

---

# 机器学习精确定位重要基因

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/15995.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！*

**机器学习精确定位重要基因。** 利用基因组数据预测农业和医学的某些结果，既是希望也是挑战。研究人员一直致力于更好地利用现有的大量基因组数据来预测生物体如何应对营养、毒素和病原体暴露的变化，从而为作物改良、疾病预后、流行病学和公共卫生提供信息。

然而，从基因组大规模的信息中准确预测农业和医学领域如此复杂的结果，仍然是一个重大挑战。美国纽约大学的研究人员利用机器学习解决了这一问题。

根据近日发表于《自然—通讯》的这项新研究，机器学习可以精确定位重要基因，帮助作物在更少的肥料条件下生长。它还可以预测植物的其他性状和动物的疾病结果，表明了它在农业以外的应用。

我们的研究表明，关注在物种间进化上表达模式保守的基因，可以提高我们学习、预测对主要作物生长性能和动物疾病结果‘具有重要意义的基因’的能力。论文资深作者、纽约大学生物学系教授Gloria Coruzzi解释道。

我们的方法利用了物种内或物种间的全基因组表达和相关表型的自然变异。论文资深作者、纽约大学基因组学与系统生物学中心的Carroll Milton Petrie补充说。

他们的结果表明，从生物学原理上讲，减少对物种内和物种间表达模式保守的基因的基因组输入是减少基因组数据维数的一种方法，这极大地提高了机器学习模型的能力来确定哪些基因对某一特征重要的能力。

氮是植物的重要养分，也是肥料的主要成分。更有效地利用氮的作物生长得更好，所需肥料更少。两种不同的植物物种——模式植物拟南芥和美国最大的作物玉米，对氮的反应性基因是在进化上保守的。

研究人员通过实验验证了8个主转录因子对氮利用效率的重要性。结果证明，改变拟南芥或玉米的基因表达可以促进植物在低氮土壤中的生长。他们在纽约大学的实验室和伊利诺伊大学的玉米地中进行了测试。

现在我们利用机器学习模型能够更准确地预测哪些玉米杂交种更善于在田间使用氮肥，便可迅速改进这一性状。提高玉米和其他作物的氮素利用效率可以带来3个主要好处：降低农民成本，减少环境污染，减少农业温室气体排放。主要作者Chia-Yi Cheng说。

---

此外，研究人员证明，这种基于进化的机器学习方法可以应用于其他性状和物种，包括拟南芥、玉米的生物量和产量。他们还表明，通过研究小鼠模型，这种方法可以预测另一种主要作物水稻关于抗旱性的重要基因，以及动物的疾病结果。

Coruzzi说：我们展示了这一方法也可以应用于动物，这表明了它揭示生物学、农业或医学中任何人们感兴趣的生理或临床特征的重要基因潜力。

许多重要的农学或临床特征在遗传上是复杂的，因此很难确定它们的靶点和遗传。这项研究成功证明，大数据和系统层面的思考可以让这些众所周知的困难挑战变得容易应对。作者之一 Ying Li说。（来源：中国科学报王方）

相关论文信息：<https://doi.org/10.1038/s41467-021-25893-w>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：Gloria Coruzzi 来源：《自然—通讯》

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](https://www.iikx.com)转发