

王锦蛇 (*Elaphe carinata*) 的基因组组装和注释*

樊佳乐¹ 兰天明² 邹红菲³

(1.东北林业大学 哈尔滨 150000; 2.东北林业大学 哈尔滨 150000; 3.东北林业大学 哈尔滨 150000)

摘要:【目的】王锦蛇 (*Elaphe carinata*) 是一种重要的经济养殖物种, 作为一种大型卵生无毒蛇类, 其对各种毒蛇的毒液可表现出拮抗作用, 包括血液循环毒素和神经毒素, 这种保护的具体免疫机制和消化途径尚不清楚。试验旨在获取王锦蛇的首个全基因组序列, 以填补现有基因组资源的空白, 为毒理学、地理种群研究和免疫遗传学研究提供材料。【方法】以采样自中国安徽省黄山市的成年王锦蛇为试验动物, 使用 MGIEasy stLFR 文库准备试剂盒构建 stLFR 相同分子标签 (co-barcode) DNA 文库, 使用 BGISEQ-500 测序仪进行测序, 使用基因组 DNA kit 分离 DNA 进行全基因组 (WGS) 测序, 使用单管长片段数据 (stLFR) 组装并结合 WGS 数据填补缺口及冗余去除, 利用生物信息学方法结合 RNA 测序数据对获得的全基因组序列进行基因功能注释。【结果】试验共获得 stLFR 长度测序数据量 209.11Gb, 测序深度 134X, WGS 二代数据 106.14Gb, 测序深度 68X, RNA 测序数据 11.38Gb。利用 stLFR 测序数据组装并结合 WGS 数据填补的基因组大小为 1.56GB, 基因组最大 Scaffold 长度为 49.75M, Scaffold N50 为 6.53M, contig N50 为 44.05Kb, GC 含量为 40.25%, BUSCO 评估为 83.20%, 预测基因组中重复元件含量占比 42.26%, 总长度为 621Mb, 在所有重复元件中, 长散在重复元件 (LINEs) 占 72.56%, 短散在重复元件占 1.33%, DNA 转座子占 4.78%, 未知类型占 0.90%。共预测了 22,065 个功能基因, 与 TrEMBL 数据库相关的注释占比最高, 达到 97.92%, 与 Swiss-Prot 数据库、KEGG 数据库、InterPro 数据库及 Gene Ontology 数据库共同注释比例达到 98.77%。此外, KEGG 注释显示在人类疾病、生物系统和代谢类型中基因数量占比较高; 基于 GO 数据库注释到的 25 个生物过程类别中, 有 251 个基因与免疫系统过程有关, 有 2 个基因与解毒相关。基于最大似然法构建的系统发育树表明, 本研究数据与玉米锦蛇亲缘关系最近, 系统发育关系合理。【结论】研究生成的王锦蛇基因组是蛇类现有基因组资源的有益补充, 为进一步研究提供了重要的基因组数据。试验数据可与其他蛇类基因组数据结合, 进行系统发育研究, 构建蛇类和其他爬行动物的发育进化史。此外, 这些数据对理解王锦蛇的免疫和消化机制具有重要意义, 特别是在研究蛇类进化和毒素产生方面。本研究成果不仅丰富了蛇类基因组数据库, 也为相关领域的科学研究提供了新的工具和视角。

关键词: 基因组; 动物遗传; 王锦蛇