

文章编号: 1674-7054(2022)05-0457-07



# 白边侧足海天牛全基因组 Survey 分析

李歆毓<sup>1,3</sup>, 丁梦莹<sup>2</sup>, 冯尔辉<sup>1</sup>, 张源源<sup>4,5</sup>, 章可兰<sup>3</sup>, 万迎朗<sup>3</sup>

(1. 海南东寨港国家级自然保护区管理局, 海口, 570100; 2. 海南大学海洋学院, 海口, 570228;  
3. 海南大学热带作物学院, 海口, 570228; 4. 中国热带农业科学院橡胶研究所, 海口, 571101;  
5. 农业农村部橡胶树生物学与遗传资源利用重点实验室, 海口, 571101)

**摘要:** 为揭示白边侧足海天牛基因组的信息, 以便为后续构建全基因组精细图谱提供基础, 利用流式细胞术与高通量测序技术对白边侧足海天牛全基因组进行了 Survey 分析。结果表明, 流式细胞术预测白边侧足海天牛基因组是 84K 杨的 1.69 倍, 约为 794.6 Mb。白边侧足海天牛高通量测序得到 25 Gb 的过滤数据, 总测序深度约为 30 x, 根据 17-mer 分析, 预测海天牛基因组大小约为 724.8 Mb, 基因组重复率为 52.8%, 杂合度为 1.55%, 模型拟合值为 99.38%; 根据 19-mer 分析, 预测海天牛基因组大小约为 730.8 Mb, 基因组重复率为 35.1%, 杂合度为 1.68%, 模型拟合值为 99.72%。最后, 对白边侧足海天牛尝试初步组装后得到总长为 628 Mb 的基因组。研究结果显示, 白边侧足海天牛是一个高度杂合的物种, 且其基因组大小超过 700 Mb。

**关键词:** 白边侧足海天牛; 基因组 Survey 分析; 基因组大小; 杂合度

**中图分类号:** Q 789; Q 959.1 **文献标志码:** A

**引用格式:** 李歆毓, 丁梦莹, 冯尔辉, 等. 白边侧足海天牛全基因组 Survey 分析 [J]. 热带生物学报, 2022, 13(5): 457-463. DOI: [10.15886/j.cnki.rdsxb.2022.05.005](https://doi.org/10.15886/j.cnki.rdsxb.2022.05.005)

白边侧足海天牛 (*Elysia leucolegnote*) 属于软体动物门囊舌目 (Sacoglossa) 海天牛超科 (Elysioidea) 海天牛属 (*Elysia*), 该种分布于中国的香港<sup>[1]</sup>、海口<sup>[2]</sup> 以及泰国<sup>[3]</sup>、菲律宾<sup>[4]</sup> 等地红树林。目前, 在我国已报道有 8 种海天牛的分布记录<sup>[1,5]</sup>。海天牛属部分种具有保留藻类叶绿体并使其发挥光合作用的能力<sup>[6-7]</sup>。这种被吸收后且继续发挥功能的叶绿体被称为盗质体<sup>[8]</sup>。不同物种的盗质体寿命的决定因素和盗质体发挥功能的分子机制等相关的研究<sup>[9-13]</sup> 对叶绿体的内共生理论的揭示<sup>[14]</sup> 以及指导植物抗逆方案的设计都具有重要的意义<sup>[15-17]</sup>, 但受限于光合软体动物材料的稀缺性, 此相关研究的规模十分有限。白边侧足海天牛在我国的分布有利于科研人员就地取材, 为我国在此方向的深入研究提供了独特的材料。分子机制的揭示离不开物种基因组的信息。海天牛属目前尚未有染色体水平上的高质量基因组的

报道。基因组已经成为深入研究分子机制的关键内容。基因组含量在同一个物种里是保持稳定的<sup>[18-20]</sup>。高通量测序为一种广泛用于多种动植物基因组测序的技术<sup>[21]</sup>, 将高通量获得的原始数据通过质控以后, 利用 K-mer 分析法评估基因组大小、杂合度和重复率等特征<sup>[22-23]</sup>。基因组测序大大促进了动植物的遗传进化及功能基因研究, 但深度测序之前的低覆盖度的全基因组调查尤为重要, 因为其可以决定基因组测序中的最合适的测序、拼装方式<sup>[24]</sup>。本实验旨在采用高通量测序技术结合 K-mer 分析<sup>[25]</sup>, 对白边侧足海天牛基因组进行测定及评估, 为后续深度测序提供基础信息。

## 1 材料与方法

**1.1 材料** 供试材料白边侧足海天牛(以下简称海天牛)收集于海南省海口市东寨港红树林保护区(110°38'26" E, 19°56'31"N), 并于海南大学热

收稿日期: 2021-11-04 修回日期: 2022-04-12

基金项目: 海南省自然科学基金(No.2019RC155)

第一作者: 李歆毓(1994-), 女, 海南大学热带作物学院 2019 级博士研究生. E-mail: [xinyul@163.com](mailto:xinyul@163.com)

通信作者: 万迎朗(1976-), 男, 教授. 研究方向: 热带作物遗传育种. E-mail: [ylwan@hainanu.edu.cn](mailto:ylwan@hainanu.edu.cn)

带作物学院饲养,84K 杨树由北京林业大学细胞生物学实验室王鑫伟提供。

**1.2 流式细胞分析** 取海天牛和 84K 杨树幼嫩新鲜叶片 1 g,置于盛有 500  $\mu\text{L}$  细胞裂解液 Galbraith 中,并迅速将其切碎。细胞核通过 30  $\mu\text{m}$  滤膜后,与 1 mL 浓度为 10  $\mu\text{g}\cdot\text{mL}^{-1}$  的 PI(碘化丙啶)染色液混合,室温孵育 30 s 后上机检测。使用 CyFlow<sup>®</sup>Cube8(希森美康,日本)流式细胞仪对海天牛基因组大小进行评估,变异系数控制在 5% 以内。每个样品重复 3 次。

**1.3 文库构建及基因组测序** 粉碎合格的 DNA 样品为 350 bp 左右的目的片段,构建文库,经过末端修复、加 A、加接头、目标片段选择和 PCR 等步骤,用安捷伦 2100 和定量 PCR 的方法检测文库片段大小和文库定量,以确定文库是否符合测序标准,通过桥式 PCR 的方法将文库固定到测序芯片上;将这些两端的片段在 Illumina HiSeq Xten(Illumina, U.S)测序仪上进行双末端 (Paired-End)测序,获得全基因组测序数据,去除未成对匹配的读长(reads)、低质量读长、接头遭受污染以及过滤掉接头重复(duplication)的读长等对测序所产生的数据进行预处理。

**1.4 测序数据的质量控制** 比较原始数据与过滤后数据的质量控制数据后,使用 FastQC(v 0.11.9)软件对过滤后的数据展开质量控制,包括对数据量的概览,并统计了读长每个位置测序质量,总体 reads 测序质量趋势, A、T、G、C 碱基含量估计测序是否存在偏差,是否存在污染,数据处理时是否需要去冗余;从而实现了对前期数据处理时,尽量高标准,严格质量控制。

**1.5 K-mer 分析以及基因组特征估计** 通过 jellyfish-2(v10.7.7)软件对序列文件进行 K-mer 的计数和统计;随后,利用负二项式模型(Negative binomial model)对应的软件 GenomeScope(v1.0)对基因组大小及其杂合度的评估,并生成最终基因组评估结果。选用 K-mer 值 17、19 两种条件对评估结果进行比较。

**1.6 基因组初步组装** 使用 SOAPdenovo2 (v2.03)软件对过滤后的数据进行拼接,拼接出 Contigs 序列,然后组装基因组。SOAPdenovo2 的 K-mer 参数设置 29,其他参数选择默认值。再将 Contigs 序列构图形成 Scaffolds 序列,并利用不同插入片段估计出 Contigs 间的距离,用 N 碱基填起来。最后,再利用测序的双末端数据之间的配对关系 (Paired-End)以及短 Reads 数据对已组装的 Contigs 的覆盖信息,对 Contigs 间空隙("N")进行局部组装,补充 Contigs 信息,适当延长 Contigs 序列。有效数据与原始序列进行对比后获得碱基深度,在序列上以 5 kb 为窗口,无重复前进,从而得到 GC depth 点图,对组装后的基因组进行评估。

## 2 结果与分析

**2.1 流式细胞结果预测基因组大小** 基于流式细胞术分析海天牛基因组大小,当变异系数控制在 5% 以内时,以 84K 杨作为对照样品(图 1-A)信号峰清晰集中,84K 杨与海天牛的混合样品的信号峰独立分离且距离较近(图 1-B)。利用 84K 杨作为对照样品,根据混合样品 PI 荧光强度以及峰值的倍数关系,计算海天牛基因组是 84K 杨的 1.69 倍,84K 杨的核 DNA 相对含量为 1.129 20,基

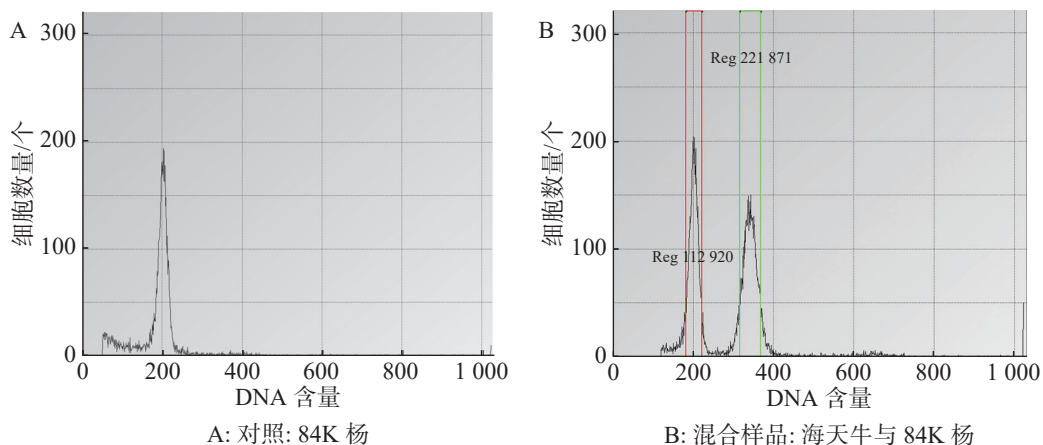


图 1 海天牛流式细胞分析

因组平均值为 470.155 Mb; 估算出海天牛的核 DNA 相对含量为 2.218 71, 基因组平均值为 794.562 Mb。

**2.2 建库信息及数据量统计** 基因组调查利用第二代高通量测序技术进行双末端测序, 获得全基因组的序列结果。测序共得到海天牛原始数据约 25.8 Gb, 共 171 847 064 条原始序列; 过滤后约 25.1 Gb, 原始数据 Q30 比例为 91.33%, 过滤后 Q30 比例为 91.78%, 满足基因组调查需要的测序数据量(表 1)。比较原始数据与过滤数据(表 2)的碱基的分布情况(图 2-A、2-B), 过滤前后除了测序时前几个 bp 碱基含量略有波动属正常现象, 其余每个测序位置 A 碱基和 T 碱比例相等, G 碱基和 C 碱基比例相等, N 碱基的数量为 0。原始数据

表 1 基因组序列数据量统计

数据类型	数值
原始序列数/条	171 847 064
原始碱基数/bp	25 777 059 600
过滤序列数/条	167 523 764
过滤序列所占比例/%	97.48
过滤碱基数/bp	25 128 564 600
原始数据Q30比值/%	91.33
过滤数据Q30比值/%	91.78

表 2 过滤数据的基本信息

参数	数值
总序列数/条	83 761 882
低质量序列数/条	0
序列长度/bp	150
G、C碱基含量/%	36

与过滤后数据的测序质量分布在 Q30 到 Q40 之间, Q30 序列占比高, 表明测序结果质量高可用于后续分析。

过滤数据所有读长上的碱基质量值大于 30 且波动小, 说明过滤后数据质量稳定(图 3-A); 实际 G、C 碱基含量与理论 G、C 碱基含量都在 38% 左右出现峰值, 且没有明显的偏差, 表明测序结果没有偏向性(图 3-B); 过滤后所有的数据读长都为 150 bp(图 3-C); 全部序列达到 Q20, 超过 95% 序列达到 Q30, 且集中在 Q36(表 2)。以上结果表明, 过滤后的数据读长长, 质量高, 没有碱基偏好性适用于 K-mer 分析。

### 2.3 K-mer 分析以及基因组大小、杂合率的估计

使用 K-mer 的分析方法可以预测白边侧足海天牛的基因组特征。选择 K-mer 的条件为 17 和 19 展开分析, 样本 17-mer 和 19-mer 分布曲线为非正常泊松分布, 呈现双峰分布, 在 17×和 27×附

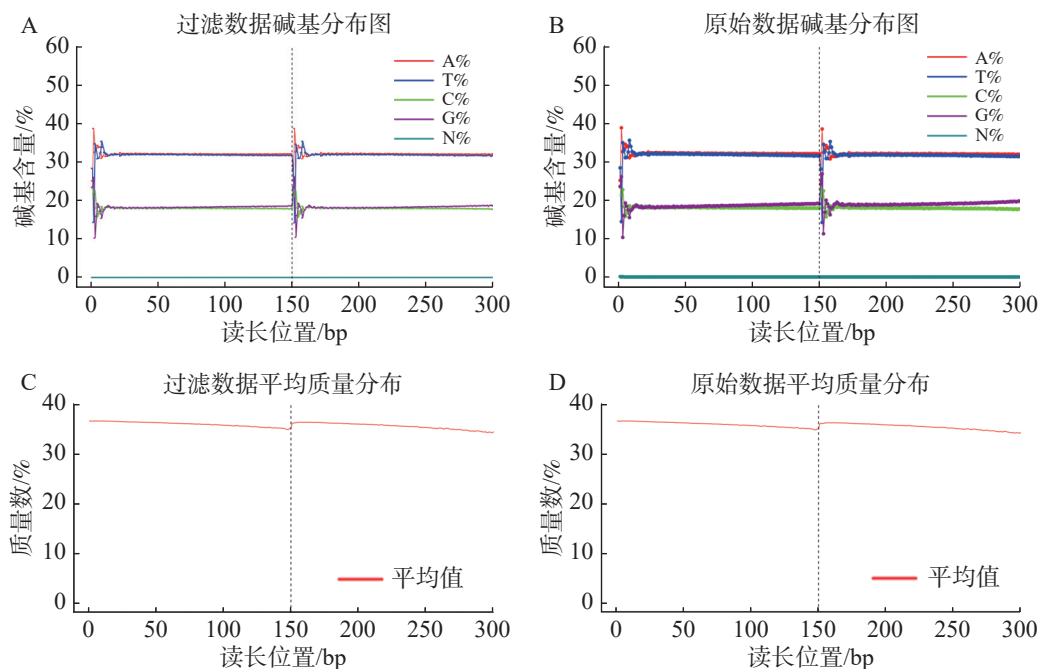


图 2 白边侧足海天牛过滤前后测序情况

A: 过滤后序列碱基分布图; B: 原始序列碱基分布图; C: 过滤后序列测序质量分布图; D: 原始序列测序质量分布图。

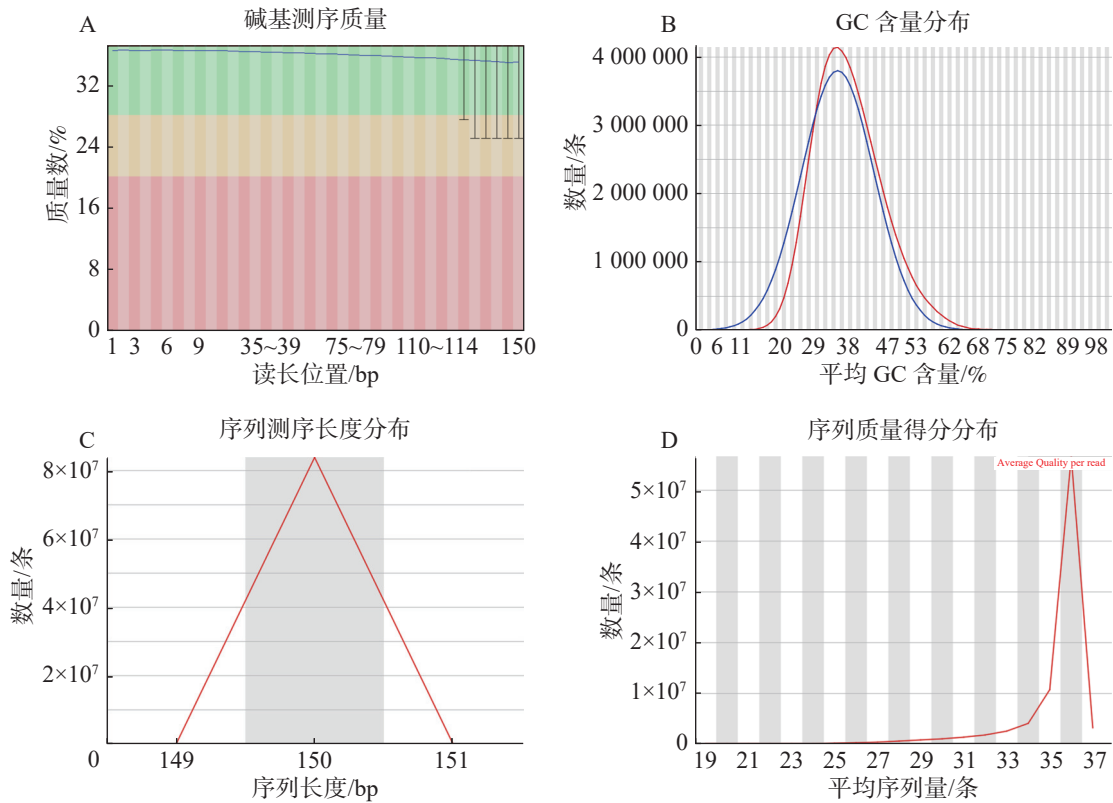


图3 白边侧足海天牛过滤数据情况图

A: 每个位置的碱基的测序质量; B: 序列平均 GC 含量分布图, 红线代表实际的 GC 分布情况, 蓝色代表理论的 GC 分布情况; C: 序列测序长度分布; D: 全序列的质量得分分布图。

近各有 1 个峰值(图 3-A、图 3-B)。总测序深度约为  $30\times$ , 根据 17-mer 分析, 预测海天牛基因组大小约为 724.8 Mb, 基因组重复率为 52.8%, 杂合度为 1.55%, 模型拟合值为 99.38%; 19-mer 分析预测海天牛基因组大小约为 730.8 Mb, 基因组重复率为 35.1%, 杂合度为 1.68%, 模型拟合值为 99.72% (表 3)。

表 3 白边侧足海天牛的 K-mer 数据统计

项目	17-mer	19-mer
深度	$30\times$	$30\times$
单倍体基因组长度/ bp	724,811,356	730,805,267
基因重复序列长度/ bp	382,653,068	256,787,169
基因特有序列长度/ bp	342,158,287	474,018,098
杂合度/ %	1.55104	1.67714
模型拟合/ %	99.3794	99.7184
读错误率/ %	0.148872	0.206886

**2.4 白边侧足海天牛基因组的预组装** 选用 SOAPdenovo2 软件对海天牛样本进行预组装, 设置 K-mer 参数为 29 时, 在 scaffold 尺度上, 得到

含 N 碱基的基因组大小 628 574 653 bp, 不含 N 的基因组大小 627 289 254 bp; Scaffold N50 长度为 373 bp, 共 405 072 条; Scaffold 数量 2 258 693 条, 最长的 scaffold 长度为 22 424 bp。在 contig 的尺度上, 以 contig N50 为 358 bp 数量有 419 361 条。得到含 N 碱基的基因组大小 624 854 764 bp, 不含 N 的基因组大小 624 854 764 bp, 最长的 contig 为 22 424 bp(表 4)。组装成 Scaffold 的 contig 的数量为 168 878 条, 每个 scaffold 的平均 contig 数目为 1.5。除此, 还得到 scaffold 尺度上的各碱基的含量, 碱基 A 数量为 207 638 986 bp, 占总的碱基

表 4 白边侧足海天牛预组装结果统计

类型	Scaffold		Contig	
	长度/ bp	数量/条	长度/ bp	数量/条
N50	373	405 072	358	419 361
N60	273	602 827	262	624 669
N70	204	870 539	195	903 051
N80	155	1 225 846	150	1 271 270
N90	122	1 684 983	120	1 739 030
最长片段	22 424	1	22 424	1
总计	628 574 653	2 258 693	624 854 764	2 313 267



数目的 33.03%; 碱基 C 数量为 110 706 106 bp, 占总的碱基数目的 17.61%; 碱基 G 数量为 109 134 885 bp, 占总的碱基数目的 17.36%; 碱基 T 与碱基 A 的数量及占比基本相同, 碱基 T 数量为 199 809 277 bp 占总的碱基数目的 31.79%; 剩下的所有为碱基 N, 数量为 1 285 399 bp, 占总的碱基数目的 0.20%。最后计算得到 G、C 碱基含量为 35.05%。GC-depth 分析显示, 测序无偏向性; 平均深度集中在 30×, GC 深度分布被分为 2 层。

### 3 讨论

目前, 在我国已报道有 8 种海天牛的分布记录<sup>[1, 5]</sup>, 其中部分海天牛具有利用藻类叶绿体进行光合作用的能力。盗质体寿命是不等的, 有的盗质体能维持长达 9 个月<sup>[26]</sup>, 而有的只能维持短短的几个小时。根据叶绿体在海天牛中停留的时间将海天牛分为三类, 第一类为长时间保存叶绿体物种(long-term retention (LtR) slugs), 已报道的包括 *E. chlorotica*, *E. timida*, *E. crispata*, *E. clarki*, *E. viridis*, *Plakobranchnus ocellatus* and *Costasiella ocellifera*<sup>[27-30]</sup>; 第二类是短时间保存叶绿体物种(short-term retention species, StR), 其对叶绿体的保留时间不超过两周; 第三类是不保存叶绿体物种(non-retention species, NR), 在食用藻类后迅速分解叶绿体<sup>[31]</sup>。在本研究中发现, 白边侧足海天牛至少能保持盗质体活性 2 个月以上, 属于能长时间保存叶绿体的海天牛。

对盗质体活性的长期保持依赖于宿主核基因编码的功能基因与盗质体基因的协调表达。例如 *E. chlorotica* 与 *E. timida* 食用藻类的叶绿体基因组中存在一种特定基因(*ftsH*, 一种对光系统 II 修复至关重要的 D1 质控蛋白酶)其中 M41 金属蛋白酶结构域是维持盗质体长期活动的关键<sup>[32-33]</sup>。同时动物内源的脂肪酸合酶-(FAS)样聚酮合酶(PKS)蛋白也可以提供光保护能力, 盗质体固定二氧化碳, 固定碳被转化为甲基丙二酰辅酶 a, 并被软体动物 EcPKS1 酶修饰, 合成紫外线-氧化阻断吡喃, 保护软体动物及其叶绿体免受光合损伤<sup>[34]</sup>。另外, 有观点认为吞食叶绿体后从植物中摄取的酶的丰度可能限制盗质体发挥功能。也有一种观念认为, 盗质体寿命的维持是通过从藻类细胞核到动物细胞核的广泛水平基因转移(HGT)来实

现。但是对于该假说还存在很大的争议, 早期研究中, TORRES 等证实了核编码的基因在质体核糖体抑制剂存在的条件下可以合成 *LHC I*, 并提出可以通过病毒或逆转录病毒实现 HTG 的假说<sup>[35]</sup>。这些假说的讨论, 都必须基于对海天牛核基因组及其摄取的盗质体基因组的分析研究。

*E. chlorotica* 的全基因组测序与组装是以二代为主, 三代 PacBio 辅助的 scaffold 的基因组组装水平, 其全基因组大小为 557 Mb, scaffold N50 为 442 kb, BUSCO 注释率为 93.3%<sup>[36]</sup>。海天牛科目前还没有染色体水平的基因组组装结果, 所以对海天牛基因组的检测仍旧是一个具有新颖性的课题。基因组调查, 也称作 Survey, 基于深度达到 20~30×以上的高质量二代测序数据对物种的基因组大小与特征进行解读, 可以为物种基因组测序方案提供重要的指导<sup>[37-38]</sup>。在基因组调查的基础上, 结合流式细胞仪可提升基因组大小预测结果的精准性<sup>[39-40]</sup>。本研究中, 利用 84K 杨为对照, 通过流式细胞术预测海天牛基因组大小均值为 794.562 Mb, K-mer 分析的结果显示其基因组大小为 724~730 Mb, 两者的结果偏差较小, 不影响对基因组测序决策的判断<sup>[41]</sup>。所有的结果显示, 白边侧足海天牛是一个高度杂合的物种, 且基因组大小超过 700 Mb。为了达到染色体级别的组装水平, 全基因组测序建议使用以三代测序技术为主, Hi-C 或 Hi-Fi 技术相结合的测序手段<sup>[42-44]</sup>, 测序量达到 80×~100×的深度足够完成海天牛基因组的精细组装。

### 参考文献:

- [1] ASSUYUTI Y, YUSLI W. New distribution record of *Elysia leucolegnote* (Jensen, 1990) (Sacoglossa Plakobranchnidae) in mangrove ecosystem of Biak Numfor [J]. Papua-Indonesia. Biodiversity Journal, 2020, 11: 259-262.
- [2] 李歆毓, 崔闯, 安启源, 等. 海南省海口市发现白边侧足海天牛[J]. 动物学杂志, 2021, 56(4): 631-635.
- [3] SANCHEZ-ESCALONA, KATHERINE. *Elysia leucolegnote* (Ophistobranchia: Sacoglossa) Jensen 1990, a new record for the Verde Island Passage [J], Philippines, 2021, 11: 42-48.
- [4] JENSEN K. Three new species of *Ascoglossa* (Mollusca, Opisthobranchia) from Hong Kong, and a description of the internal anatomy of *Costasiella pallida* Jensen, 1985 [M]. //Proceedings of the 2nd International Marine Biological Workshop: The marine Flora and Fauna of

- Hong Kong and Southern . Hong Kong: Hong Kong University Press, 1986:419-432. .
- [5] 尤仲杰, 林光宇. 中国近海海天牛科(后鳃类)的研究[J]. 浙江海洋学院学报(自然科学版), 2007, 26(4): 363 – 370.
- [6] DE VRIES J, ARCHIBALD J M. Plastid autonomy vs nuclear control over plastid function [J]. Elsevier Ltd. Adv Bot Res., 2018, 85: 1 – 28.
- [7] MAEDA T, TAKAHASHI S, YOSHIDA T, et al. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchnus ocellatus* [J]. *Elife.*, 2021, 10: e60176.
- [8] CHRISTA G, HÄNDELER K, SCHÄBERLE T F, et al. Identification of sequestered chloroplasts in photosynthetic and non-photosynthetic sacoglossan sea slugs (Mollusca, Gastropoda) [J]. *Front Zool*, 2014, 11(1): 15.
- [9] MUJER C V, ANDREWS D L, MANHART J R, et al. Chloroplast genes are expressed during intracellular symbiotic association of *Vaucheria litorea* plastids with the sea slug *Elysia chlorotica* [J]. *Proc Natl Acad Sci USA*, 1996, 93: 12333 – 12338.
- [10] RUMPHO M E, WORFUL J M, LEE J, et al. Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica* [J]. *Proc Natl Acad Sci USA*, 2008, 105: 17867 – 17871.
- [11] BHATTACHARYA D, PELLETREAU K N, PRICE D C, et al. Genome analysis of *Elysia chlorotica* egg DNA provides no evidence for horizontal gene transfer into the germ line of this kleptoplastic mollusk [J]. *Mol Biol Evol*, 2013, 30: 1843 – 1852.
- [12] CAI H, LI Q, FANG X, et al. Data descriptor: A draft genome assembly of the solar-powered sea slug *Elysia chlorotica* [J]. *Sci Data*, 2019, (6): 1 – 13.
- [13] CHIHARA S, NAKAMURA T, HIROSE E. Seasonality and longevity of the functional chloroplasts retained by the sacoglossan sea slug *Plakobranchnus ocellatus* van Hasselt, 1824 inhabiting a subtropical back reef off Okinawa-jima Island, Japan [J]. *Zool Stud.*, 2020, 59: e65.
- [14] MELO CLAVIJO J, FRANKENBACH S, FIDALGO C, et al. Identification of scavenger receptors and thrombospondin-type-1 repeat proteins potentially relevant for plastid recognition in Sacoglossa [J]. *Ecol Evol*, 2020, 10(21): 12348 – 12363.
- [15] HAVURINNE V, HANDRICH M, ANTINLUOMA M, et al. Genetic autonomy and low singlet oxygen yield support kleptoplast functionality in photosynthetic sea slugs [J]. *J Exp Bot.*, 2021, 72(15): 5553 – 5568.
- [16] HAVURINNE V, TYYSTJÄRVI E. Photosynthetic sea slugs induce protective changes to the light reactions of the chloroplasts they steal from algae [J]. *Elife.*, 2020, 9: e57389.
- [17] STAMATAKIS K, VAYENOS D, KOTAKIS C, et al. The extraordinary longevity of kleptoplasts derived from the Ross Sea haptophyte *Phaeocystis antarctica* within dinoflagellate host cells relates to the diminished role of the oxygen-evolving Photosystem II and to supplementary light harvesting by mycosporine-like amino acids [J]. *Biochim Biophys Acta Bioenerg*, 2017, 1858(2): 189 – 195.
- [18] 李蔚, 刘莉莎, 李仁, 等. 十字花科蔬菜基因组含量的测定与分析[J]. 植物遗传资源学报, 2011, 12(1): 103 – 106.
- [19] 杨勇, 陈克成, 孙天恩. 对几种百合科植物基因组大小的评价[J]. 武汉植物学研究, 1996, 14(3): 199 – 203.
- [20] MOYANO J, RODRIGUEZ-CABAL M A, NUÑEZ M A. Invasive trees rely more on mycorrhizas, countering the ideal-weed hypothesis [J]. *Ecology*, 2021, 102(5): e03330.
- [21] SHANGGUAN W, YONGJIU D, BAOYUAN L, et al. A China dataset of soil properties for land surface modeling [J]. *Adv Model. Earth Syst*, 2013, 5: 212 – 224.
- [22] 陈双双, 齐香玉, 冯景, 等. 基于流式细胞术和基因组 Survey 的绣球基因组大小及特征分析[J]. 江苏农业科学, 2021, 49(12): 6.
- [23] 霍恺森, 曹清河, 王珧, 等. 甘薯近缘野生种 *Ipomoea littoralis* 全基因组 Survey 分析[J]. *热带作物学报*, 2019, 40(10): 2001 – 2005.
- [24] LU MIN, AN HUAMING, LI LIANGLIANG. Genome survey sequencing for the characterization of the genetic background of *Rosa roxburghii* tratt and leaf ascorbate metabolism genes [J]. *PloS one*, 2016, 11(2): e0147530.
- [25] 许亚龙, 金静静, 赵艳珍, 等. 基于高通量测序的烟草甲基因组调查分析[J]. 烟草科技, 2020, 53(11): 1 – 6.
- [26] CHAN C X, VAYSBERG P, PRICE D C, et al. Active host response to algal symbionts in the sea slug *Elysia chlorotica* [J]. *Mol Biol Evol*, 2018, 35: 1706 – 1711.
- [27] CHRISTA G, ZIMORSKI V, WOEHLER C, et al. Plastid-bearing sea slugs fix CO<sub>2</sub> in the light but do not require photosynthesis to survive [J]. *Proc Biol Sci*, 2013, 281(1774): 2493.
- [28] TRENCH R K. Chloroplasts as functional endosymbionts in the mollusc *Tridachia crispata* (Bërgh), (Opisthobranchia, Sacoglossa) [J]. *Nature*, 1969, 222(5198): 1071 – 1072.
- [29] HÄNDELER K, GRZYMBOWSKI Y P, KRUG P J, et al. Functional chloroplasts in metazoan cells – a unique evolutionary strategy in animal life [J]. *Front Zool.*, 2009, 6(1): 28.
- [30] PIERCE S K, CURTIS N E, MASSEY S E, et al. A morphological and molecular comparison between *Elysia crispata* and a new species of kleptoplastic sacoglossan sea slug (Gastropoda: Opisthobranchia) from the Florida Keys, USA [J]. *Mollus Res*, 2006, 26(1): 23 – 38.
- [31] CHRISTA G, GOULD S B, FRANKEN J, et al. Functional kleptoplasty in a limapontioidean genus: phylogeny, food preferences and photosynthesis in *Costasiella* with a focus on *C. ocellifera* (Gastropoda: Sacoglossa) [J]. *J Mollus Stud*, 2014, 80(5): 499 – 507.

- [32] FAN X, QIAO H, XU D, et al. Short-term retention of kleptoplasty from a green alga (*Bryopsis*) in the sea slug *Placida* sp. *YS001* [J]. *Biologia*, 2014, 69: 635 – 643.
- [33] DE VRIES J, CHRISTA G, GOULD S B. Plastid survival in the cytosol of animal cells [J]. *Trends Plant Sci*, 2014, 19(6): 347 – 350.
- [34] CHRISTA G, PÜTZ L, SICKINGER C, et al. Non-photochemical quenching does not prevent kleptoplasts from net photoinactivation [J]. *Frontiers in Ecology and Evolution*, 2018, 6: 121.
- [35] TORRES J P, LIN Z, WINTER J M, et al. Animal biosynthesis of complex polyketides in a photosynthetic partnership [J]. *Nat Commun*, 2020, 11(1): 2882.
- [36] HANTEN J J, PIERCE S K. Synthesis of several light-harvesting complex I polypeptides is blocked by cycloheximide in symbiotic chloroplasts in the sea slug, *Elysia Chlorotica* (Gould): a case for horizontal gene transfer between alga and animal? [J]. *Biological Bulletin*, 2001, 201(1): 34 – 44.
- [37] CAI H, LI Q, FANG X, et al. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica* [J]. *Sci Data*, 2019, 6: 190022.
- [38] MENG X, FU Q, LUAN S, et al. Genome survey and high-resolution genetic map provide valuable genetic resources for *Fenneropenaeus chinensis* [J]. *Sci Rep*, 2021, 11(1): 7533.
- [39] 江思容, 夏志强, 张辰笈, 等. 芭蕉芋和美人蕉全基因组 Survey 分析 [J]. *分子植物育种*, 2022, 20(3): 765 – 771.
- [40] MIYASHITA T, ARAKI H, HOSHINO Y. Ploidy distribution and DNA content variations of *Lonicera caerulea* (Caprifoliaceae) in Japan [J]. *J Plant Res*, 2011, 124(1): 1 – 9.
- [41] BENNETT M D, LEITCH I J. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow [J]. *Annals of botany*, 2011, 107(3): 467 – 590.
- [42] VURTURE, GREGORY W, SEDLAZECK, et al. Genome scope: fast reference-free genome profiling from short reads [J]. *Bioinformatics*, 2017, 33(14): 2202 – 2204.
- [43] 刘贵明. 农作物参考基因组的组装和基于高通量测序的表观调控技术及应用 [C] // “农业健康与环境”组学大数据整合生物信息学研讨会. 蒙古民族大学、浙江省生物信息学学会、浙江省科学技术协会, 2017.
- [44] IMAKAEV M, FUDENBERG G, MCCORD R P, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization [J]. *Nat Methods*, 2012, 9: 999 – 1003.

## Whole genome survey analysis of *Elysia leucolegnote*

LI Xinyu<sup>1,3</sup>, DING Mengying<sup>2</sup>, FENG Erhui<sup>1</sup>, ZHANG Yuanyuan<sup>4,5</sup>, ZHANG Kelan<sup>3</sup>, WAN Yinglang<sup>3</sup>  
 (1. Hainan Dongzhaigang National Natural Reserve Authority, Haikou, Hainan 570100; 2. College of Marine Biology, Hainan University, Haikou, Hainan 570228; 3. College of Tropical Crops, Hainan University, Haikou, Hainan 570228; 4. Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101; 5. Key Laboratory of Biology and Utilization of Genetic Resources of Rubber Tree, Ministry of Agriculture and Rural Affairs, Haikou, Hainan 571101, China)

**Abstract:** *Elysia leucolegnote* are sacoglossan sea slugs, belonging to the family Elysiidae and the genus *Elysia*. They are distributed in mangroves throughout China, Thailand and the Philippines. Photosynthetic sea slugs are famous for their ability of photosynthesis. Stolen chloroplasts are called kleptoplasts. However, there has been no important breakthrough in the molecular mechanism research of kleptoplasts. In this context whole genome survey analysis of *E. leucolegnote* was conducted to find the clue about the regulation of photosynthesis and maintenance of chloroplast activities by high-throughput sequencing technology. The high-throughput sequencing produced 25 Gb high-quality data in *E. leucolegnote*, with a total sequencing depth of about 30x. According to 17-mer analysis, the predicted genome size was 724.8 Mb, the genome duplication rate was 52.8%, the heterozygosity was 1.55%, and the model fitting value was 99.38%. The 19-mer analysis predicted that the genome size was 730.8 Mb, with the genome duplication rate of 35.1%, the heterozygosity of 1.68%, and the model fitting value of 99.72%. A draft genome of *E. leucolegnote* Ker was obtained by preliminary assembly with a total length of 628 Mb. These results showed that *E. leucolegnote* is a highly heterozygous species with a genome size of more than 700 Mb.

**Keywords:** *Elysia leucolegnote*; genomic survey analysis; genome size; heterozygosity