北京基因组所等开发出叶绿体基因组综合数据库

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/20136.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！

叶绿体是植物将光能转化为化学能的重要细胞器，具有独立的基因组。自植物叶绿体基因组被发现以来，被广泛应用于植物系统进化关系研究、光合作用调控机制研究、叶绿体基因工程等方面。随着基因测序技术的发展，尽管已发布了海量的植物叶绿体基因组序列，但如何整合应用这些数据目前仍面临数据命名标准不统一、数据信息不全以及较高经济价值的物种尚未进行测序等问题。

近日，中国科学院北京基因组研究所（国家生物信息中心）国家基因组科学数据中心章张、宋述慧团队，联合中国中医科学院中药资源中心袁媛、黄璐琦团队，开发了迄今为止物种数量最多的叶绿体基因组综合数据库Chloroplast Genome Information Resource（CGIR）。CGIR收录了来自11,946个物种的19,388条叶绿体基因组序列，包括利用全国第四次中药资源普查标本自测的718种未发表的叶绿体基因组序列，按照基因组（Genomes）、基因（Genes）、微卫星序列（SSRs）、DNA条形码（Barcodes）、DNA特征序列（DSSs）五个功能模块对数据进行组织与管理。相关研究成果以*Towards comprehensive integration and curation of chloroplast genomes*为题，发表在*Plant Biotechnology Journal*上。

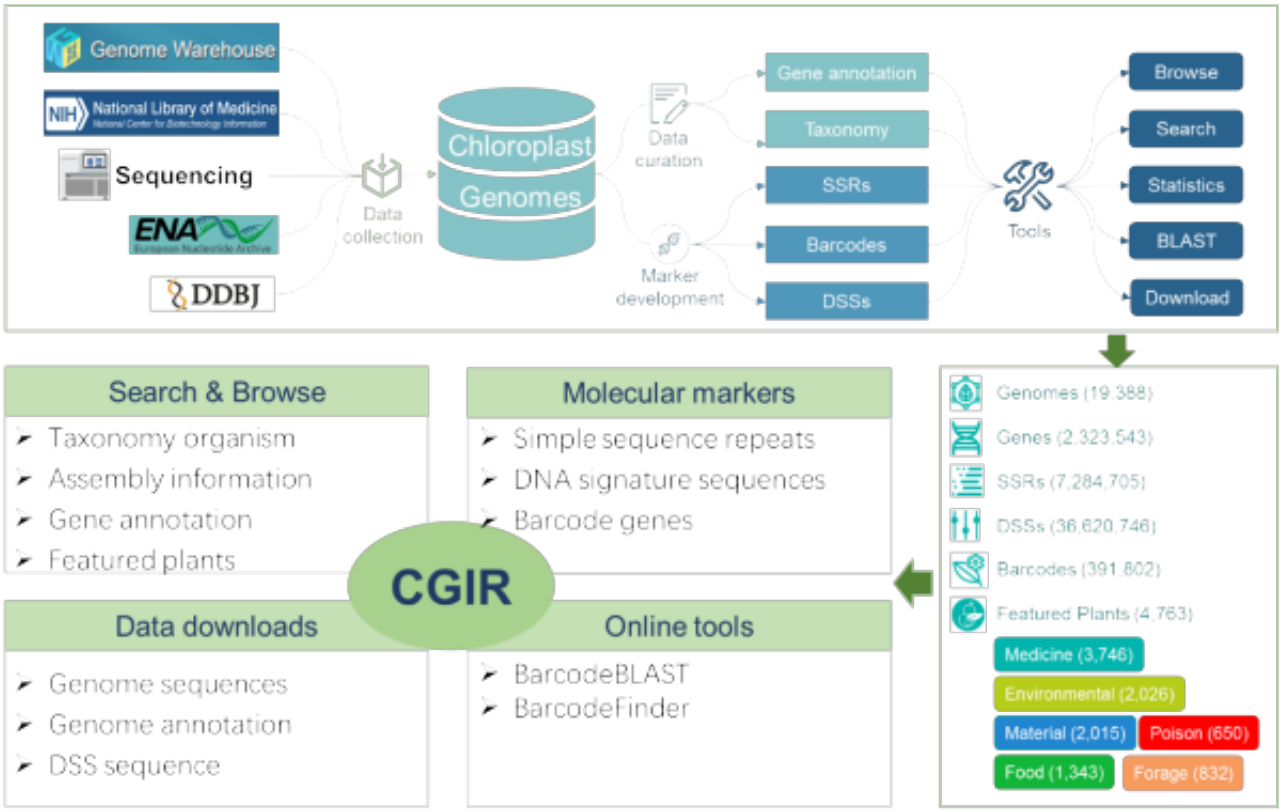
根据生物物种名录（The Catalogue of Life），经过大规模人工审编，CGIR对所收录叶绿体基因组的物种分类信息进行审编，按照纲、目、科、属、种不同分类层级进行整理，并依据权威植物研究机构邱园发布的世界功能植物名录（World Checklist of Useful Plant Species）对药用植物、食用植物、环境植物、能源植物、有毒植物、能源植物等进行标注。同时，CGIR审编修正基因名的不规范命名、异名、错误注释等情况。在此基础上，CGIR系统整理各基因组的基因注释信息，为用户检索、浏览和信息获取提供便利。

针对分子标记开发这一叶绿体基因组最为常见的应用情景，CGIR使用生物信息学方法计算了所收录叶绿体基因组的微卫星序列、DNA条形码和DNA特征序列三种不同类型分子标记信息，同时，开发了相应的树型视图方便用户根据分类层级信息快速寻找目标标记，简化了科研人员开发分子标记的流程。

CGIR通过自主测序、整合公开基因组资源和人工数据审编向用户提供了目前最全面、物种数量最多的叶绿体基因组数据。经审编的物种分类、物种功能、基因名称与序列、分子标记等保证了数据的高度可靠，对植物系统发育、物种鉴定、叶绿体基因工程的发展均具有重要意义。

研究工作得到科技基础资源调查专项、中国中医科学院科技创新工程项目、中央本级重大增减支项目“ 名贵中药资源可持续利用能力建设项目 ”的支持。

论文链接



CGIR数据处理示意图及主要功能模块的数据统计

研究团队单位：北京基因组研究所

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转发