



2024年第24期总376期

## 农业生物技术专题

### 本期导读

#### ▶ 前沿资讯

1. 北京理工大学基于迁移学习生成物种特异性启动子方面取得进展

#### ▶ 学术文献

1. 中科院拓宽基因工程工具箱后生动物基因组揭示新活性DNA转座子

2. 山东农大构建了玉米叶枕的单细胞转录组图谱

3. 中国农科院畜牧所基于机器学习模型预测苜蓿群体适应性

4. 北大-清华合作开发蛋白质“扩增测序”新方法

中国农业科学院农业信息研究所

联系人：周诚昊；顾亮亮

联系电话：010-82109850

邮箱：[agri@ckcest.cn](mailto:agri@ckcest.cn)

更多资讯 尽在农业专业知识服务系统：<http://agri.nais.net.cn/>

2024年6月10日

## 前沿资讯

### 1. 北京理工大学基于迁移学习生成物种特异性启动子方面取得进展

**简介:** 近日,北京理工大学霍毅欣教授与郭淑元教授团队在Nucleic Acids Research发表论文,题为“Species-specific design of artificial promoters by transfer-learning based generative deep-learning model”。该工作在物种数据集较少的条件下,训练了较高质量的生成模型PromoGen,用于从头生成物种特异性启动子。启动子是在转录水平上调节基因表达的关键元件,能够启动基因转录、调节基因表达,并影响代谢途径中的代谢流分布。尽管天然启动子已被用于基因调控,但其缺乏连续的调控强度和广泛的调控范围。目前,深度学习在蛋白质设计、调控元件生成等领域已经取得了一定的进展,但是在数据集缺乏的条件下还不能生成质量较高的调控元件。为了解决原核生物启动子数据量不足的问题,团队基于迁移学习的策略,开发了一系列核苷酸语言模型 PromoGen,用于在数据缺乏的条件下从头生成物种特异性的启动子。通过位置权重矩阵、6聚体频率相关性和 -10 区域分布分别对枯草芽胞杆菌(*Bacillus subtilis*)的PromoGen-bsu模型生成启动子的能力进行分析。并对PromoGen-bsu生成的启动子进行湿实验验证,结果表明72.7%的生成启动子的启动活性高于天然启动子PlepA的3倍,18%的启动子与天然强启动子活性水平相当。为了证明迁移学习策略的有效性,分别使用27个物种的启动子的数据,在PromoGen-pre上对其进行微调,得到了27个物种的生成模型。并对所有的模型进行预训练和微调性能评估,微调后32%的模型相关性超过0.8。此外,团队开发了一个在线平台 (<https://promogen1.cloudmol.org/>),该平台针对27种原核生物提供了微调后模型来从头生成启动子。

**来源:** 科微学术

**发布日期:**2024-06-05

**全文链接:**

<http://agri.nais.net.cn/file1/M00/10/41/Csgk0GZhvVSAOauwADUIvirP6tk598.pdf>

## 学术文献

### 1. 中科院拓宽基因工程工具箱后生动物基因组揭示新活性DNA转座子

**简介:** 2024年6月5日,中国科学院动物研究所张勇及王皓毅共同通讯在Cell 在线发表题为“Heterologous survey of 130 DNA transposons in human cells highlights their functional divergence and expands the genome engineering toolbox”的研究论文,该研究从102个后生动物基因组中预测了130个活性DNA TEs,并评估了它们在人类细胞中的活性。在20世纪40年代,Barbara McClintock发现了第一个转座因子(TE),由一对DNA转座子组成,其中包含末端倒置重复序列(TIR)、自主Ac和非自主Ds。McClintock发现TEs通过影响邻近基因的表达作为“控制元件”,通过诱导宿主基因组的变化作为诱变剂。在随后的70年里,该领域经历了一次扩展,发现TE几乎存在于所有真核生物基因组中。在各种类型的TEs中,DNA TEs包括许多超家族,如以Ac/Ds为创始成员的hAT和Tc1/mariner。DNA TEs对宿主的贡献已经引起了广泛的兴趣,它们在各种重要途径中被驯化。同时,从转座活性的潜在因素、进化动力学和基因组工程工具的发展等三个角度对DNA TEs进行了

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

研究。尽管这些研究的规模有限，但它们提供了基本的知识。首先，研究了转位的编码和非编码决定因素。残基包括D(天冬氨酸)，D和E/D(谷氨酸)已被确定为催化核心。TIR由转座结合，一些Tc1/mariner元件在每个TIR(IR-DR)内具有两个不完全直接重复(DR)作为结合位点的特征结构。其次，DNA TEs被认为经历了“水平转移、垂直失活和随机丢失”。自主TE侵入宿主后，会抑制和积累失活突变。此外，包括微型倒置重复TEs(MITEs, 50-800 bp内部缺失的非自主TEs)在内的非自主TEs。由于TIRs的存在和它们的小尺寸而有效地竞争由相关自主TEs编码的转座酶。第三，多种DNA TEs已经发展成为多种应用的工具，包括插入诱变和转基因。其中，Tc1/mariner超家族的Sleeping Beauty(SB)和piggyBac超家族的piggyBac(PB)最受关注。随着编码和/或非编码序列的广泛优化，活性变异(如SB100X)已被开发为有效的非病毒载体。基于SB100x的嵌合抗原受体(CAR) T细胞疗法已被用于治疗血液肿瘤。显然，对少量DNA TEs的偏向和实验研究的异质性使得很难确定TE超家族或家族的一般规则。具体来说，DNA TEs在人细胞中异源表达后的转座活性仍然是不可预测的。进化特征(如超家族类型)和序列特征(如IR-DR)如何影响转位在很大程度上是未知的。在进化动力学方面，虽然已经有报道将MITEs扩增为自主TE的缺失衍生物，但TE超家族是否与MITEs普遍相关尚不清楚。此外，基于TEs的具有多种功能特征的基因组工程工具，特别是有效治疗血液病和实体瘤的高效CAR-T载体，仍然不发达。该研究发现了40个活跃的(整合能力强的)TEs，超过了之前发现的累计数量(20个)。通过这些统一的比较数据，发现Tc1/mariner超家族表现出较高的活性，这可能解释了它们普遍存在的水平转移。TEs的进一步功能表征显示了插入偏倚等特征的额外差异。值得注意的是，在血液和实体肿瘤的CAR-T治疗中，TE鉴定的最活跃的DNA Mariner2\_AG(MAG)在很大程度上优于两种广泛使用的载体，慢病毒载体和基于TE的载体SB100X。总的来说，该研究强调了DNA TE的不同转位特征和进化动力学，并增加了TE工具箱的多样性。

来源: Cell

发布日期:2024-06-05

全文链接:

<http://agri.nais.net.cn/file1/M00/03/6F/Csgk0WZhutWAJ-kAAJQ8L9AKpM8727.pdf>

## 2. 山东农大构建了玉米叶枕的单细胞转录组图谱

简介: 2024年6月3日, 山东农业大学生命科学学院李刚教授团队与北京大学现代农业研究院李博生研究员团队在Molecular Plant发表题为: Histological and single-nucleus transcriptome analyses reveal the specialized functions of ligular sclerenchyma cells and key regulators of leaf angle in maize的研究论文。本研究通过组织学和单核转录组分析, 揭示了玉米叶枕(LA)调节中特殊功能和关键调控因子。研究发现LA形成受两个步骤的调控过程显著影响: 首先是细胞伸长, 然后是随后的厚壁木素细胞(SCs)的木质化。通过大量和单核RNA测序, 研究者们生成了详尽的转录组图谱, 并鉴定了在低表皮细胞中富集的基因, 这些基因可能影响它们分化为SCs。此外, 功能特征化了两个编码非典型基本螺旋-环-螺旋(bHLH)转录因子的基因, bHLH30及其同源基因bHLH155, 它们在伸长的腹侧细胞中高度表达, 并正向调控LA扩展。这一研究不仅加深了人们对玉米叶夹角建成和株型调控的理解, 还为耐密理想株型的遗传改良提供了潜在的靶点, 具有重要的理论价值和实践意义。

来源: Molecular Plant

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

发布日期:2024-06-03

全文链接:

<http://agri.nais.net.cn/file1/M00/03/6F/Csgk0WZhviOAI1OhAEBVahMoQoo593.pdf>

### 3. 中国农科院畜牧所基于机器学习模型预测苜蓿群体适应性

简介: 2024年6月3日, 中国农科院杨青川/周永锋团队在Molecular Plant发表题为:

“Evolutionary genomics of climatic adaptation and resilience to climate change in alfalfa” 研究发现了苜蓿属种内与种间基因交流引入有助于适应当前和未来气候的基因, 共鉴定出1671个环境适应性相关的候选基因, 基于机器学习模型, 预测出不同地域苜蓿群体应对气候变化的适应性。该工作鉴定了多个苜蓿属物种气候适应的重要基因, 为苜蓿全基因组设计育种奠定基础, 在全球气候变化的大背景下, 对培育具有气候变化适应性的优异苜蓿新品种具有重要意义。

来源: Molecular Plant

发布日期:2024-06-03

全文链接:

[http://agri.nais.net.cn/file1/M00/10/41/Csgk0GZhuzKAeo6KADrfY\\_SquQU846.pdf](http://agri.nais.net.cn/file1/M00/10/41/Csgk0GZhuzKAeo6KADrfY_SquQU846.pdf)

### 4. 北大-清华合作开发蛋白质“扩增测序”新方法

简介: 2024年5月30日, 北京大学化学与分子工程学院、北大-清华生命联合中心陈鹏课题组与王初课题组合作, 在National Science Review杂志上发表了题为“Amplifiable protein identification via residue-resolved barcoding and composition code counting”的论文。在这项工作中, 作者结合氨基酸特异的化学偶联反应、核酸定量扩增技术和计算机辅助的蛋白质指纹识别算法, 开发了一种蛋白质“扩增测序”的新方法 (AmproCode), 为开发新一代的蛋白质测序方法提供了全新的思路。蛋白质是生命功能的基石, 单分子、单细胞水平的蛋白质测序技术能够为科学研究带来突破性的进展, 并推动人类疾病诊断和治疗的进步。尽管基于质谱的蛋白质组学技术能够快速、高通量地鉴定生物样品中的蛋白质, 但是与更为成熟的核酸测序技术相比, 尚不能借助类似聚合酶链式反应 (PCR) 的技术直接扩增蛋白质分子, 低丰度蛋白质的测序依然存在巨大的挑战。近些年, 为了解决这一问题, 研究者们努力将各类高灵敏的分析技术应用到蛋白质测序领域, 例如单分子光学技术、单分子力学技术、DNA纳米技术、纳米孔技术等等。这些单分子蛋白质测序的尝试大多依赖于高灵敏的单分子分析设备。而本工作中, 作者则借助氨基酸侧链特异的偶联反应, 将DNA“条码”标记到目标氨基酸的侧链上, 结合qPCR定量放大DNA信号, 实现蛋白质中特定氨基酸的“编码”与“扩增”。同时, 作者还为之“量身定做”了蛋白质解析的计算机程序, 以分析qPCR定量的氨基酸比例信号, 完成对数据库中蛋白质的指认。这个全新的高灵敏蛋白质“扩增测序”技术被命名为AmproCode。作者首先设计了AmproCode技术的蛋白质解析程序。作者用一个个矩阵记录蛋白质组中每条蛋白质的选定氨基酸的比例信息, 并计算它们与输入信号矩阵的余弦距离, 以评价蛋白质是否被成功鉴定。作者尝试利用计算机模拟评估在此算法下AmproCode技术对蛋白质组的覆盖程度。计算模拟结果表明, 四到五类的氨基酸的比例信息便足以成功覆盖超过90%的人类全蛋白质组和分泌蛋白质组。这为AmproCode技术提供了坚实的理论依据。随后, 作者综合考虑反应速率、效率、特异性等因素, 为AmproCode技术选择了五类氨基酸特异的偶联反应, 包括半胱氨酸和马来酰亚胺的反应、赖氨酸和活化酯的反应、甲

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>



硫氨酸和氮杂环氧丙烷的反应、天冬氨酸/谷氨酸和胺基的反应、以及酪氨酸和重氮苯甲醛的反应。作者在小分子和多肽上分别测试了上述反应，并且设计了对应的DNA标记策略，完成了多肽上氨基酸侧链特异的DNA标记。其中前四类反应效率较高，作者优先选用它们进行后续的实验，而酪氨酸的反应效率和特异性偏低，作为备选。作者在两条合成的分泌组多肽ELA和URP上展示了AmproCode技术的应用潜力。作者对多肽分别进行了半胱氨酸、赖氨酸、甲硫氨酸、天冬氨酸/谷氨酸特异的DNA标记和qPCR定量，获得了它们的比例信息，并借助蛋白质解析程序成功将ELA和URP肽从分泌组数据库中分别指认。AmproCode能够在低至飞摩尔每升的浓度（fmol/L）指认数据库中的目标多肽，比质谱和ELISA灵敏10到10000倍，展示了其超灵敏的蛋白质鉴定能力。如果借助合适的分离与富集技术，AmproCode技术也能用于混合样品中目标蛋白质的识别。作者借助SrtA  $\beta$  酶从混合物中分离A  $\beta$  肽，扩增并定量了其半胱氨酸、赖氨酸、甲硫氨酸和酪氨酸的比例，并在分泌组数据库中完成了A  $\beta$  肽的鉴定，展示了AmproCode技术在生物标志物的检测和发现方面的潜在应用价值。最后，作者再次借助计算机模拟，评估了不同实验参数对AmproCode性能的影响。模拟结果表明，氨基酸定量的误差比较容易影响蛋白质识别的成功率，但是标记定量更多的氨基酸则能够显著削弱实验误差的影响，提高AmproCode技术的鉴定成功率。因此，在未来利用更多氨基酸特异的化学反应标记并定量更多种类的氨基酸，能够显著提升AmproCode技术的准确性、对蛋白质组覆盖率和应用普适性。综上，本工作开发了一种蛋白质“扩增测序”新技术——AmproCode。作为一种将蛋白质转化为可扩增分子的突破性技术，它为开发下一代蛋白质测序技术提供了全新的思路，在未来可能推动单细胞蛋白质组学以及临床标志物鉴定的技术突破。作者也认为这项工作能够吸引更多化学家关注蛋白质科学，开发更多氨基酸特异的化学反应，为蛋白质测序以及蛋白质化学生物学的提供更多优秀的新工具。

**来源:** National Science Review

**发布日期:**2024-05-30

**全文链接:**

<http://agri.nais.net.cn/file1/M00/03/6F/Csgk0WZhvieAZ8rQADla1WkLn1Q287.pdf>