

Why sequence all eukaryotes?

为什么要测序所有真核生物？

朱玉锋 董马丹妮 蒋运 张子栋

华中农业大学
信息学院

2023 年 5 月 5 日



华中农业大学
HUAZHONG AGRICULTURAL UNIVERSITY

分工

- 董马丹妮
 -
- 朱玉锋
 -
- 蒋运
 -
- 张子栋
 -



大纲

- 
- ① 发现生命之树
 - ② 定义真核细胞的起源
 - ③ 追踪共生的基因组变化
 - ④ 解密染色体进化
 - ⑤ 揭示真核基因调控的深层逻辑
 - ⑥ 探索性系统的多样性
 - ⑦ 探索物种基因组学中的多样性
 - ⑧ 探索物种基因组学中的多样性

引言

● 作者

- 作者来自不同国家和机构的多位科学家，均为地球生物基因组计划（Earth BioGenome Project）的成员。该计划的目标是对已命名的真核生物（约 200 万种）的基因组进行高质量的测序，以尽力构建一个地球生命的数字图书馆。

● 背景

- 文章探讨了为什么要测序所有真核生物，而不仅仅是一些代表性的物种。

● 观点

- 文章认为，只有当拥有了完整的生命树或者自然生态系统中所有物种的全基因组数据时，才能回答一些关于进化和生态学的重要问题。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

发现生命之树

需要准确描述生命树的原因

- 正确的系统发育树是理解表型起源及其多样化的重要数据。
- 新植物物种的杂交起源很常见，杂交、多倍体化和水平基因转移也在其他分支的多样化中发挥了重要作用。
- 即使没有杂交，由于谱系排序不完整和遗传漂移的随机性，整个基因组的基因树通常与总体物种树不同。

发现生命之树

系统发育树构建的难题

- 国际核苷酸数据库中物种序列少，物种之间同源基因重叠低。
- 基因组序列已被用于探索真核生物的起源，并用于修正原生生物、植物和动物的深层分裂，但大多数处于草稿形式的基因组仅适用于所有物种的 0.4%。特别是，这棵树的根和许多单细胞真核生物深层谱系（多系「原生生物」）的位置仍未解决。
- 可以解决的问题（关于真核生物进化的深层问题）
 - ① 杂交和多倍体化在生命树中的普遍性、重要性；
在基因组中留下了哪些古老的杂交、多倍体化和渐渗特征？
 - ② 新基因是何时、何地以及如何产生的？
不同谱系中是否存在可预测的基因家族起源和多样化模式？
 - ③ 水平基因转移在真核生物进化中有多重要？

发现生命之树

完整基因组序列的作用及意义

- ① 为「真核生物树所有分支的起源和多样化的产生」这一假设提供充分支持，绘制网状事件和基因复制、分化和丢失的时间。
- ② 完整的序列更可能包含与感兴趣的节点、过程或事件有关的信息。
- ③ 为不同问题选择合适的数据，例如选择具有与被评估节点的系统发育深度相匹配分歧率的子集。
- ④ 同源基因家族、其蛋白质序列和非编码调控元件的完整清单将突出功能 DNA、RNA 和氨基酸残基的差异和限制，并为生物工程提供新功能的信息。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

定义真核细胞的起源

需要解决的问题

- ① 对于真核细胞复杂性的前因以及所涉及的环境和选择性因素知之甚少。
- ② 目前远不清楚参与以下过程中的单个蛋白同源物如何、何时、以何种顺序进化：
 - 细胞核
 - 时空分离的转录和翻译
 - 线粒体
 - 线性染色体
 - 减数分裂
 - 有丝分裂
 - 内膜和蛋白质运输系统以及鞭毛

定义真核细胞的起源

完整真核基因组可以解决的问题

- 所有现存真核生物的祖先中都存在哪些基因？
- 真核细胞的复杂性是以逐步的方式进化还是一次进化？
- 真核细胞的复杂性是以逐步的方式进化还是一次进化？
- 哪些基因组过程限制和促进了次生和次生内共生生物光合作用的水平传播？

猜测：对真核细胞起源的变革性见解很可能来自古菌和微生物真核结构域内目前未知的细胞多样性序列。高分辨率真核生物生命树的生成将为核心真核生物工具包和器官功能中基因的起源和多样性分析奠定基础。

总结

一套完整的真核基因组可以解决关于真核细胞进化起源的基本问题。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

追踪共生的基因组变化

共生的互惠关系

需要解决的问题

共生对真核细胞生态多样性和生物复杂性的影响。

● 共生的互惠关系

- ① 土地上的初级生产力是基于维管植物和定根微生物之间的互惠共生关系。昆虫营养生态位是通过获得有益的微生物共生体来定义的。
- ② 相互作用的合作伙伴往往将基本过程的责任委托给彼此，即使面对免疫和反编辑系统施加的障碍，也会强制执行相互依赖。
- ③ 物种存在于有助于食物消化和其他过程的微生物组中，微生物组的基因组可能与理解宿主生理学有关。

追踪共生的基因组变化

所有物种的基因组可以解决关于共生体的重大问题

- 宿主及其微生物组是否应该被视为全基因组超有机体？
- 寄生共生体和互惠共生体基因组的进化有共同的主题吗？
- 在基因和分子水平上，实现共生的机制的范围和多样性是什么？
- 有益但非强制性的共生体如何影响双方的进化？
- 共生基因转移有多普遍？共生体消失的遗产对当今的生物功能有哪些贡献？
- 大型宿主生物及其微生物群之间的共同进化在多大程度上推动了多样化？

追踪共生的基因组变化

作用

- ① 获取所有真核生物物种共生伙伴的基因组将有助于发现新的基因和调节机制，这些基因和机制是不同生物整合、开辟新的生态位和促进物种多样化的基础。
- ② 了解宿主及其较小伙伴的相对进化轨迹，将有助于了解这些生物之间的必要联系，并可能导致有利于社会或环境的共生体的定向增强。
- ③ 了解害虫和寄生虫及其互惠共生体的基本生理依赖性也可能导致新的干预策略。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化**
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

解密染色体进化

解密染色体进化的原因

- 大多数细菌和古菌都有单一的圆形基因组，并且是单倍体，但大多数真核生物的基因组是二倍体的，并以线性染色体组织。
- 真核染色质折叠成三维构象，这对基因表达和细胞分化至关重要。
- 真核生物通过添加各种非基因组模板 DNA，保护其染色体末端端粒免受不可避免的复制相关侵蚀。
- 基于特定的序列特征或表观遗传信号，着丝粒功能确保染色体组正确分离到子细胞。基因在不同染色体上的分配也为基因调控网络的进化提供了机会和约束。
- 在许多真核生物类群中，核型和同源性通常是保守的，而在其他类群中，即使在短的进化时间内，重排也很猖獗。

解密染色体进化

真核生物基因组的完整采样可以解决关于染色体的重要问题

- 在真核生物树的所有祖先节点上，基因组的可能核型、3D 组织和同源关系是什么？
- 在一个新物种的形成过程中，染色体数量是如何稳定的？
- 有丝分裂和减数分裂对染色体数量和组织有哪些限制？
- 染色体重排是真核生物进化中主要转变和适应的原因吗？
- 端粒是如何进化的，不同端粒维持机制的意义是什么？
- 着丝粒的种类（例如，全着丝粒与着丝粒）是否影响基因组进化？
- 核型对基因调控网络的进化有哪些限制？

解密染色体进化

作用

- ① 通过在真核生物生命树上生成染色体规模的基因组组装，可以明确地解决所有规模的染色体组织进化问题，从局部同源性的保守性分析到核型进化，以及着丝粒和端粒重新定位的动力学。
- ② 这些数据可以支持开发具有工程染色体稳定性和其他行为的新型合成基因组，用于生物医学和工业应用。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑**
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

揭示真核基因调控的深层逻辑

揭示真核基因调控深层逻辑的必然性

- ① 每个真核基因组都包含数千个蛋白质编码基因，每个基因都有调控元件和电路。这些元素中的大多数（非编码转录物（短和长）、启动子、增强子、绝缘体和基因组本身的 3D 结构）可以与非功能 DNA 区分开来，因为它们的序列在进化上受到限制。
- ② 特定谱系变化速度加快的区域可能与新的进化适应有关。对于单个物种，参考基因组可以测定组织和细胞中的调节。
- ③ 比较基因组学可以将在一个物种中实验确定的功能注释映射到相关物种，从而提供整个分支的快速、深入的基因组注释。

揭示真核基因调控的深层逻辑

可以解决的具体问题

- 所有真核基因的综合列表是什么？
- 真核生物基因组调控的组成部分是什么？它们的稳定性如何？
- 新的监管制度是如何产生并取代现有制度的？
- 深层多物种全基因组比对能揭示新的保守元素类别吗？
- 哪些基因组特征区分保守的调控网络？
- 调控网络是如何限制染色体和基因组进化的？

揭示真核基因调控的深层逻辑

所有物种的基因组的作用与意义

● 作用

- ① 可以以更高的灵敏度和更高的分辨率绘制功能序列，并将功能变异与表型联系起来。
- ② 可以绘制长程染色体相互作用图，并解开重复、移动元件插入和重复积累在修改先前存在的基因调控中的影响。

● 意义

- ① 将有可能绘制蛋白质编码基因和长非编码 RNA 的调控启动子区、重复、重复和转座元件插入，是基因调控创新的关键来源。
- ② 将每个物种单独考虑，可以更深入地了解调控基因组及其功能。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性**
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

探索性系统的多样性

性系统的多样性

● 性系统的多样性

- ① 一些原生生物群体（纤毛虫）有多种等效的交配类型，但大多数多细胞生物有两性。
- ② 分化单倍体配子（大卵子和精子）的产生多次独立进化，父母对卵子或精子的贡献导致了形态、生活史策略和行为上极端性别差异的进化。
- ③ 决定性别的机制多种多样，从单基因座驱动因素和分化的性染色体到表观遗传和环境决定。

● 真核谱系完整测序可以解决的问题

- 雄性-雌性系统多久独立进化一次？具有多种交配类型的系统如何进化和维持？
- 性别决定系统的变化会驱动物种形成吗？如何在基因组中反映？
- 在具有分离种系的物种（纤毛虫和大多数动物）中，种系如何指定？种系基因组如何维持？
- 染色质减少如何在生命树中分布及调节的？

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性**
- 8 探索物种基因组学中的多样性

探索物种基因组学中的多样性

● 争议

- ① 物种形成的过程。
- ② 许多情况下物种边界是泄漏或不完整的。
 - 生物物种概念认为姐妹分类群之间存在稳定的繁殖隔离。然而，基因组数据表明，在分类学上被接受的物种形成后，通过杂交和渗入进行的基因交换是常见的。

● 探索物种基因组学多样性的原因

- 命名物种的概念隐藏了多个神秘的分类群，许多物种尚未被发现和命名。基因组数据对发现和定义新的分类群至关重要。

- 1 发现生命之树
- 2 定义真核细胞的起源
- 3 追踪共生的基因组变化
- 4 解密染色体进化
- 5 揭示真核基因调控的深层逻辑
- 6 探索性系统的多样性
- 7 探索物种基因组学中的多样性
- 8 探索物种基因组学中的多样性

探索物种基因组学中的多样性

真核生物谱系完整测序可解决物种形成问题

- 物种形成过程的多样性以及它们是如何形成基因组的？
- 哪些基因组信号区分不完全谱系分类、渗入和杂交？
- 某些类型的基因和基因网络更有可能与物种形成有关吗？
- 染色体重排在物种形成中扮演什么角色？
- 不同的繁殖策略是否会驱动不同的基因组结构，从而影响物种形成？
- 物种辐射等宏观进化现象是如何塑造基因组的？
- 从目前的基因组来看，未来的物种形成是可以预测的吗？

Q&A

谢谢!

请老师批评指正!