



北京基因组所（国家生物信息中心）开发分子序列组分动态图谱数据库CompoDynamics

作者： 发布时间：2021-12-29 | 【大】 【中】 【小】 | 【打印】 【关闭】



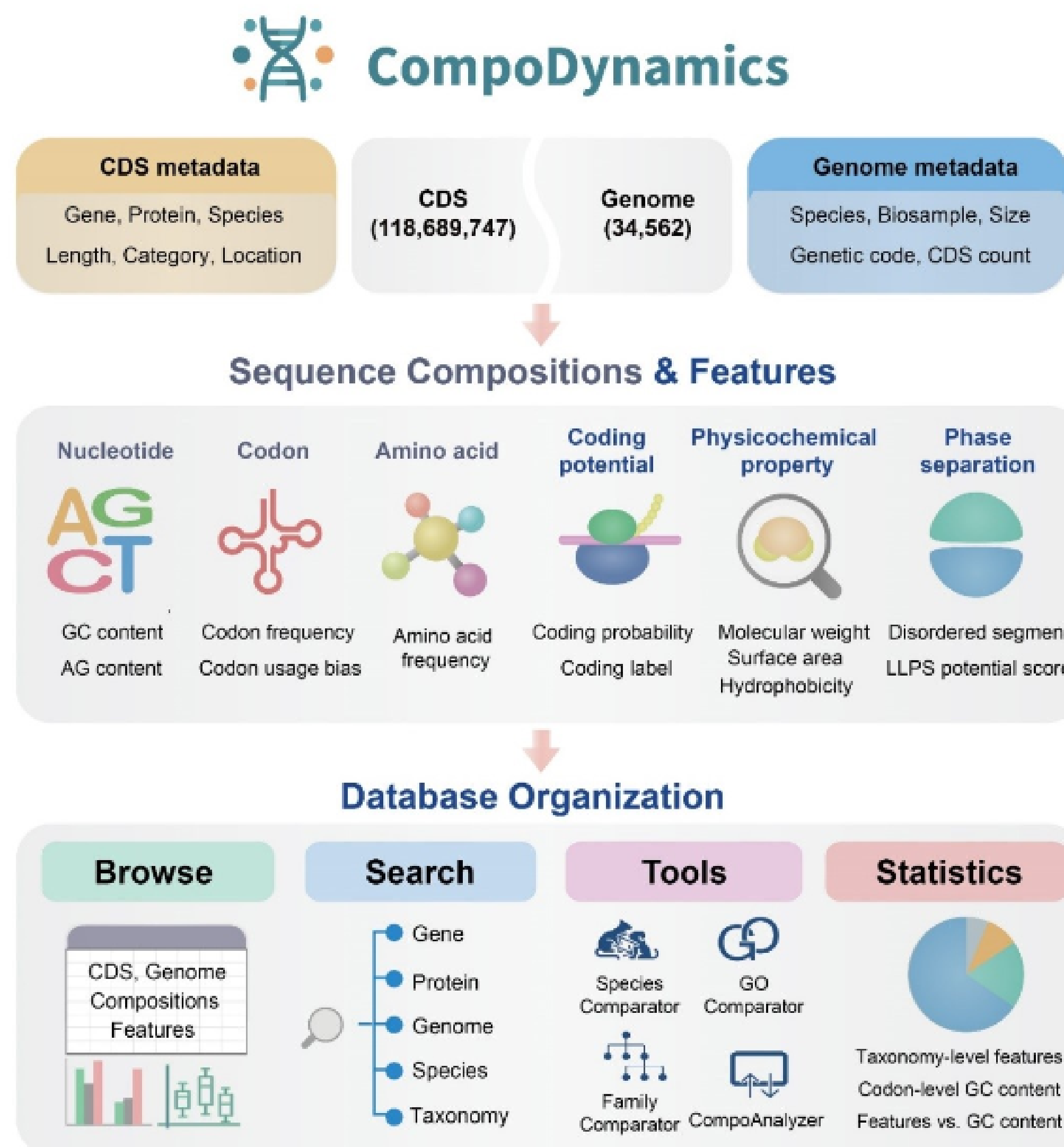
生物体的核酸与蛋白质分子的序列组分及相关特征（如GC/AG含量、密码子使用偏好、蛋白质物理化学性质等）对于研究基因功能和物种演化具有重要意义，是理解不同物种、不同基因家族以及不同功能基因之间差异的数据基础。

近日，由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开发的分子序列组分动态图谱数据库CompoDynamics正式上线，旨在对序列组分的动态变化提供综合、全面的展示，以多物种的系统性比较分析为基础，为深入的分子演化研究提供重要的参考和启示。该研究成果以“CompoDynamics: a comprehensive database for characterizing sequence composition dynamics”为题在国际学术期刊*Nucleic Acids Research* 在线发表。

CompoDynamics针对RefSeq数据库的基因组注释信息，分别计算分析了基因和基因组层面的3类序列组分特征（碱基组成、密码子使用偏好、氨基酸组成）和3类相关序列特征（编码潜能、蛋白物理化学性质、相分离特性）。目前，CompoDynamics共包含24,995个物种、34,562个基因组、1,692,647个基因，以及118,689,747条开放读码框序列。每条序列或每个基因组均有专门页面对各项特征进行详尽展示，并在主页以组分/特征分类展示。用户可通过物种分类、物种名、装配号、基因序列号、蛋白名称等进行检索。此外，CompoDynamics还提供了SpeciesComparator、FamilyComparator、GOComparator和CompoAnalyzer 4个在线比较分析工具，分别用于物种间、基因家族间、基因功能层面的各项组分/特征比较分析，支持对用户输入序列的计算分析，促进多组分特征和多维度的分子演化研究。

北京基因组所（国家生物信息中心）助理研究员降帅、硕士研究生杜强和冯祖瑞为本文共同第一作者，马利娜副研究员与章张研究员为共同通讯作者。该研究得到了科技部、中科院战略性先导科技专项、国家自然科学基金委、中科院青促会项目资助。

[论文链接](#)



CompoDynamics数据库内容与结构

