peptide analysis of mouse brain *Striatum* identifies novel sORF-encoded polypeptides[J]. *Proteomics*, 2018, 18(10): e1700218.
- [39] VERBRUGGEN S, NDAH E, van CRIEKINGE W, GESSULAT S, KUSTER B, WILHELM M, van DAMME P, MENSCHAERT G. PROTEOFORMER 2.0: further developments in the ribosome profiling-assisted proteogenomic hunt for new proteoforms\*[J]. *Molecular & Cellular Proteomics*, 2019, 18(8): S126-S140.
- [40] MA J, SAGHATELIAN A, SHOKHIREV MN. The influence of transcript assembly on the proteogenomics discovery of microproteins[J]. *PLoS One*, 2018, 13(3): e0194518.
- [41] LU YY, BILMES J, RODRIGUEZ-MIAS RA, VILLÉN J, NOBLE WS. DIAMeter: matching peptides to data-independent acquisition mass spectrometry data[J]. *Bioinformatics*, 2021, 37(Supplement\_1): i434-i442.
- [42] OGURI T, KWON Y, WOO JKK, PREHNA G, LEE H, NING MR, WON KJ, LEE J, MEI S, SHI YX, JEONG H, LEE H. A family of small intrinsically disordered proteins involved in flagellum-dependent motility in *Salmonella enterica*[J]. *Journal of Bacteriology*, 2018, 201(2): e00415-e00418.
- [43] NEUHAUS K, LANDSTORFER R, FELLNER L, SIMON S, SCHAFFERHANS A, GOLDBERG T, MARX H, OZOLINE ON, ROST B, KUSTER B, KEIM DA, SCHERER S. Translatomics combined with transcriptomics and proteomics reveals novel functional, recently evolved orphan genes in *Escherichia coli* O157:H7 (EHEC)[J]. *BMC Genomics*, 2016, 17(1): 133.
- [44] van ORSDEL CE, KELLY JP, BURKE BN, LEIN CD, OUFIERO CE, SANCHEZ JF, WIMMERS LE, HEARN DJ, ABUIKHDIAIR FJ, BARNHART KR, DULEY ML, ERNST SEG, KENERSON BA, SERAFIN AJ, HEMM MR. Identifying new small proteins in *Escherichia coli*[J]. *Proteomics*, 2018, 18(10): e1700064.
- [45] 陈碧燕, 文李. 酿酒酵母全基因组学及其应用研究进展[J]. *食品与机械*, 2020, 36(11): 223-227, 232. CHEN BY, WEN L. Progress in *Saccharomyces cerevisiae* genome research and relative application[J]. *Food & Machinery*, 2020, 36(11): 223-227, 232 (in Chinese).
- [46] 何崔同. 酵母小蛋白的蛋白质基因组学研究[D]. 合肥: 安徽医科大学硕士学位论文, 2018. HE CT. Proteogenomics study on small proteome of

- Saccharomyces cerevisiae*[D]. Hefei: Master's Thesis of Anhui Medical University, 2018 (in Chinese)
- [47] ERPF PE, FRASER JA. The long history of the diverse roles of short ORFs: SPEPs in fungi[J]. *Proteomics*, 2018, 18(10): e1700219.
- [48] KASTENMAYER JP, NI L, CHU A, KITCHEN LE, AU WC, YANG H, CARTER CD, WHEELER D, DAVIS RW, BOEKE JD, SNYDER MA, BASRAI MA. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*[J]. *Genome Research*, 2006, 16(3): 365-373.
- [49] BRAR GA, YASSOUR M, FRIEDMAN N, REGEV A, INGOLIA NT, WEISSMAN JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling[J]. *Science*, 2012, 335(6068): 552-557.
- [50] HOLLERER I, HIGDON A, BRAR GA. Strategies and challenges in identifying function for thousands of sORF-encoded peptides in meiosis[J]. *PROTEOMICS*, 2018, 18(10): 1700274.
- [51] HE CT, JIA CX, ZHANG Y, XU P. Enrichment-based proteogenomics identifies microproteins, missing proteins, and novel smORFs in *Saccharomyces cerevisiae*[J]. *Journal of Proteome Research*, 2018, 17(7): 2335-2344.
- [52] TAYMAZ-NIKEREL H, CANKORUR-CETINKAYA A, KIRDAR B. Genome-wide transcriptional response of *Saccharomyces cerevisiae* to stress-induced perturbations[J]. *Frontiers in Bioengineering and Biotechnology*, 2016, 4: 17.
- [53] KESSLER MM, ZENG QD, HOGAN S, COOK R, MORALES AJ, COTTAREL G. Systematic discovery of new genes in the *Saccharomyces cerevisiae* genome[J]. *Genome Research*, 2003, 13(2): 264-271.
- [54] LÓPEZ-MARTÍNEZ G, RODRÍGUEZ-PORRATA B, MARGALEF-CATALÀ M, CORDERO-OTERO R. The STF2p hydrophilin from *Saccharomyces cerevisiae* is required for dehydration stress tolerance[J]. *PLoS One*, 2012, 7(3): e33324.
- [55] LÓPEZ-MARTÍNEZ G, PIETRAFESA R, ROMANO P, CORDERO-OTERO R, CAPECE A. Genetic improvement of *Saccharomyces cerevisiae* wine strains for enhancing cell viability after desiccation stress[J]. *Yeast: Chichester, England*, 2013, 30(8): 319-330.
- [56] VARGAS-MAYA NI, GONZÁLEZ-HERNÁNDEZ GA, PADILLA-GUERRERO IE, TORRES-GUZMÁN JC. Overexpression of smORF YNR034W-A/*EGO4* in *Saccharomyces cerevisiae* increases the fermentative efficiency of *Agave tequilana* Weber must[J]. *Journal of Industrial Microbiology & Biotechnology*, 2017, 44(1): 63-74.
- [57] HEIDTMAN M, CHEN CZ, COLLINS RN, BARLOWE C. Yos1p is a novel subunit of the Yip1p-Yif1p complex and is required for transport between the endoplasmic reticulum and the Golgi complex[J]. *Molecular Biology of the Cell*, 2005, 16(4): 1673-1683.
- [58] REICHOW SL, VARANI G. Nop10 is a conserved H/ACA snoRNP molecular adaptor[J]. *Biochemistry*, 2008, 47(23): 6148-6156.
- [59] DIMMER KS, PAPIĆ D, SCHUMANN B, SPERL D, KRUMPE K, WALTHER DM, RAPAPORT D. A crucial role for Mim2 in the biogenesis of mitochondrial outer membrane proteins[J]. *Journal of Cell Science*, 2012, 125(Pt 14): 3464-3473.
- [60] WENDLER F, GILLINGHAM AK, SINKA R, ROSA-FERREIRA C, GORDON DE, FRANCH-MARRO X, PEDEN AA, VINCENT JP, MUNRO S. A genome-wide RNA interference screen identifies two novel components of the metazoan secretory pathway[J]. *The EMBO Journal*, 2010, 29(2): 304-314.
- [61] WANG Q, HE J, LYNN B, RYMOND BC. Interactions of the yeast SF3b splicing factor[J]. *Molecular and Cellular Biology*, 2005, 25(24): 10745-10754.
- [62] GABLE K, SLIFE H, BACIKOVA D, MONAGHAN E, DUNN TM. Tsc3p is an 80-amino acid protein associated with serine palmitoyltransferase and required for optimal enzyme activity[J]. *Journal of Biological Chemistry*, 2000, 275(11): 7597-7603.
- [63] ESSER K, PRATJE E, MICHAELIS G. SOM 1, a small new gene required for mitochondrial inner membrane peptidase function in *Saccharomyces cerevisiae*[J]. *Molecular & General Genetics: MGG*, 1996, 252(4): 437-445.
- [64] PIERREL F, KHALIMONCHUK O, COBINE PA, BESTWICK M, WINGE DR. Coa2 is an assembly factor for yeast cytochrome c oxidase biogenesis that facilitates the maturation of Cox1[J]. *Molecular and Cellular Biology*, 2008, 28(16): 4927-4939.
- [65] DANG NX, HINCHA DK. Identification of two hydrophilins that contribute to the desiccation and freezing tolerance of yeast (*Saccharomyces cerevisiae*) cells[J]. *Cryobiology*, 2011, 62(3): 188-193.
- [66] ZHU WJ, YU MX, XU R, BI K, YU S, XIONG C, LIU ZG, SHARON A, JIANG DH, WU MD, GU QN, GONG L, CHEN WD, WEI W. *Botrytis cinerea*

- BcSSP2 protein is a late infection phase, cytotoxic effector[J]. *Environmental Microbiology*, 2022, 24(8): 3420-3435.
- [67] PANG XM, TIAN D, ZHANG T, LIAO LS, LI CX, LUO XM, FENG JX, ZHAO S. G protein  $\gamma$  subunit modulates expression of plant-biomass-degrading enzyme genes and mycelial-development-related genes in *Penicillium oxalicum*[J]. *Applied Microbiology and Biotechnology*, 2021, 105(11): 4675-4691.
- [68] MAT-SHARANI S, FIRDAUS-RAIH M. Computational discovery and annotation of conserved small open reading frames in fungal genomes[J]. *BMC Bioinformatics*, 2019, 19(13): 551.
- [69] DECOENE T, PETERS G, de MAESENEIRE SL, de MEY M. Toward predictable 5'UTRs in *Saccharomyces cerevisiae*: development of a yUTR calculator[J]. *ACS Synthetic Biology*, 2018, 7(2): 622-634.
- [70] de NIJS Y, de MAESENEIRE SL, SOETAERT WK. 5' untranslated regions: the next regulatory sequence in yeast synthetic biology[J]. *Biological Reviews of the Cambridge Philosophical Society*, 2020, 95(2): 517-529.
- [71] ZHANG T, WU AQ, YUE YP, ZHAO Y. uORFs: Important *cis*-regulatory elements in plants[J]. *International Journal of Molecular Sciences*, 2020, 21(17): 6238.
- [72] KURIHARA Y. uORF shuffling fine-tunes gene expression at a deep level of the process[J]. *Plants*: Basel, Switzerland, 2020, 9(5): 608.
- [73] 陈燕君, 王坤. 植物小开放阅读框编码肽的研究进展[J]. *植物科学学报*, 2020, 38(5): 707-715.  
CHEN YJ, WANG K. Advances in studies on small open reading frames in plants[J]. *Plant Science Journal*, 2020, 38(5): 707-715 (in Chinese).
- [74] XU GY, YUAN M, AI CR, LIU LJ, ZHUANG E, KARAPETYAN S, WANG SP, DONG XN. uORF-mediated translation allows engineered plant disease resistance without fitness costs[J]. *Nature*, 2017, 545(7655): 491-494.
- [75] AI G, LIU J, FU XW, LI TL, ZHU H, ZHAI Y, XIA CY, PAN WY, LI JL, JING MF, SHEN DY, XIA A, DOU DL. Making use of plant uORFs to control transgene translation in response to pathogen attack[J]. *BioDesign Research*, 2022, 2022: 1-14.
- [76] GUO XG, CHAVEZ A, TUNG A, CHAN Y, KAAS C, YIN Y, CECCHI R, GARNIER SL, KELSIC ED, SCHUBERT M, DICARLO JE, COLLINS JJ, CHURCH GM. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR-Cas9 in yeast[J]. *Nature Biotechnology*, 2018, 36(6): 540-546.
- [77] CHEN J, BRUNNER AD, COGAN JZ, NUÑEZ JK, FIELDS AP, ADAMSON B, ITZHAK DN, LI JY, MANN M, LEONETTI MD, WEISSMAN JS. Pervasive functional translation of noncanonical human open reading frames[J]. *Science*, 2020, 367(6482): 1140-1146.
- [78] CHU Q, RATHORE A, DIEDRICH JK, DONALDSON CJ, YATES JR, SAGHATELIAN A. Identification of microprotein-protein interactions via APEX tagging[J]. *Biochemistry*, 2017, 56(26): 3299-3306.
- [79] KOH M, AHMAD I, KO Y, ZHANG YX, MARTINEZ TF, DIEDRICH JK, CHU Q, MORESCO JJ, ERB MA, SAGHATELIAN A, SCHULTZ PG, BOLLONG MJ. A short ORF-encoded transcriptional regulator[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(4): e2021943118.
- [80] LAFRANCHI L, SCHLESINGER D, KIMLER KJ, ELSÄSSER SJ. Universal single-residue terminal labels for fluorescent live cell imaging of microproteins[J]. *Journal of the American Chemical Society*, 2020, 142(47): 20080-20087.
- [81] GUERRERO-GONZÁLEZ MDLL, ORTEGA-AMARO MA, JUÁREZ-MONTIEL M, JIMÉNEZ-BREMONT JF. *Arabidopsis* polyamine oxidase-2 uORF is required for downstream translational regulation[J]. *Plant Physiology and Biochemistry*, 2016, 108: 381-390.
- [82] LIN YZ, MAY GE, KREADY H, NAZZARO L, MAO M, SPEALMAN P, CREEGER Y, MCMANUS CJ. Impacts of uORF codon identity and position on translation regulation[J]. *Nucleic Acids Research*, 2019, 47(17): 9358-9367.
- [83] MAY GE, MCMANUS CJ. High-throughput quantitation of yeast uORF regulatory impacts using FACS-uORF[J]. *Methods in Molecular Biology*: Clifton, N J, 2022, 2404: 331-351.
- [84] de JONGH RPH, van DIJK ADJ, JULSING MK, SCHAAP PJ, de RIDDER D. Designing eukaryotic gene expression regulation using machine learning[J]. *Trends in Biotechnology*, 2020, 38(2): 191-201.