时空三极环境大数据平台

**黑河流域典型荒漠植物耐旱机理的基因组学研究——数据集II（2014-2015）**

英文标题：Genomic studies of drought tolerance mechanisms of a typical plant in Heihe basin - dataset II (2014-2015)

1、摘要

一、数据概述
 此数据汇交是“黑河流域生态-水文过程集成研究”重大研究计划重点项目“黑河流域典型荒漠植物耐旱机理的基因组学研究”的第二次数据汇交。本项目的主要研究目标是以典型荒漠植物沙冬青为材料，利用目前国际上先进的新一代基因测序技术对沙冬青的全基因组序列及基因转录组序列进行解码，从而发掘与抗旱相关的基因和基因群组，并用转基因技术在模式植物（如拟南芥和水稻）中验证其抗旱性。
二、数据内容
1．沙冬青基因组和转录组的序列测定：
 前期基因组预测序测得蒙古沙冬青的基因组大小约为926 Mb，GC含量36.88%，重复序列比例66%，基因组杂合率0.56%，表明其基因组重复序列多，杂合度较高，属复杂基因组。 基于这一预测序结果，我们随后开展了沙冬青基因组的深度测序，所得数据经组装后得到937 Mb的全基因组序列（表一），与前期预测的基因组大小基本一致。通过对沙冬青的转录组测序和序列组装（表二），获得了77,000余个基因编码序列（Unigene），对这些基因序列进行注释发现，绝大部分基因序列与豆科植物大豆、鹰嘴豆和菜豆等有较高的相似度（图一），与沙冬青属豆科植物的事实相符。
2．沙冬青简单重复序列（SSR）分子标记的发掘：
 网络公共数据库已有公开发表的沙冬青转录组数据集，其样品采集地点是宁夏中卫市。而本项目组样品采集的地点是甘肃民勤县，为了研究这不同地区的沙冬青的序列是否具有序列多态性，我们首先鉴定了民勤县植物样品的基因组中的简单重复序列（SSR）分子标记（表三），随后与中卫市植物样品的转录组序列进行比较，发现部分SSR分子标记具有多态性（表四），这些分子标记可用于该物种植物的遗传图谱构建、QTL定位和遗传多样性分析等研究中。
三、数据处理说明
 样品采集地点：甘肃民勤县，经纬度：北纬N38°34′25.93″ 东经E103°08′36.77″。基因组测序：共构建8个不同大小的基因组DNA文库，使用Illumina HiSeq 2500仪器测定。转录组测序：共构建24个转录组mRNA的文库，使用Illumina HiSeq 4000仪器测定。
四、数据的使用说明和意义
 我们选定一种典型的荒漠植物作为研究对象，从基因组学的角度解析该荒漠植物的全基因组和转录组序列，发掘其中蕴藏的宝贵抗旱基因资源，并研究他们的抗旱机理，有利于沙冬青这一古老而重要植物资源的有效利用，以及黑河流域抗旱植物的遗传培育、生态恢复和可持续发展。

2、关键词

主题关键词：沙漠/荒漠,植被,基因组学,荒漠生态系统
学科关键词：陆地表层
地点关键词：黑河流域, 民勤, 阿拉善盟
时间关键词：2015

3、数据细节

1.比例尺：10000

2.投影：4326

3.文件大小：1.0MB

4.数据格式：EXCEL

4、空间范围

|  |  |  |
| --- | --- | --- |
| - | 北：38.79 | - |
| 西：102.5 | - | 东：103.57 |
| - | 南：37.87 | - |

5、时间范围2014-01-12 17:00:00+00:00--2016-01-11 17:00:00+00:00

6、引用方式

数据的引用:

何军贤. 黑河流域典型荒漠植物耐旱机理的基因组学研究——数据集II（2014-2015）. 时空三极环境大数据平台, DOI:10.3972/heihe.0034.2016.db, CSTR:18406.11.heihe.0034.2016.db, 2016.[HE Junxian. Genomic studies of drought tolerance mechanisms of a typical plant in Heihe basin - dataset II (2014-2015). A Big Earth Data Platform for Three Poles, DOI:10.3972/heihe.0034.2016.db, CSTR:18406.11.heihe.0034.2016.db, 2016]

文章的引用:

7、资助项目信息

8、数据资源提供者

姓名: 何军贤
单位: 香港中文大学深圳研究院
电子邮件: jxhe@cuhk.edu.hk