

秋海棠属基因组揭示其多样性形成及适应性进化机制

华大智造stLFR技术搭配DNBSEQ测序平台助力复杂物种基因组高质量组装

由深圳市仙湖植物园与深圳华大生命科学研究院牵头完成了秋海棠属植物基因组研究，并在 *New Phytologist* 发表了题为 “Genomes shed light on the evolution of Begonia, a mega-diverse genus” 的研究成果¹。

该研究基于多种长读长测序技术，完成了4种秋海棠属植物全基因组的测序和组装工作，更好地解释了秋海棠属的起源、进化和荫蔽适应性。其中铁十字秋海棠和黑武士秋海棠的基因组是采用华大智造stLFR技术构建文库测序获得的高质量基因组，并辅助组装成杂合度较高的基因组。该研究中部分测序工作是在华大智造DNBSEQ测序平台完成。

推荐应用：DNA长读长测序

推荐机型：MGISEQ-2000RS，DNBSEQ-T7RS

• 更全面的长片段DNA信息读取

MGIEasy stLFR试剂盒可以检测分析10kb-300kb长度的DNA分子，同时兼顾二代三代测序技术的优势。

• 各类变异检测表现极佳

通过一次检测可以获得多种类型的变异信息，整体变异检测表现达到相同深度的常规全基因组测序文库表现的最佳水平。

• 覆盖均匀性

仅1 ng DNA起始的stLFR文库即可获很好的基因组覆盖均匀性。

• 更好的高度同源基因组对比

可以获得高度同源基因组的更多信息，从而产生更好的基因组比对结果。

• 提供配套产品组合

华大智造stLFR技术搭配自动化建库设备与stLFR *de novo* 组装软件，可为客户提供全套产品方案。



背景介绍

秋海棠属是植物中的一类大属，其种类高达2000种，其进化枝的多样化机制仍是植物生物学的谜团之一²。最大的10个被子植物属中仅有三个具有代表性的完全组装的核基因组已发表，即茄属³、石斛属⁴和海棠⁵。铁十字秋海棠属是一种泛热带的，包括 2000 多种目前公认的草本植物和偶尔的亚灌木物种。因此，它代表了一个极好的进化研究系统，用于产生许多密切相关的物种的过程。

秋海棠在美洲和亚洲有着很高的物种多样性，而在其被推测的起源地非洲的物种数量反而较少。秋海棠属多为阴生种属，在阳光充足的情况下很容易受到损伤。但秋海棠的不同物种表现出了适应阳光从极少到长期暴晒的连续光适应能力，使得对不同水平的光的适应机制有机会被解开，利用秋海棠的这一机制了解使作物如何适应阴影优化光合作用和物理防御有着巨大价值。有研究比较了四种秋海棠的基因组并以此重建了秋海棠的古基因组，对秋海棠史前全基因组加倍的时间对其进化的意义进行了分析。此外，Brennan 等人基于细胞核的不一致分析了秋海棠光适应度的分子基础⁶。这些研究基础对秋海棠属植物多样性形成及适应性进化都提供了有意义的参考。

本研究通过长度长测序的方法和基因组的组装，从进化的角度更好地解释了秋海棠属基因组的多样性。其中秋海棠属的高质量基因组组装是一个难点，而华大智造研发的stLFR(single tube Long Fragment Read)技术有明显的优势，可以助力完成这一难点。stLFR是一种无分割共标签长片段读取技术（图1），搭配华大智造DNBSEQ™系列测序平台，可通过短读长测序获取长片段DNA信息，从而实现一次测序获得高准确度的SNP/InDel/CNV/SV变异检测结果，以及单倍体分型结果⁷。stLFR技术在长片段读取的基础上，提供了更全面更均匀的基因组测序，辅助完成秋海棠属植物基因组的组装，为秋海棠属适应机制和进化机制的测序分析提供了强有力的技术。

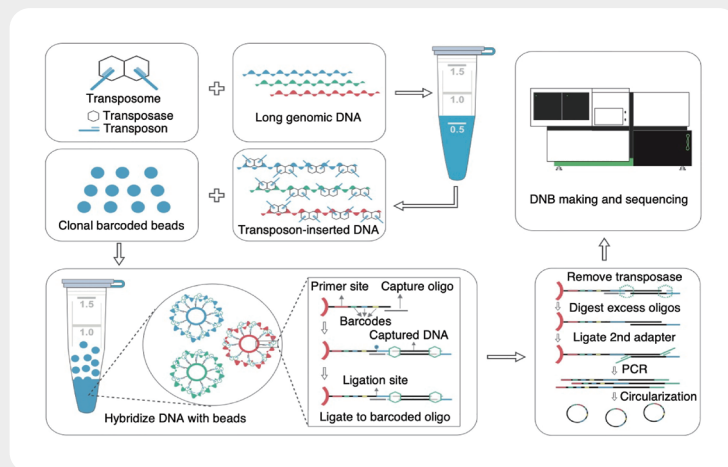


图1. stLFR 文库构建流程示意图

实验方法

样本收集和DNA/RNA提取

所有秋海棠样品均取自仙湖植物园的温室，通过基于十六烷基三甲基溴化铵的方法提取幼叶基因组DNA进行全基因组测序（WGS）；用植物组织DNA分离试剂盒分离获得的高分子量基因组DNA进行stLFR测序；分别提取四种秋海棠的根、茎、叶、花梗和花的mRNA，进行转录组测序。

文库制备和测序

用于WGS的DNA文库均使用MGIEasy FS DNA文库制备套装构建以获得主峰为300–500bp片段大小的文库后进行双端150bp(PE150)测序。转录组文库则使用RNA文库制备试剂盒构建以获得主峰为200–400bp片段大小的文库，并进行双端100bp(PE100)测序。每个文库生成了超过5Gb的序列数据。

stLFR 文库使用MGIEasy stLFR文库制备试剂盒构建，然后进行双端100bp+42bp的读长测序，每个文库产生>150Gb的原始序列数据(图2)。10x Genomics Chromium基因组文库的插入片段大小为350-500bp，使用配套试剂盒构建而成，后通过引入合适的测序引物在DNBSEQ平台进行PE150测序。同时，其中部分数据是基于PacBio进行测序，相关SMART文库构建以及测序细节可参阅原文附件进行了了解。

流程分析

对于 10xGenomics 的文库和 stLFR 文库的序列组装，使用该研究团队自研的内部脚本获得准确的读数，后使用 SPERNOVA 进行从头组装获得，在 '输出' 阶段指定了 100bp 的最小快速记录大小，用于

以 'pseudohap' 样式输出程序集。而 PacBio 文库测序所获数据则由 CANU 进行从头组装，使用 RACON 软件进行 PacBio 长读长的两轮迭代修正和 PILON 修正。对于 stLFR 相关的序列组装，stLFR de novo 基因组从头组装发挥了重要作用。经过基因组的组装后，该研究随后进行了变异检测分析、系统发育分析、基因组共线性分析、叶绿素荧光分析和捕光 Chl a/b 结合蛋白超家族的鉴定与系统发育分析等。

结果

基因组测序和基因组特征

对桑寄生状秋海棠 (*Begonia loranthoides*)、铁十字秋海棠 (*Begonia masoniana*)、黑武士秋海棠 (*Begonia darthvaderiana*)、和盾叶秋海棠 (*Begonia peltatifolia*) 4种秋海棠属植物进行全基因组组装，并完成了74个全球代表性的种类的浅层基因组测序 (图2)。组装结果显示，4个物种基因组大小介于331Mb-799Mb，基因数量介于22,059-23,444之间，BUSCO评估均达到91%以上，高质量的基因组组装为秋海棠属植物功能基因组学研究，其中stLFR长片段从头组装优势明显。

全基因组加倍事件

秋海棠属在大约三千五百万年前发生了一次全基因组加倍事件 (WGD)，这次事件对秋海棠属的植物多样性演化有重要的作用，并且使得秋海棠属植物的基因得以保留与富集，保留了多个拷贝 (图3)，同时秋海棠属对环境的适应性在这个时期也出现了变化。

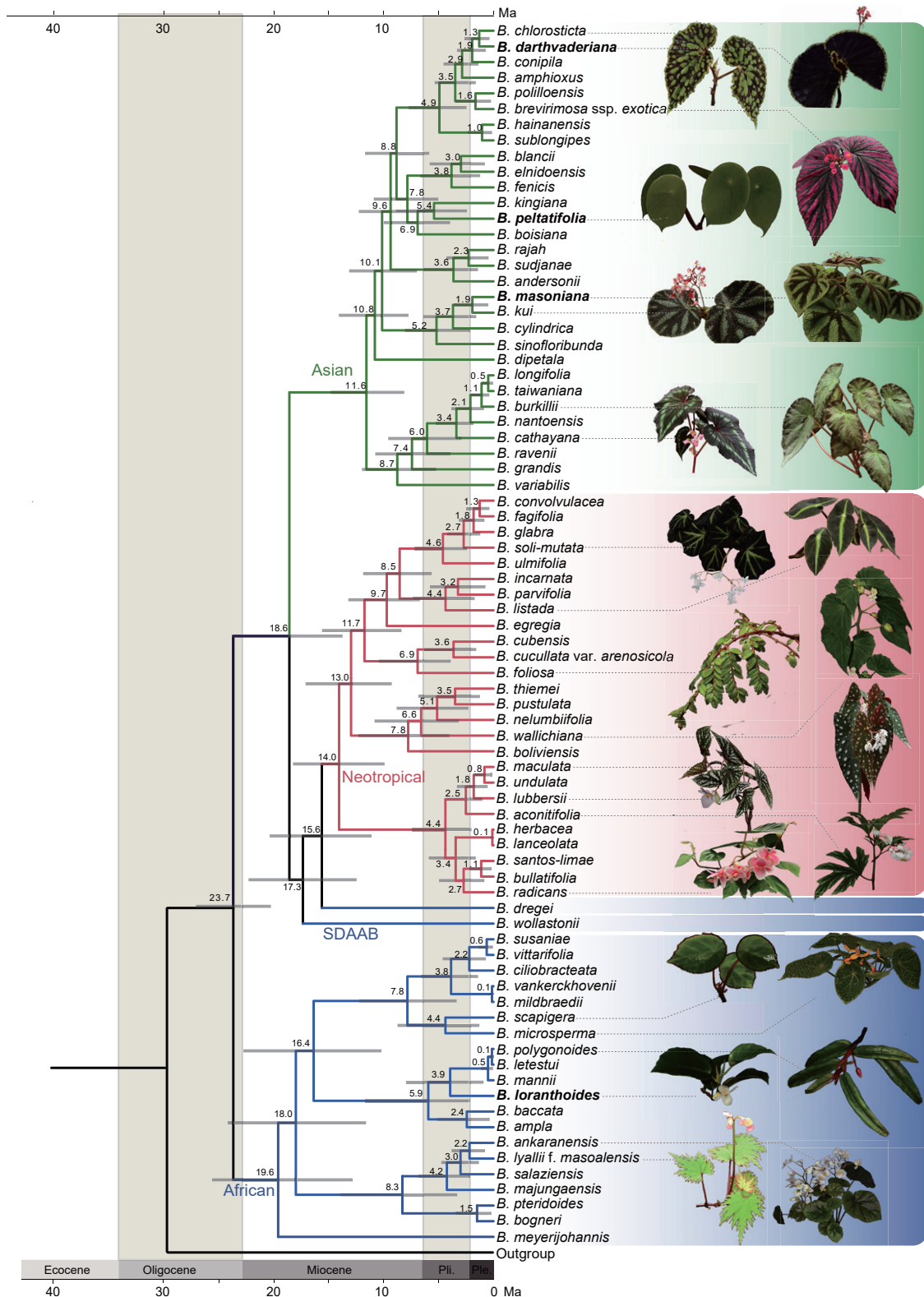


图2. 秋海棠属植物发育树

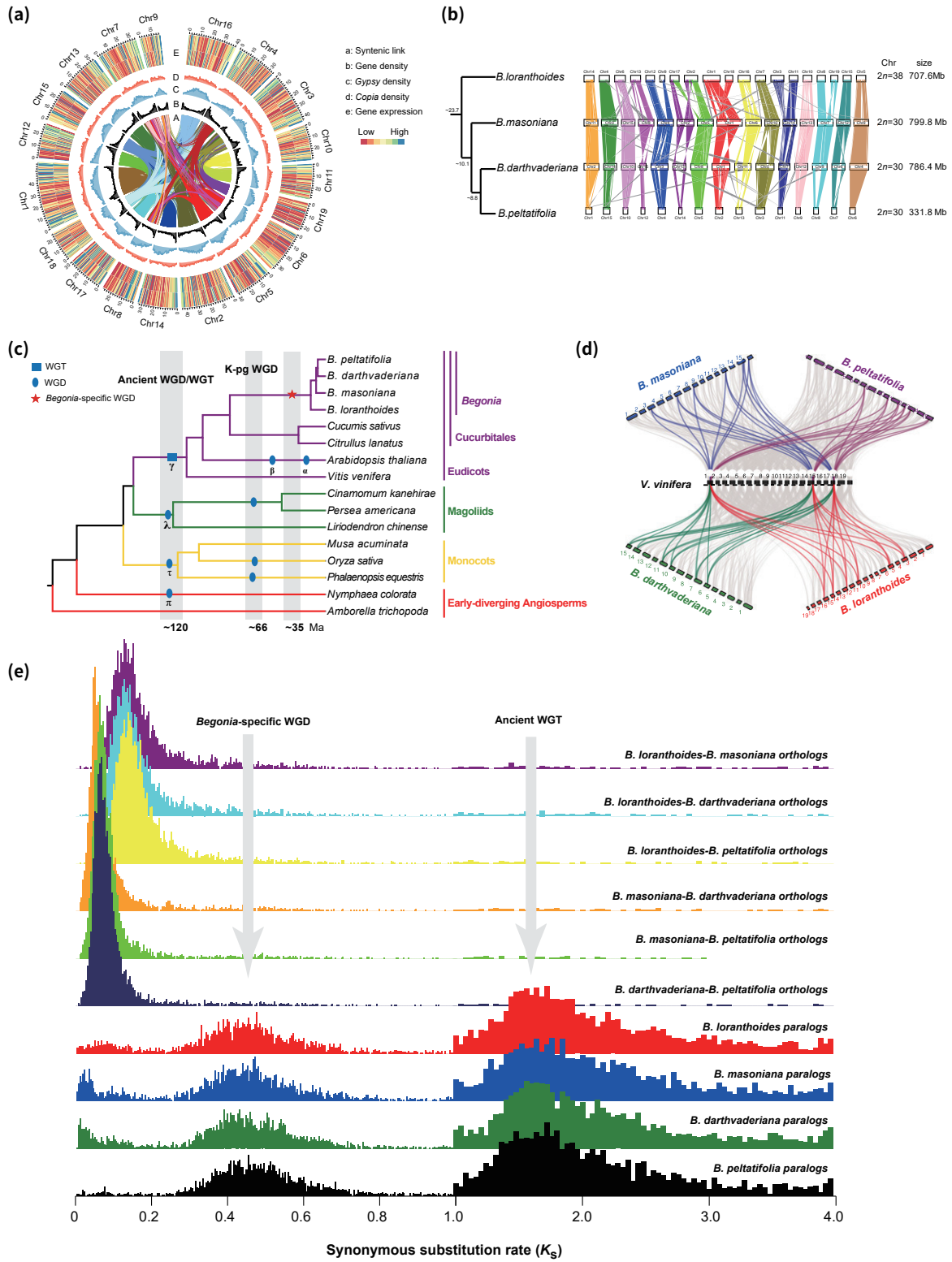


图3. 全基因组加倍事件

转座因子进化和分布

转座因子在秋海棠属植物基因组中占有很大比例，且种间特异性较高。转座因子在功能基因启动子及内含子等不同位置插入的模式及数量在不同物种间存在明显的差异，杂交与基因渐渗推动

了秋海棠属植物多样化形成。研究发现在美洲分布的类群，秋海棠属的祖先可能发生了多次杂交事件，揭示了转座因子与秋海棠物种的形成及适应性关系十分密切(图4)。

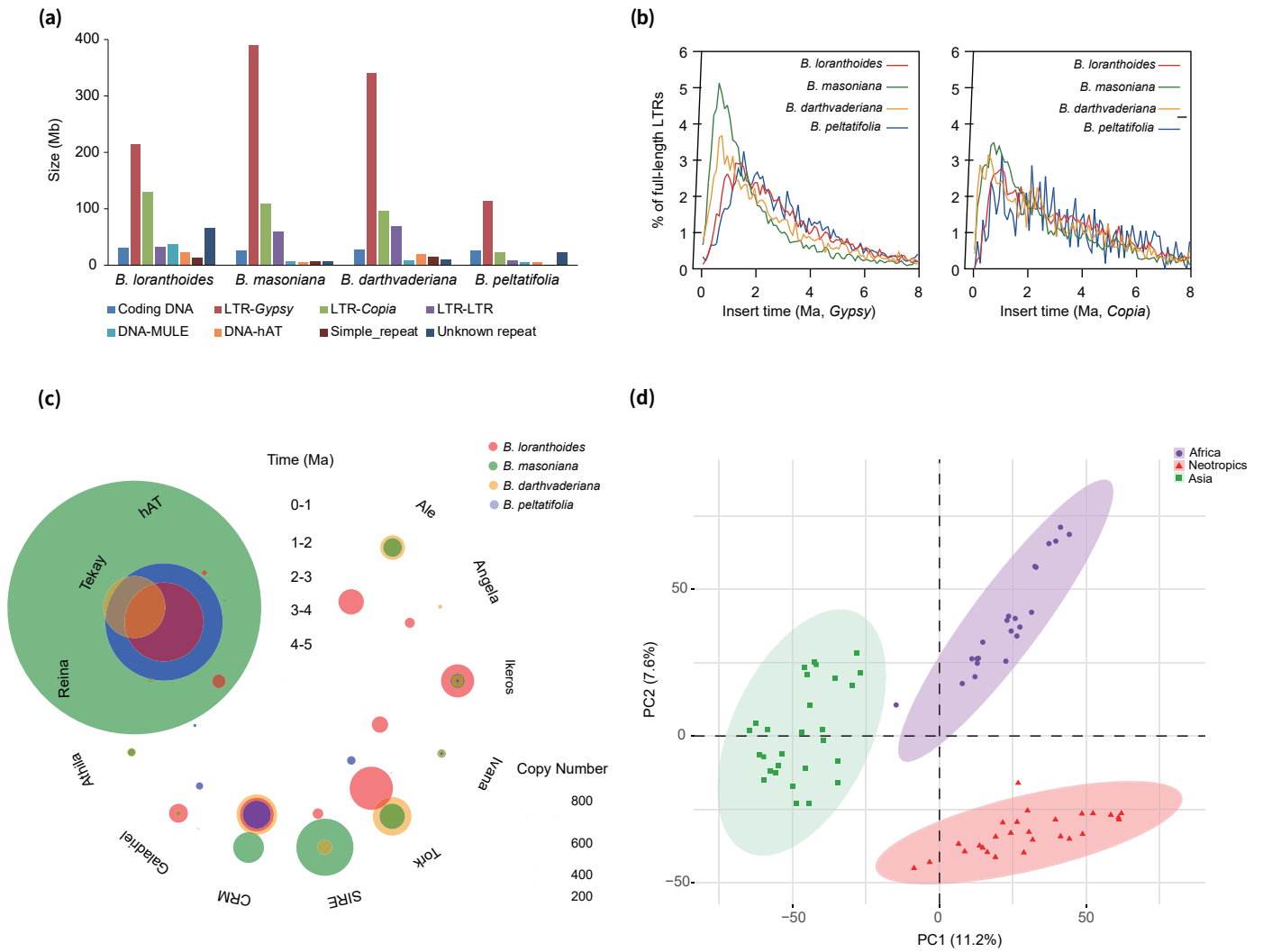


图4. 转座因子进化和分布

阴影适应的演变

WGD使得很多与光合作用及能量代谢相关的基因得以保留与富集，研究发现与红光、蓝光及紫外光接收有关的受体基因（PHOT、CRY1/2、PHY及UVR8）均保留了多个拷贝，且部分已经发生了功能分化，这对秋海棠属植物适应阴生环境、提高光合利用率具有重要意义（图5）。

及UVR8）均保留了多个拷贝，且部分已经发生了功能分化，这对秋海棠属植物适应阴生环境、提高光合利用率具有重要意义（图5）。

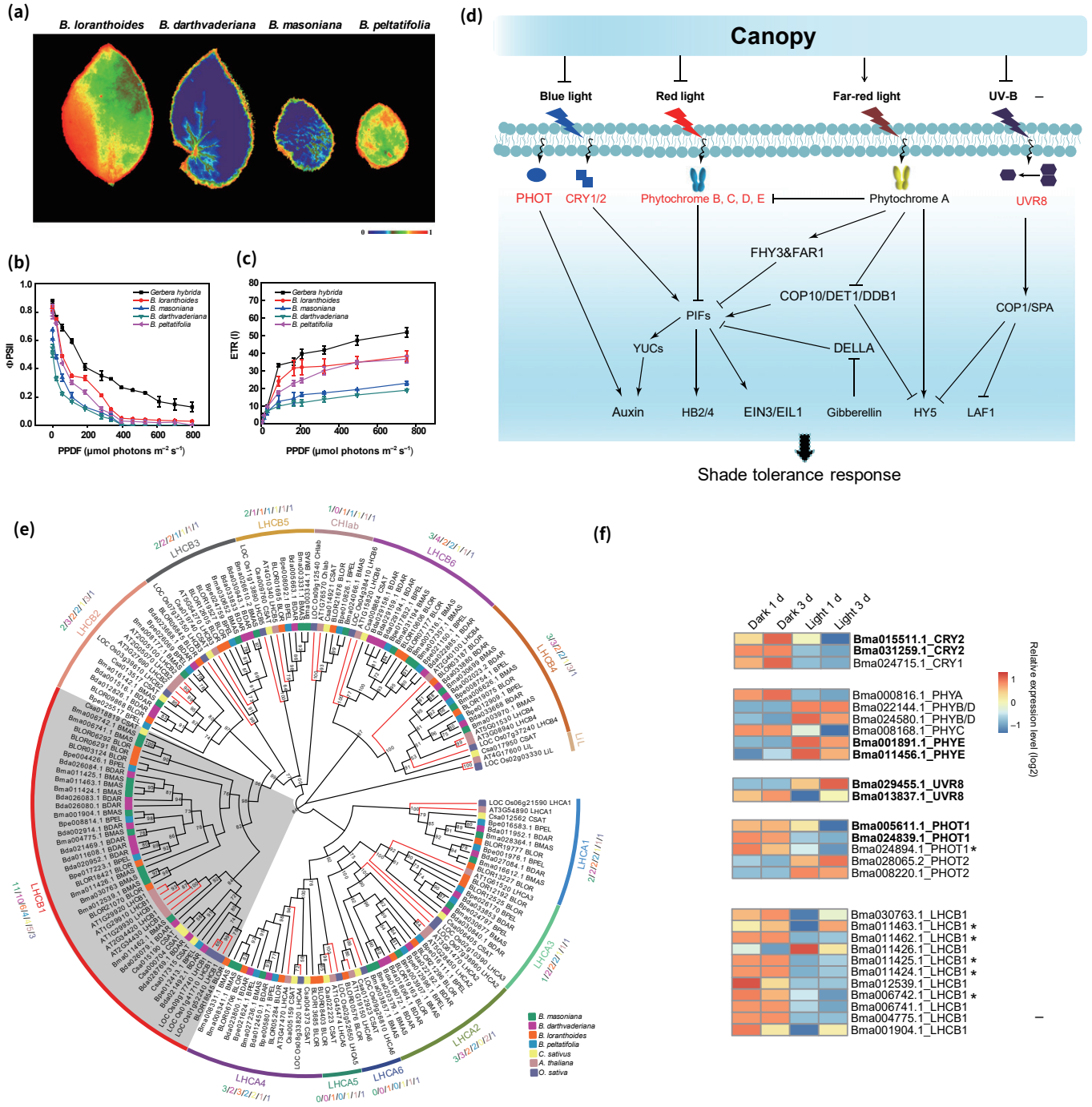


图5. 秋海棠属适应阴性机制

仅供研究使用，不适用于临床诊断

总结

该研究成果为揭示秋海棠属植物多样性形成及适应性进化提供了新的见解，秋海棠属基因组组装使用的stLFR技术可以给其他类似基因组学研究提供技术参考，华大智造DNBSEQ测序平台在长片段读长测序和高质量基因组组装方面的优势也是其他基因组测序研究的参考选择之一。



基因测序仪MGISEQ-2000

参考文献

1. Li, L. et al. Genomes shed light on the evolution of Begonia, a mega-diverse genus. *New Phytol* **234**, 295-310, doi:10.1111/nph.17949 (2022).
2. Frodin, D. G. History and concepts of big plant genera. *Taxon* **53**, 753-776, doi:Doi 10.2307/4135449 (2004).
3. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
4. Yan, L. et al. The Genome of *Dendrobium officinale* Illuminates the Biology of the Important Traditional Chinese Orchid Herb. *Mol Plant* **8**, 922-934, doi:10.1016/j.molp.2014.12.011 (2015).
5. Griesmann, M. et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, doi:10.1126/science.aat1743 (2018).
6. Gommers, C. M., Visser, E. J., St Onge, K. R., Voeselek, L. A. & Pierik, R. Shade tolerance: when growing tall is not an option. *Trends Plant Sci* **18**, 65-71, doi:10.1016/j.tplants.2012.09.008 (2013).
7. Wang, O. et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res* **29**, 798-808, doi:10.1101/gr.245126.118 (2019).

推荐订购信息

产品类型	产品名称	产品货号
仪器	基因测序仪MGISEQ-2000RS	90000003500
	基因测序仪DNBSEQ-T7RS	900-000236-00
	MGISP-960RS 自动化样本制备系统	900-000100-00
软件	MegaBOLT生信分析加速器（工作站式服务器）	970-000085-00
	ZTRON Pro 一体机（T500）	900-000444-00
	MGI-tech-bioinformatics/stLFR_v1	https://github.com/MGI-tech-bioinformatics/stLFR_v1
建库试剂	MGIEasy stLFR文库制备试剂盒(16 RXN)	94000019300
	MGIEasy Fast 酶切DNA制备试剂套装(16 RXN)	94000002900
	MGIEasy RNA 文库制备试剂盒(16 RXN)	1000005274
测序试剂	MGISEQ-2000RS 高通量测序试剂套装（stLFR FCL PE100）	1000016984
	DNBSEQ-T7RS 高通量测序试剂套装(stLFR FCL PE100)	1000019251
	MGISEQ-2000RS 高通量测序试剂套装（FCL PE100）	1000012554
	MGISEQ-2000RS 高通量测序试剂套装（FCL PE150）	1000012555

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋

☎ 4000-688-114

🌐 www.mgi-tech.com

✉ MGI-service@mgi-tech.com

股票简称：华大智造

股票代码：688114



仅供研究使用

版权声明：本手册版权属于深圳华大智造科技股份有限公司所有,未经本公司书面许可,任何其他个人或组织不得以任何形式将本手册中的各项内容进行复制拷贝、编辑或翻译为其他语言。本手册中所有商标或标识均属于深圳华大智造科技股份有限公司及其提供者所有。

版本：2023年6月版

撰稿：吴居业 张嘉乐

责任编辑：王其伟

审稿：江遥