

基于机器学习的植物抗盐碱基因挖掘模型研究

孙善文, 庞韶姿, 解莉楠

东北林业大学, 哈尔滨, 150006

摘要: 土壤盐碱化会导致植物死亡并降低全球作物产量。挖掘新的抗盐碱基因、培育抗盐碱品系是应对盐碱胁迫、保障粮食安全以及物种保护的关键。湿实验是挖掘目标功能基因的主要方法, 但成本高、耗时长。相比之下, 机器学习可以利用大数据进行学习来加速基因挖掘的过程, 使其更加高效, 降低人力和物力成本。但是, 目前利用机器学习挖掘植物功能基因的研究很少。因此, 能够用于抗逆基因挖掘的智能工具还很匮乏。针对这一问题, 我们开发了第一个已知的机器学习模型 **SaGP** 用于挖掘新的植物抗盐碱基因。利用独立验证集对 **SaGP** 进行评估后发现, 它对抗盐碱基因的挖掘能力要高于传统的计算工具 **BLAST** 和 **HMMER**。此外, **SaGP** 正确识别了 3 个最新发表的抗盐碱基因。最后, 我们利用 **SaGP** 在白桦中发现了一个新的抗盐碱基因——*BpHKT1*。这一结果得到了湿实验的验证。这些结果表明 **SaGP** 可用于大规模地挖掘植物抗盐碱基因, 从而促进作物育种和植物保护。