
研究开发植物tRNA大语言模型

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/39342.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

研究开发植物tRNA大语言模型。近日，中国科学院成都生物研究所张韬、张勇团队及其合作者首次将DNA大语言模型引入植物tRNA功能挖掘，开发出专用于植物tRNA识别与功能预测的人工智能系统——植物tRNA大语言模型（tLLMs）。该研究突破传统生物信息学算法依赖保守结构特征的限制，实现对非典型结构tRNA的高精度识别，为基因组编辑工具开发提供了全新的AI驱动范式。相关成果4月7日发表于《生物技术趋势》上。

植物基因组编辑技术已成为现代生物学研究和作物改良的核心工具。然而，传统CRISPR-Cas系统依赖U3或U6启动子表达向导RNA（gRNA），在多重编辑应用中面临载体容量受限，难以实现高阶多重编辑，同时表达效率不一致，导致编辑效果不稳定。内源性tRNA作为Pol III启动子资源，具有启动子功能与RNA加工单元的双重优势，被视为构建紧凑型高效多重编辑系统的理想选择。但全球范围内缺乏系统性的功能性tRNA挖掘方法，尤其是难以识别结构变异较大的非典型tRNA，这限制了其在植物基因组编辑中的广泛应用。

研究团队基于前期构建的基座模型训练tLLM，使其系统学习植物tRNA的序列-结构-功能关联规律，突破了传统生物信息学算法对保守结构特征的依赖，成功实现了对非典型结构tRNA的高精度识别。借助该模型，团队从植物基因组中挖掘出数千个被传统算法遗漏的功能性tRNA，其中包含大量非典型结构成员。实验验证显示，tLLM预测的17个新型tRNA均展现出基因组编辑活性，预测准确率达100%。这些高活性tRNA广泛跨越拟南芥、水稻、玉米、高粱、大豆、小麦等多个物种，其中AtAsp-tRgtc01和OsAsp-tRgtc01的编辑效率显著优于传统使用的AtGly-tRgcc。

基于上述发现，团队进一步构建了T-tR-sgR-pT紧凑型多重编辑架构，巧妙利用tRNA的双重功能。该架构在水稻中成功实现10个内源位点的同时高效编辑，并在大豆毛状根中构建了五重基因编辑系统，平均突变率达78%，性能显著优于传统系统。此外，将tLLM挖掘的强启动子tRNA拓展应用于先导编辑（Prime Editing），在OsIPA1位点将编辑效率从5%大幅提升至75%，在OsACC1位点更实现了100%的精准碱基替换效率，有效突破了传统方法难以编辑位点的技术限制。

tLLM的建立不仅提供了高效的基因组编辑工具元件，更重要的是确立了“AI预测-实验验证-工程应用”的tRNA功能挖掘新范式。该模型揭示的tRNA序列-功能关联规律，为深入理解Pol III启动子进化与表达调控机制提供了全新视角，也为设计兼顾多重编辑效率与载体稳定性的新一代基因组编辑系统奠定了坚实的理论与技术基础。（来源：中国科学报 杨晨）

相关论文信息：<https://doi.org/10.1016/j.tibtech.2026.02.016>

作者：张韬等 来源：《生物技术趋势》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发