

泛基因组图谱揭示转座子在非洲栽培稻驯化中的作用

李笑霞^{1,2,#} 张天驰^{1,2,#} 贺文闯² 贺慧英² 盘楠² 刘祎凡² 钱前^{1,2,3} 商连光^{2,3,*}

[¹ 沈阳农业大学 水稻研究所, 沈阳 110161; ² 中国农业科学院 深圳农业基因组研究所(岭南现代农业科学与技术广东省实验室深圳分中心), 深圳 518120; ³ 崖州湾国家实验室, 三亚 572024; #共同第一作者; *通信作者, email: shanglianguang@caas.cn]

A Pan-genome Map Reveals the Contribution of Transposable Elements to Domestication of *Oryza glaberrima*

LI Xiaoxia^{1,2,#}, ZHANG Tianchi^{1,2,#}, HE Wenchuang², HE Huiying², PAN Nan², LIU Yifan², QIAN Qian^{1,2,3}, SHANG Lianguang^{2,3,*}

[¹Rice Research Institute, Shenyang Agricultural University, Shenyang 110161, China; ²Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China; ³Yazhouwan National Laboratory, Sanya 572024, China; #These authors contributed equally to this work; *Corresponding author, email: shanglianguang@caas.cn]

Abstract: [Objective] Transposable elements (TEs) are a significant source of genetic diversity in rice and play a crucial role in the domestication process of Asian cultivated rice. However, due to the highly repetitive nature of TE sequences, their identification typically requires high-quality genome sequences. Currently, the limited genomic resources for African cultivated rice (*Oryza glaberrima*, *Og*) hinder systematic investigations into TE diversity and its contributions to the *Og* domestication process. [Method] We constructed a high-resolution pan-TE variation map using the chromosome-level genomes of 21 accessions (11 *Og* and 10 *Oryza barthii* [*Ob*]). We estimated the divergence of each TE variation across the genome using *F_{ST}* analysis and identified 854 selected TE variations putatively involved in the *Og* domestication process. [Result] The pan-TE variation map comprises 17,097 non-redundant high-quality TE variations with a total length of 39.1 Mb, with 61.7% of the TE variations larger than 1 kb. The total number of TE variations in the *Ob* accessions was greater than that in the *Og* accessions. A total of 658 genes harboring selected TE variations were identified as potentially involved in the *Og* domestication process, which were significantly enriched in biological processes related to basic metabolism. Notably, 30 genes with known functions were involved in the independent domestication of *Og*. [Conclusion] This study highlights the significant role of TE variations as a source of phenotypic diversity during the *Og* domestication process. The high-quality pan-TE variation map not only enriches the available variation library for rice breeding but also provides critical theoretical foundations and resource support for the genetic improvement of *Og* germplasm.

Key words: *Oryza glaberrima*; *Oryza barthii*; transposable element; pan-genome; domestication

摘要: 【目的】转座子(TE)是水稻遗传多样性的重要来源,在亚洲栽培稻(*Oryza sativa*)的驯化过程中发挥了关键作用。然而,由于TE序列的高度重复性,其鉴定通常依赖于高质量的基因组序列。目前,非洲栽培稻(*Oryza glaberrima*, *Og*)的基因组资源有限,导致对其TE的多样性及其在驯化过程中的贡献仍然缺乏系统性研究。【方法】本研究基于11份*Og*和10份短舌野生稻(*Oryza barthii*, *Ob*)的染色体级别的基因组序列,构建了*Og*和*Ob*群体的高精度泛TE变异图谱。通过对该图谱进行选择分析,鉴定了854个在*Og*驯化过程中受到选择的TE变异。【结果】该泛TE图谱共包含17,097个高质量的TE变异,覆盖基因组总长度39.1 Mb,其中61.7%的TE变异长度超过1 kb。*Ob*群体的TE变异总数高于*Og*群体。在*Og*驯化过程中,共鉴定出658个含受选择TE变异的基因,这些基因显著富集于基础代谢相关的生物过程,其中包括30个在*Og*独立驯化过程中发挥作用的已知功能基因。【结论】本研究揭示了TE变异是*Og*驯化过程中表型变异的重要来源,构建的高质量泛TE变异图谱不仅丰富了水稻育种的可用变异库,也为*Og*种质资源的遗传改良提供了重要的理论依据和资源支持。

关键词: 非洲栽培稻; 短舌野生稻; 转座子; 泛基因组; 驯化

水稻作为全球近半数人口的主食来源, 其产量的提升对于应对全球人口不断增长所带来的食物需求至关重要^[1]。非洲栽培稻(*Oryza glaberrima*, *Og*)与亚洲栽培稻(*Oryza sativa*)的亲缘关系较近, 是非洲特有的栽培稻种。大约 3,000 年前, 在非洲西部, *Og* 由短舌野生稻(*Oryza barthii*, *Ob*)独立驯化而来, 并因其对非生物胁迫(如干旱、盐碱等)的显著耐受性而著称^[2]。*Og* 的发展对于促进非洲的粮食安全、经济增长、技术合作、品种改良以及社会变革具有深远的意义。

转座子(transposable elements, TE)是广泛存在于真核生物基因组中的一类可移动、自主复制的 DNA 重复序列, 其序列长度从数百碱基对(bp)至数万 bp 不等。TE 不仅能够影响基因功能、创造新的性状、增加个体的独特性, 还能够在压力条件下被激活, 从而协助生物体适应复杂多变的自然环境^[3]。1950 年, 著名遗传学家 Barbara McClintock 首次观察到 TE 与玉米籽粒和叶片颜色变化之间的关联, 进而揭示了 TE 在表型变异中的核心作用^[4]。根据转座机制的不同, TE 可分为逆转录 TE 和 DNA TE 两大类^[5]。逆转录 TE 首先转录为 RNA, 再由 RNA 中间体逆转录为 DNA, 并最终插入到新位置进行转座, 主要包括长末端重复(LTR, 包括 *Gypsy*、*Copia* 和 *unknown*)元件以及非 LTR(短散在逆转录转座元件[SINEs]和长散在逆转录转座元件[LINES])元件。而 DNA TE 则通过 DNA 中间体, 采用“剪切和粘贴”的机制进行转座, 主要包括微型反向重复转座元件(MITE, 包括 *Stowaway* 和 *Tourist*)、DTC(*CACTA*)、DTA(*hAT*)、DTT(*Tcl-Mariner*)、DTM(*Mutator*)、DTH(*PIF-Harbinger*)以及 *Helitron* 元件。与单核苷酸多态性(SNP)相比, TE 变异通常能够解释更多的表型变异, 且越来越多的证据表明, TE 变异与亚洲栽培稻的驯化和分化过程紧密相关^[6]。例如 *MYB61* 启动子区域的 *Helitron* 会影响水稻对氮的利用^[7]; 在 *GY3* 启动子区域插入的 LTR 会抑制其表达, 从而导致幼穗中活性细胞分裂素含量增加^[8]; 而 *PIF-Harbinger* 衍生的基因 *PANDA* 则通过表观遗传调控影响水稻的穗数和籽粒大小^[9]。前期研究还发现, 在 *Og* 驯化过程中, *GLAG5* 启动子区域一个 *hAT* 的插入引起该区域 DNA 甲基化, 进而抑制基因表达, 导致籽粒呈现光壳表型^[10]。

然而, 截至目前, *Og* 和 *Ob* 的泛 TE 变异图谱尚未构建完成, TE 变异对 *Og* 驯化过程的影响仍未得到充分研究。在本研究中, 我们利用 11 份 *Og* 和 10 份 *Ob* 的高质量基因组序列, 成功构建了一个高

质量、群体水平的泛 TE 变异图谱。通过该变异图谱进行群体选择分析, 我们揭示了 TE 变异在 *Og* 驯化中的重要作用, 为基因克隆和分子育种提供了丰富的遗传信息。

1 材料与方法

1.1 材料和数据收集

我们从课题组网站上(<http://www.ricesuperpir.com>)下载日本晴端粒到端粒(T2T)的参考基因组^[11]、11 个 *Og* 和 10 个 *Ob* 的染色体级别、高质量的基因组序列及其对应的 TE 注释文件^[12]。

1.2 泛 TE 变异图谱的构建

为构建高质量的 *Og* 和 *Ob* 的泛 TE 变异图谱, 我们使用 minimap2(version 2.24-r1122)^[13]将 21 个材料的基因组序列与参考基因组日本晴^[11]进行全基因组比对分析。随后, 利用 CHAIN/NET/NETSYNTENY^[14]软件和 Perl 脚本(https://github.com/yiliaol022/LASTZ_SV_pipeline)进一步处理和过滤数据, 鉴定出存在/缺失变异(PAV, >50 bp)。接着, 使用 SVJedi(version 1.1.6)^[15]软件结合纳米孔长读长测序数据, 进一步鉴定区分 PAV 的基因型(包括纯合和杂合基因型)。最后, 再整合 21 个基因组(11 个 *Og* 和 10 个 *Ob*)的 TE 注释文件, 判断如果一个 PAV 序列中, 超过 90%的序列是 TE 序列时, 则该 PAV 被视为 TE 变异, 最终构建一个高质量的 TE 变异图谱。

1.3 基因型和群体结构分析

根据已发布基因组数据^[1]下载 21 份水稻材料的基因型数据。根据 SNP 位点信息, 使用 plink 1.90b6.2^[16]对材料群体进行主成分分析(PCA), 使用 Admixture 1.3.0 评估群体结构^[17]。

基于 TE 变异数据, 采用 FastTree(version 2.1.11)软件^[18]中 Jones-Taylor-Thornton CAT 模型, 构建最大似然系统发生树。用 iTOL(<https://itol.embl.de/>)将系统发生树可视化。

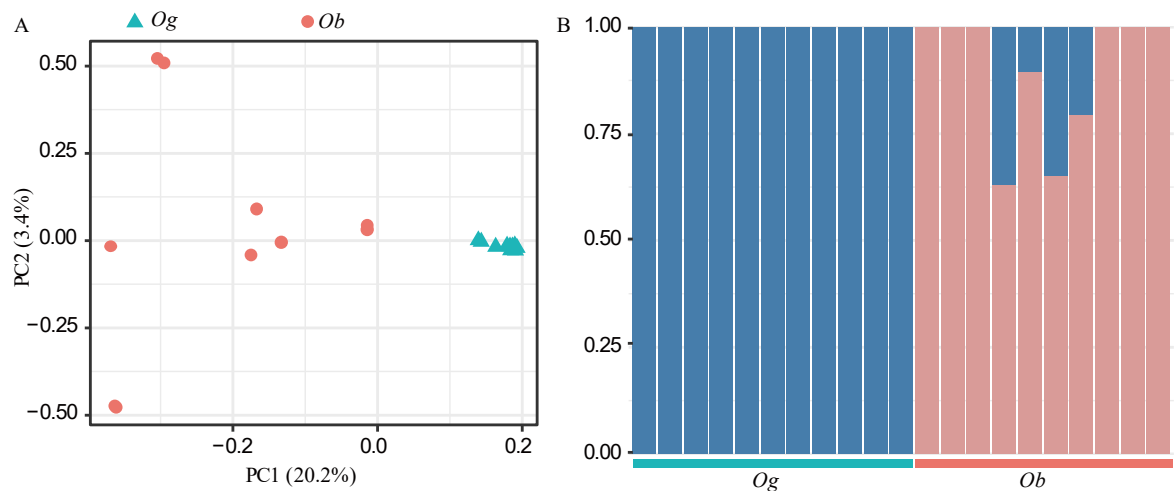
1.4 非洲栽培稻驯化分析

采用 VCFtools(version 0.1.16)软件^[19]进行群体 F_{ST} 分析, 鉴定在 *Og* 驯化过程中受选择的 TE 变异。通过计算每个 TE 变异的 F_{ST} 值并排序, 选取前 5% 的 F_{ST} 值对应的 TE 变异作为受选择的 TE 变异。

2 结果与分析

2.1 *Og* 和 *Ob* 的泛 TE 变异图谱构建

基于主成分分析和群体结构将 21 份材料分为

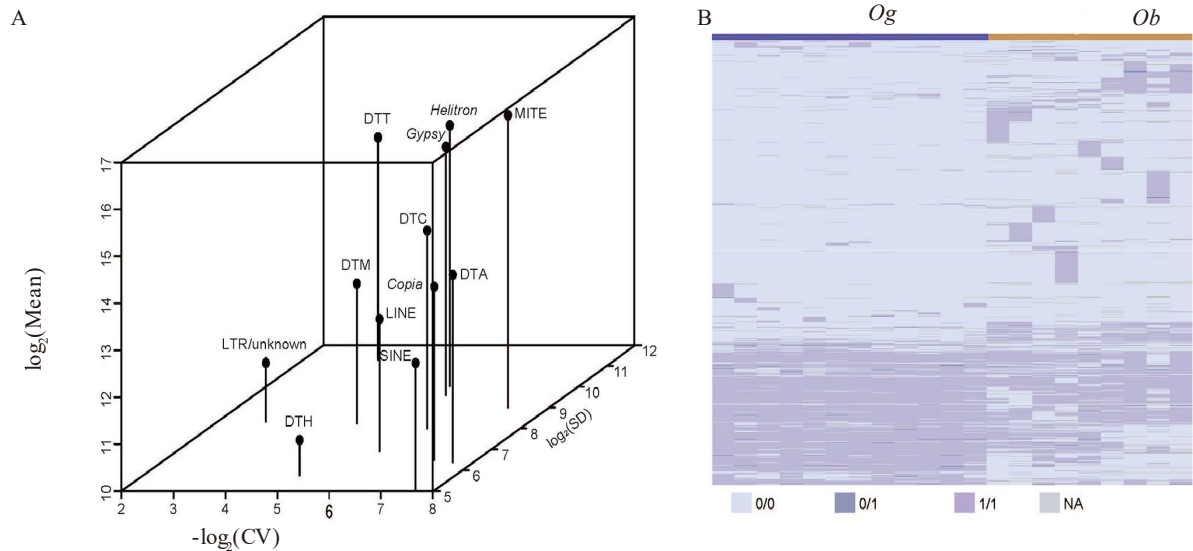


A: 主成分分析(PCA); B: 群体结构分布。
A, Principal component analysis(PCA); B, Population structure distribution.

图1 21份非洲稻材料的群体结构分析
Fig. 1. Population structure analysis of 21 African rice accessions

11个 *Og* 和 10个 *Ob*(图1)。基于这11个 *Og* 和10个 *Ob* 的高质量基因组序列及其对应的TE注释文件^[11], 我们发现这些材料中TE数量的变异系数(CV)和标准差(SD)分别为0.005 ~ 0.131和32 ~ 2506(图2-A), 表明这21份材料中的TE数目高度多样化, 为我们研究TE对 *Og* 和 *Ob* 进化的影响提供了重要基础。

为构建 *Og* 和 *Ob* 的泛TE变异图谱, 我们使用 minimap2^[12]将21个染色体水平的基因组序列与参考基因组日本晴^[11]进行全基因组比对, 并鉴定出长度超过50 bp的PAV。当一个PAV序列中, 有超过90%的序列是TE序列时, 我们将该PAV视为TE变异。最终, 我们成功构建了一个包含11个 *Og* 和10个 *Ob* 材料、非冗余的高质量泛TE变异图谱(图



A: 21个材料中不同TE家族的TE数目。X轴表示每个TE家族在单个材料中TE数目差异, 以变异系数(CV)评估; Y轴表示每个TE家族在单个材料中总TE数目的差异, 以标准差(SD)评估; Z轴表示每个TE家族在单个材料中平均TE数目的差异; B: 21个样品构建的泛TE图谱, *Og* 和 *Ob* 分别指非洲栽培稻和短舌野生稻。

A, Number of TE for each TE family in 21 African accessions. X axis represents the number of TE for each TE family across accessions, evaluated as coefficient of variation (CV); Y axis represents the differences in total TE number for each family among accessions, evaluated as standard deviation (SD); Z axis represents the average number of TEs for each family; B, Landscape of TE variations across the 21 African accessions.

图2 泛TE变异图谱的构建
Fig. 2. Construction of pan-TE variation map

2-B)。

2.2 泛 TE 图谱中的 TE 变异特征

该 TE 图谱共包含 17,097 个 TE 变异, 覆盖基因组总长度 39.1 Mb(大约是日本晴基因组的 10%)(图 3-A), 其中, TE 插入的总数(9,455)和总长度(24.3 Mb)均超过 TE 缺失的总数(7,642)和总长度(14.7 Mb)(图 3-A)。值得注意的是, 36.9% TE 插入和 21.3 % TE 缺失为私有变异, 即仅存在于少数几个材料中(频率<0.05)(图 3-B)。超过 3/5 的 TE 变异较长(即 61.7%的 TE 变异长度超过 1 kb, 图 3-C), 并且 11.1%的 TE 变异包含至少两个 TE 序列。

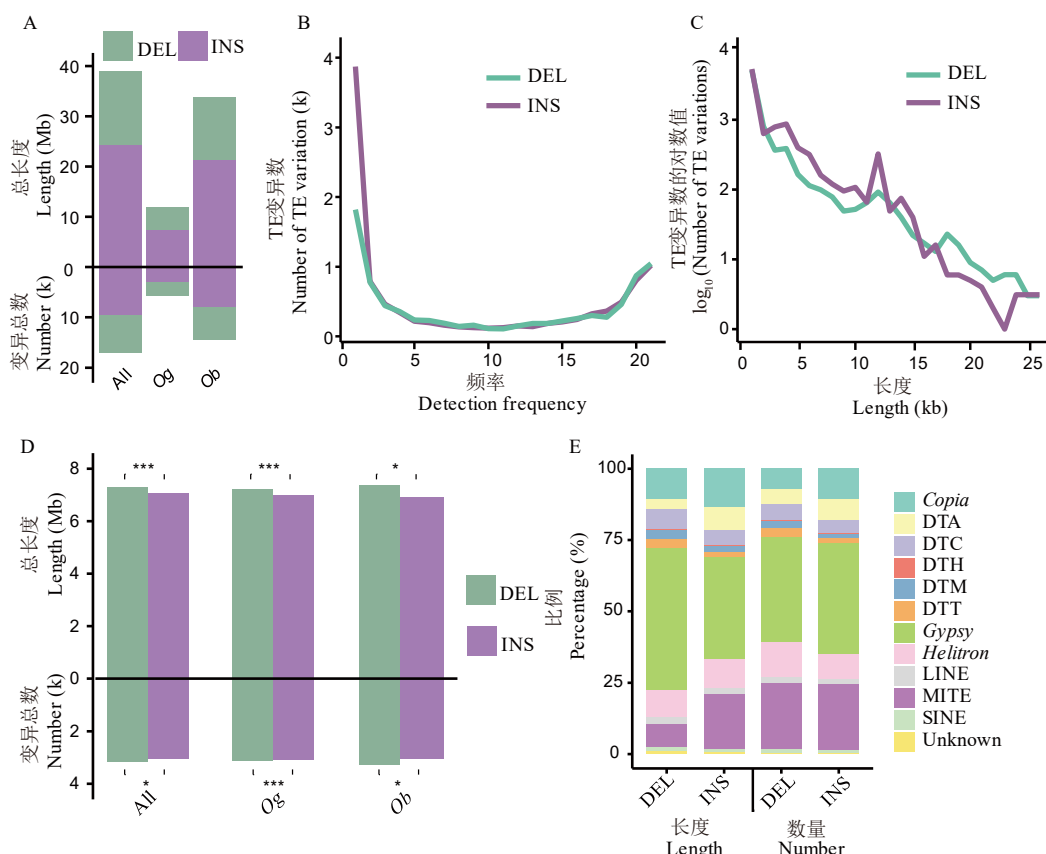
基于该泛 TE 变异图谱比较物种间的 TE 变异, 我们发现 *Ob* 中 TE 变异总数(14,357)和总长度(33.8 Mb)均高于 *Og*(分别为 5,582 和 11.8 Mb)(图 3-A)。为了排除群体大小不同的影响, 我们进一步分析了每个材料中 TE 变异的总数和总长度, 发现 *Ob* 群

体中单个样品的 TE 变异总数和总长度均高于 *Og*(图 3-D)。

进一步分析每个 TE 家族中的 TE 变异数目, 发现该泛 TE 变异图谱中, 主要为 *Gypsy*(40.4%)、*MITE*(18.5%)和 *Helitron*(10.1%)家族(图 3-E), 这与各基因组中这些家族的高比例一致^[1]。

2.3 TE 变异在全基因组中的分布

根据泛 TE 变异图谱中 TE 变异与全基因组基因的相对位置, 我们计算了每个基因附近(基因区域 ± 2 kb 内)的 TE 变异数量, 结果显示, 28.1%的基因在所有材料中均为保守(即未发现附近存在 TE 变异), 而约 3.8%的基因至少含有两个 TE 变异。另外, 76.5%的 TE 变异与基因及附近区域(基因区域 ± 2 kb 内)存在重叠(图 4)。与参考基因组日本晴^[1]中 TE 序列的分布相比, 该泛 TE 变异图谱中的 TE 变异在转录调控区显著富集, 尤其是 DNA TE(包括

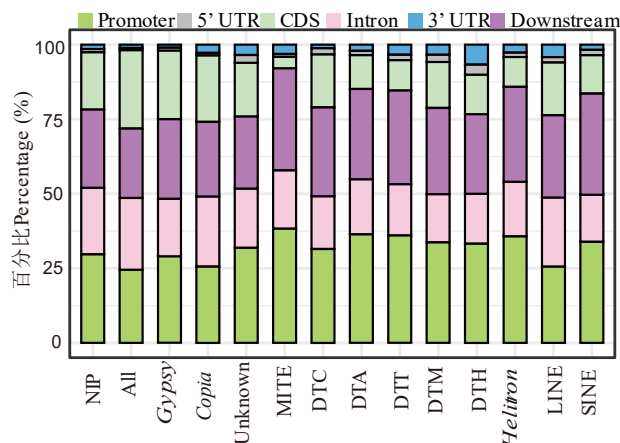


A: 不同群体中鉴定到的非冗余的 TE 变异总数和总长度, “All”指本研究中所有材料; B: 21 个材料构建的泛 TE 图谱中每个 TE 变异的频率; C: 泛 TE 图谱中 TE 变异的长度分布。D: 不同亚群中每个材料的 TE 变异平均长度和平均数目。*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ (t 检验); E: 泛 TE 图谱中 TE 变异的百分比, 未知指 LTR/未知家族。

A, Total length and number of non-redundant TE variations and sequences detected across different subpopulations. All refers to all the accessions in the present study; B, Frequencies of each TE variation in 21 African accessions; C, Length distributions of TE variations in the non-redundant TE dataset for African accessions. D, Average length and number of TE variations for each accession across different subpopulations. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$ (t -test); E, Percentage of TE variations in the non-redundant TE dataset for African accessions. Unknown refers to LTR/unknown family.

图 3 泛 TE 变异图谱中 TE 变异特征

Fig. 3. Characteristics of TE variation in pan-TE variation map



基因及其附近区域包括 CDS、3'UTR、5'UTR 和内含子，以及上游 2 kb（以下称为“启动子”）和下游 2 kb（以下称为“下游”）的区域。未知指的是 LTR/未知家族。“All”指的是本研究中总 TE 变异图谱中所有 TE 变异中与基因及其附近区域内重叠的百分比，“NIP”指的是日本晴基因组中 TE 序列与基因及其附近区域重叠的百分比。

Genic regions include gene body (CDS, 3'UTR, 5'UTR, and intron), and the regions within 2 kb upstream (henceforth termed “promoter”) and 2 kb downstream (henceforth termed “downstream”) of the gene body. Unknown refers to LTR/unknown family. “All” refers to the percentage of TE variations in the genic regions of all TE loci identified in the present study, and “NIP” refers the percentage of TEs overlapping with genic regions in the Nipponbare genome.

图 4 TE 变异与基因及邻近区域重叠的百分比

Fig. 4. Percentage of TE variations overlapping with genic regions

MITE 和 *Helitron* 家族)，而在编码区(CD)和内含子区(Intron)的分布则显著降低(图 4)。这些结果表明，TE 变异可能在调控基因表达中发挥重要作用。

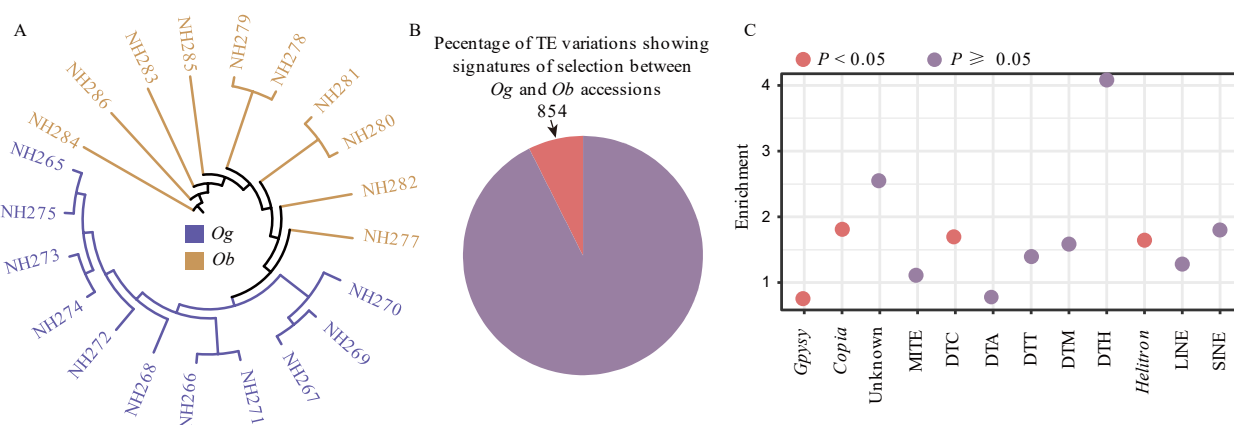
2.4 TE 变异影响 *Og* 的驯化过程

基于泛 TE 图谱中的变异数据，我们构建了 *Ob*

和 *Og* 的系统进化树，结果显示这两个物种能够被清晰区分(图 5-A)。为发掘 *Og* 驯化过程中受到选择的 TE 变异，我们用 F_{ST} 分析比较了 *Og* 和 *Ob* 中 TE 变异的基因型分布，并对 F_{ST} 值进行排序，选择前 5% F_{ST} 值作为选择的 TE 变异($F_{ST} = 0.64$)，最终在 *Og* 驯化过程中共鉴定到 854 个受选择 TE 变异(图 5-B)。对这些受选择的 TE 变异进行 Fisher 检验后，我们发现 *Copia*、DTC、*Helitron* 家族在 *Og* 的驯化过程中显著富集(图 5-C)，且这三个家族也在亚洲栽培稻的驯化过程中显著富集^[6]，这表明 TE 变异可能在亚洲栽培稻和 *Og* 的平行驯化过程中发挥了重要作用。另外，我们也发现 *Gypsy* 家族在 *Og* 的驯化过程中没有显著富集($P > 0.05$ ，图 5-C)，这与亚洲栽培稻的驯化过程不一致^[6]。

为深入探究受选择 TE 变异对 *Og* 驯化过程中基因功能的影响，我们详细分析了包含这些受选择 TE 变异的基因及其邻近区域(基因及其上下游 2 kb 范围内)，最终，共鉴定出 658 个含有受选择 TE 变异的基因，富集分析结果显示，这些基因显著富集于与基础代谢相关的生物过程($P < 0.05$ ，图 6)。值得注意的是，在 *Og* 和亚洲栽培稻驯化过程中，共有 42 个含受选择 TE 变异的基因是相同的，占 *Og* 驯化过程中含选择性 TE 变异总基因的 6.3%，这些基因包括 *Snf1* 相关蛋白激酶基因 *SnRK1A*^[20]、钙调蛋白基因 *OsCML14*^[21] 和盐胁迫基因 *OsSOS1*^[22]。

与亚洲稻栽培稻的驯化过程相比，*Og* 驯化过程中独有的含受选择 TE 变异的基因有 616 个，这表明大部分含选择性 TE 变异的基因(93.6%)可能参

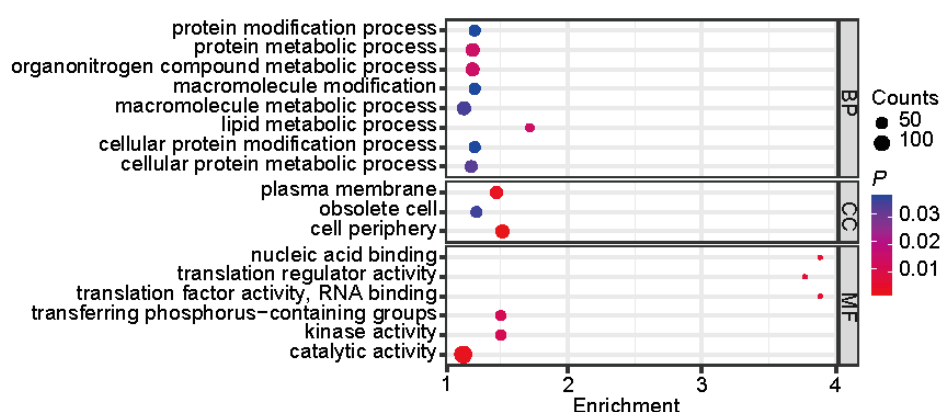


A: 基于总的泛 TE 变异图谱构建的系统发生树；B: 在 *Og* 驯化过程中受选择 TE 变异的比率，根据 TE 变异的 F_{ST} 值被认为受选择 TE 变异；C: 根据 Fisher 检验对每个 TE 家族中受选择的 TE 变异进行富集分析，Unknown 指 LTR/unknown 家族。

A, Phylogeny of 21 African accessions based on pan-TE genotypes; B, Percentage of TE variations showing signatures of selection during *Og* domestication. F_{ST} outlier based on TE variations were indicated as loci showing signatures of selection; C, Fold enrichment of selected TE variations in each TE family by the Fisher's exact test. Unknown refers to LTR/unknown family.

图 5 TE 对非洲栽培稻驯化过程的影响

Fig. 5. Contribution of TE variations to *Og* domestication

图6 *Og* 驯化过程含受选择 TE 变异的基因富集分析Fig. 6. Enrichment analysis of genes containing selected TE variations during the domestication process of *Og* accessions

与了 *Og* 的独立驯化过程, 包括 30 个已知功能基因 (图 7), 其中 12 个基因与重要农艺性状(包括粒型、每穗粒数、结实率、产量等)相关, 14 个基因与生物胁迫(白叶枯病、稻瘟病等抗性)和非生物胁迫(盐、温度等)响应相关。

例如, 穗发育的基因 *OsYABBY4* 的启动子区域存在一个 262 bp 的 TE(*Stowaway* MITE)插入。该 TE 插入存在于所有的 *Og* 中($n=11$), 但仅存在于少数 *Ob*(4/10)材料中, 亚洲栽培稻、普通野生稻(*O. rufipogon*)以及外群的稻属基因组中均无该 TE 插入 (图 8-A), 这一发现进一步证明了 TE 变异在 *Og* 驯化过程中起重要作用。

另外, 前期研究已报道耐盐性可能是 *Og* 适应当地环境的关键性状^[2]。在本研究中, 我们也发现 *Og* 驯化过程中存在 9 个与盐胁迫相关的已知功能基因。例如, 高亲和性钾离子转运蛋白基因 *OsHKT2;4*(*LOC_Os06g48800*)的启动子区域存在

一个 1.0 kb TE(*Helitron*)缺失(图 8-B), 该 TE 缺失存在于所有 *Og* 中, 但仅存在于少数 *Ob*(1/10)中, 并且该 TE 缺失存在于亚洲栽培稻和普通野生稻中, 但在 AA、BB 和 EE 类外群基因组中不存在。这一结果表明, 该 TE 缺失可能来源于 *Ob*, 并经过驯化或人工选择过程保留下来, 成为影响 *Og* 耐盐性的重要遗传变异。

3 讨论

水稻作为全球数十亿人口的主要粮食作物, 深入解析其表型变异及驯化过程对农业、世界粮食安全以及生物和基因组研究具有重要意义^[23-24]。TE 在水稻基因组中广泛存在, 约占基因组序列的 51%, 其中主要以 LTR(尤其是 *Gypsy*)元件为主^[1]。已有研究表明, TE 主要富集于基因及附近区域, 可能通过改变基因的表达量进而影响水稻表型^[6],

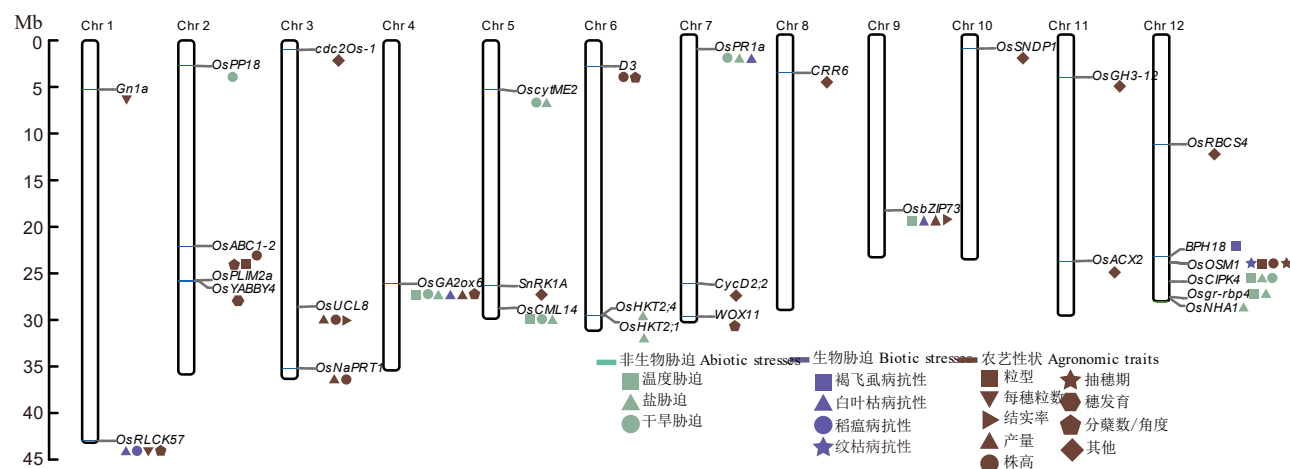
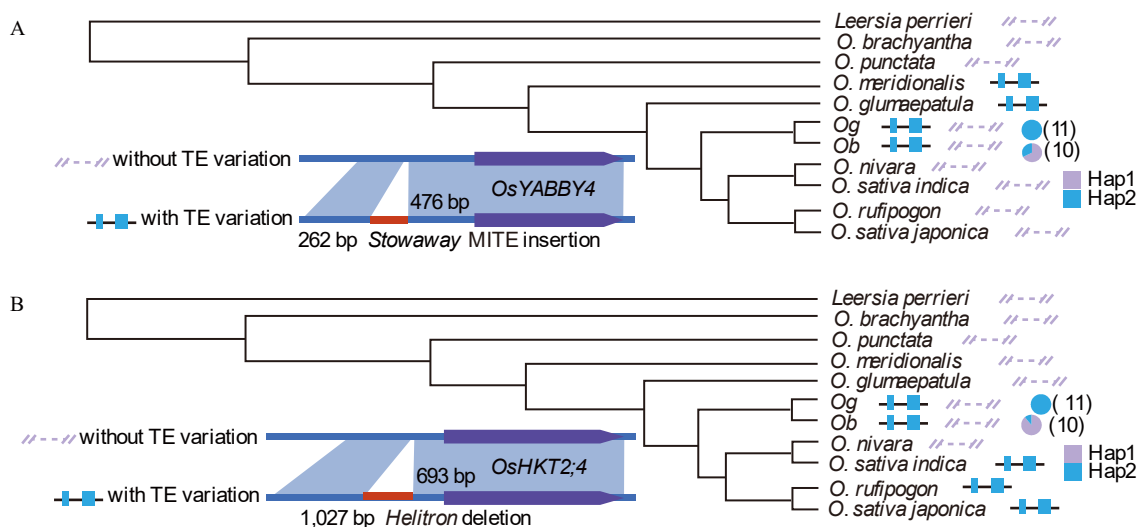


图7 水稻 12 条染色体上含受选择 TE 变异的功能基因

Fig. 7. Positions of functional genes harboring selected TE variations in their genomic regions on chromosomes 1-12



A-B: 在所有 *Og* 样品中发现穗发育基因 *OsYABBY4* 启动子区域中的 Stowaway MITE 插入 (A) 和盐胁迫基因 *OsHKT2;4* 启动子区域中的 Helitron 缺失 (B), 但在少数 *Ob* 样品中未发现。

A-B, Stowaway MITE insertion in the *OsYABBY4* promoter region (A) and Helitron deletion in the *OsHKT2;4* promoter region were found in all *Og* accessions except a few *Ob* accessions(B).

图 8 非洲栽培稻驯化过程中已知功能基因中的 TE 变异

Fig. 8. TE variations in known functional genes during *Og* domestication

这与其他植物中的研究结果一致^[25], 说明 TE 在植物基因组进化中的重要性。构建物种的泛基因组有助于解析 TE 变异所隐藏的基因组复杂性, 并重新评估它们在基因组进化和表型变异中的作用。然而, 前期的研究尚未构建 *Og* 和 *Ob* 的泛 TE 变异图谱^[6]。本研究利用 11 份 *Og* 和 10 份 *Ob* 的染色体级别的基因组序列, 构建了高质量的泛 TE 变异图谱。通过分析 *Og* 驯化过程中 TE 变异的变化, 我们鉴定了 854 个受选择的 TE 变异, 这些变异可能参与了从 *Ob* 到 *Og* 的驯化过程。在泛 TE 图谱中, 我们还发现了一些包含两个或更多 TE 序列的复合 TE 变异(如由 *Gypsy*、*Helitron* 和 *Copia* 元件组成的 TE 变异), 这些 TE 变异也可能对表型多样性有重要影响。如果有 *Og* 和 *Ob* 群体的转录组和表型的数据, 将更深入地揭示 TE 对 *Og* 驯化过程的影响。

通过比较 *Og* 和 *Ob* 的 TE 分布, 我们鉴定出一些驯化相关的 TE 变异, 这些 TE 变异很大程度上改变基因的表达量, 进一步影响水稻的驯化性状(株型、产量、抗性)。TE 在 *Og* 驯化过程中对表型变异的贡献是多纬度的, 包括改变基因表达、修饰蛋白质功能以及诱导表观遗传变化等。例如, TE 插入基因调控区域(如启动子、增强子等)可能破坏或创建新的调控元件, 从而导致基因表达水平的变化。这一点在玉米中得到了验证, 例如 TE 插入开花时间基因 *Vgt1* 附近与开花表型的变异密切相关^[26]。此外,

TE 变异还可以通过介导结构变异(如插入、缺失或染色体重排)来影响基因剂量或产生新的基因融合, 从而促进表型多样性^[27]。进一步地, TE 变异可以通过引入被 DNA 甲基化或组蛋白修饰机制靶向的序列来影响表观遗传景观, 从而调控染色质状态和基因表达^[28]。一些有利于农业生产的 TE 变异可能在驯化过程中被人类选择并固定保留下来, 从而形成 *Og* 独特的表型特征。TE 可能通过快速产生遗传变异, 帮助水稻适应不同的生境条件, 如温度、耐盐、病虫害等^[29]。本研究发现, 一些 TE 插入位于耐盐等逆境的相关基因中, 这可能增强了 *Og* 的环境适应性。这一发现为理解作物适应性进化提供了新的视角, 表明 TE 变异在作物遗传改良中具有潜在的应用价值。除了 SNP 和 TE, 其他遗传变异可能也会影响 *Og* 驯化过程, 例如倒位^[30]、串联重复^[31]及稀有变异^[32]。尽管本研究的样本量相对有限, 但样本选择具有代表性, 能够较好地反映 *Og* 和 *Ob* 群体的 TE 变异特征, 这一局限性对结果的整体可靠性影响较小。

综上, 本研究系统解析了 TE 对 *Og* 驯化过程的贡献, 为作物适应性进化的研究开辟了新的方向。利用 TE 介导的遗传变异, 可以加速水稻遗传改良的育种进程。未来的研究可进一步探索 TE 功能的具体机制, 并将其应用于其他作物的遗传改良中。另外, 本研究构建的泛 TE 变异图谱为研究人

员提供了一个高效平台, 使他们能够通过少数样本获得更多的变异信息, 这对于通过基因组编辑技术成功定位高度复杂或重复区域内的因果变异具有重要意义。

参考文献:

- [1] Shang L G, Li X X, He H Y, Yuan Q L, Song Y N, Wei Z R, Lin H, Hu M, Zhao F L, Zhang C, Li Y H, Gao H S, Wang T Y, Liu X P, Zhang H, Zhang Y, Cao S M, Yu X M, Zhang B T, Zhang Y, Tan Y Q, Qin M, Ai C, Yang Y X, Zhang B, Hu Z Q, Wang H R, Lü Y, Wang Y X, Ma J, Wang Q, Lu H W, Wu Z, Liu S L, Sun Z Y, Zhang H L, Guo L B, Li Z C, Zhou Y F, Li J Y, Zhu Z F, Xiong G S, Ruan J, Qian Q. A super pan-genomic landscape of rice[J]. *Cell Research*, 2022, 32(10): 878-896.
- [2] Meyer R S, Choi J Y, Sanches M, Plessis A, Flowers J M, Amas J, Dorph K, Barretto A, Gross B, Fuller D Q, Bimpong I K, Ndjiondjop M N, Hazzouri K M, Gregorio G B, Purugganan M D. Domestication history and geographical adaptation inferred from a SNP map of African rice[J]. *Nature Genetics*, 2016, 48(9): 1083-1088.
- [3] Feschotte C, Jiang N, Wessler S R. Plant transposable elements: Where genetics meets genomics[J]. *Nature Reviews Genetics*, 2002, 3(5): 329-341.
- [4] McClintock B. Chromosome organization and genic expression[J]. *Cold Spring Harbor Symposia on Quantitative Biology*, 1951, 16: 13-47.
- [5] Wicker T, Sabot F, Hua-Van A, Bennetzen J L, Capy P, Chalhoub B, Flavel A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman A H. A unified classification system for eukaryotic transposable elements[J]. *Nature Reviews Genetics*, 2007, 8(12): 973-982.
- [6] Li X X, Dai X F, He H Y, Lü Y, Yang L B, He W C, Liu C C, Wei H, Liu X P, Yuan Q L, Wang X M, Wang T Y, Zhang B T, Zhang H, Chen W, Leng Y, Yu X M, Qian H G, Zhang B, Guo M L, Zhang Z P, Shi C L, Zhang Q Q, Cui Y, Xu Q, Cao X L, Chen D D, Zhou Y F, Qian Q, Shang L G. A pan-TE map highlights transposable elements underlying domestication and agronomic traits in Asian rice[J]. *National Science Review*, 2024, 11(6): nwae188.
- [7] Gao Y H, Xu Z P, Zhang L J, Li S C, Wang S G, Yang H L, Liu X L, Zeng D L, Liu Q Q, Qian Q, Zhang B C, Zhou Y H. MYB61 is regulated by GRF4 and promotes nitrogen utilization and biomass production in rice[J]. *Nature Communications*, 2020, 11(1): 5219.
- [8] Wu B, Meng J H, Liu H B, Mao D H, Yin H R, Zhang Z Y, Zhou X C, Zhang B, Sherif A, Liu H Y, Li X H, Xiao J H, Yan W H, Wang L, Li X W, Chen W, Xie W B, Yin P, Zhang Q F, Xing Y Z. Suppressing a phosphohydrolase of cytokinin nucleotide enhances grain yield in rice[J]. *Nature Genetics*, 2023, 55(8): 1381-1389.
- [9] Mao D H, Tao S T, Li X, Gao D Y, Tang M F, Liu C B, Wu D, Bai L L, He Z K, Wang X D, Yang L, Zhu Y X, Zhang D C, Zhang W L, Chen C Y. The Harbinger transposon-derived gene *PANDA* epigenetically coordinates panicle number and grain size in rice[J]. *Plant Biotechnology Journal*, 2022, 20(6): 1154-1166.
- [10] Chang L Q, Hu M, Ning J, He W, Gao J Y, Ndjiondjop M N, Fu Y C, Liu F X, Sun H Y, Gu P, Sun C Q, Zhu Z F. The genetic control of glabrous glume during African rice domestication[J]. *Journal of Genetics and Genomics*, 2022, 49(5): 427-436.
- [11] Shang L G, He W C, Wang T Y, Yang Y X, Xu Q, Zhao X J, Yang L B, Zhang H, Li X X, Lü Y, Chen W, Cao S, Wang X M, Zhang B, Liu X P, Yu X M, He H Y, Wei H, Leng Y, Shi C L, Guo M L, Zhang Z P, Zhang B T, Yuan Q L, Qian H G, Cao X L, Cui Y, Zhang Q Q, Dai X F, Liu C C, Guo L B, Zhou Y F, Zheng X M, Ruan J, Cheng Z K, Pan W H, Qian Q. A complete assembly of the rice Nipponbare reference genome[J]. *Molecular Plant*, 2023, 16(8): 1232-1236.
- [12] Lü Y, Liu C C, Li X X, Wang Y Y, He H Y, He W C, Chen W, Yang L B, Dai X F, Cao X L, Yu X M, Liu J J, Zhang B, Wei H, Zhang H, Qian H G, Shi C L, Leng Y, Liu X, Guo M L, Wang X M, Zhang Z P, Wang T Y, Zhang B T, Xu Q, Cui Y, Zhang Q Q, Yuan Q L, Jahan N, Ma J, Zheng X M, Zhou Y F, Qian Q, Guo L B, Shang L G. A centromere map based on super pan-genome highlights the structure and function of rice centromeres[J]. *Journal of Integrative Plant Biology*, 2024, 66(2): 196-207.
- [13] Li H. Minimap2: Pairwise alignment for nucleotide sequences[J]. *Bioinformatics*, 2018, 34(18): 3094-3100.
- [14] Kent W J, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(20): 11484-11489.
- [15] Lecompte L, Peterlongo P, Lavenier D, Lemaitre C. SVJedi: Genotyping structural variations with long reads[J]. *Bioinformatics*, 2020, 36(17): 4568-4575.
- [16] Chang C C, Chow C C, Tellier L C, Vattikuti S, Purcell S M, Lee J J. Second-generation PLINK: Rising to the challenge of larger and richer datasets[J]. *Gigascience*, 2015, 4: 7.
- [17] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals[J]. *Genome Research*, 2009, 19(9): 1655-1664.
- [18] Price M N, Dehal P S, Arkin A P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix[J]. *Molecular Biology and Evolution*, 2009, 26(7): 1641-1650.
- [19] Danecek P, Auton A, Abecasis G, Albers C A, Banks E, DePristo M A, Handsaker R E, Lunter G, Marth G T, Sherry S T, McVean G, Durbin R. Price M N, Dehal P S,

- Arkin A P. The variant call format and VCFtools[J]. *Bioinformatics*, 2011, 27(15): 2156-2158.
- [20] Lu C A, Lin C C, Lee K W, Chen J L, Huang L F, Ho S L, Liu H J, Hsing Y I, Yu S M. The SnRK1A protein kinase plays a key role in sugar signaling during germination and seedling growth of rice[J]. *The Plant Cell*, 2007, 19(8): 2484-2499.
- [21] Boonburapong B, Buaboocha T. Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins[J]. *BMC Plant Biology*, 2007, 7(1): 4.
- [22] Zhou G A, Jiang Y, Yang Q, Wang J F, Huang J, Zhang H S. Isolation and characterization of a new Na⁺/H⁺ antiporter gene *OsNHAI* from rice (*Oryza sativa* L.)[J]. *DNA Sequence*, 2006, 17(1): 24-30.
- [23] Wang W S, Mauleon R, Hu Z Q, Chebotarov D, Tai S S, Wu Z C, Li M, Zheng T Q, Fuentes R R, Zhang F, Mansueto L, Copetti D, Sanciango M, Palis K C, Xu J L, Sun C, Fu B Y, Zhang H L, Gao Y M, Zhao X Q, Shen F, Cui X, Yu H, Li Z C, Chen M L, Detras J, Zhou Y L, Zhang X Y, Zhao Y, Kudrna D, Wang C C, Li R, Jia B, Lu J Y, He X C, Dong Z T, Xu J B, Li Y H, Wang M, Shi J, Li J X, Li J, Zhang D B, Lee S, Hu W S, Poliakov A, Dubchak I, Ulat V J, Borja F N, Mendoza J R, Ali J, Li J, Gao Q, Niu Y C, Yue Z, Naredo M E B, Talag J, Wang X Q, Li J J, Fang X D, Yin Y, Glaszmann J C, Zhang J, Li J W, Hamilton R S, Wing R A, Ruan J, Zhang G Y, Wei C C, Alexandrov N, McNally K L, Li Z K, Leung H. Genomic variation in 3,010 diverse accessions of Asian cultivated rice[J]. *Nature*, 2018, 557(7703): 43-49.
- [24] He W C, Li X X, Qian Q, Shang L G. The developments and prospects of plant super-pangenomes: Demands, approaches, and applications[J]. *Plant Communications*, 2025, 6(2): 101230.
- [25] Cai X, Lin R M, Liang J L, King G J, Wu J, Wang X W. Transposable element insertion: A hidden major source of domesticated phenotypic variation in *Brassica rapa*[J]. *Plant Biotechnology Journal*, 2022, 20(7): 1298-1310.
- [26] Lisch D. How important are transposons for plant evolution[J]? *Nature Reviews Genetics*, 2013, 14(1): 49-61.
- [27] Chuong E, Elde N C, Feschotte C. Regulatory activities of transposable elements: From conflicts to benefits[J]. *Nature Reviews Genetics*, 2017, 18(2): 71-86.
- [28] Slotkin R K, Martienssen R. Transposable elements and the epigenetic regulation of the genome[J]. *Nature Reviews Genetics*, 2007, 8(4): 272-285.
- [29] Wei H, Wang X M, Zhang Z P, Yang L B, Zhang Q Q, Li Y L, He H Y, Chen D D, Zhang B, Zheng C, Leng Y, Cao X L, Cui Y, Shi C L, Liu Y F, Lü Y, Ma J, He W C, Liu X P, Xu Q, Yuan Q L, Yu X M, Wang T Y, Qian H G, Li X X, Zhang B T, Zhang H, Chen W, Guo M L, Dai X F, Wang Y X, Zheng X M, Guo L B, Xie X Z, Qian Q, Shang L G. Uncovering key salt-tolerant regulators through a combined eQTL and GWAS analysis using the super pan-genome in rice[J]. *National Science Review*, 2024, 11(4): nwae043.
- [30] He W C, He H Y, Yuan Q L, Zhang H, Li X X, Wang T Y, Yang Y X, Yang L B, Yang Y T, Liu X P, Wei H, Zhang H, Zhang B, Guo M L, Leng Y, Shi C L, Lü Y, Chen W, Wang X M, Zhang Z P, Yu B H, Zhang B T, Xu Q, Qian H G, Zhou Y F, Wang S K, Qian Q, Shang L G. Widespread inversions shape the genetic and phenotypic diversity in rice[J]. *Science Bulletin*, 2024, 69(5): 593-596.
- [31] He H Y, Leng Y, Cao X L, Zhu Y W, Li X X, Yuan Q L, Zhang B, He W C, Wei H, Liu X P, Xu Q, Guo M L, Zhang H, Yang L B, Lü Y, Wang X M, Shi C L, Zhang Z P, Chen W, Zhang B T, Wang T Y, Yu X M, Qian H G, Zhang Q Q, Dai X F, Liu C C, Cui Y, Wang Y X, Zheng X M, Xiong G S, Zhou Y F, Qian Q, Shang L G. The pan-tandem repeat map highlights multiallelic variants underlying gene expression and agronomic traits in rice[J]. *Nature Communications*, 2024, 15(1): 7291.
- [32] Wang T Y, He W C, Li X X, Zhang C, He H Y, Yuan Q L, Zhang B, Zhang H, Leng Y, Wei H, Xu Q, Shi C L, Liu X P, Guo M L, Wang X M, Chen W, Zhang Z P, Yang L B, Lv Y, Qian H G, Zhang B T, Yu X M, Liu C C, Cao X L, Cui Y, Zhang Q Q, Dai X, Guo L B, Wang Y X, Zhou Y F, Ruan J, Qian Q, Shang L G. A rice variation map derived from 10548 rice accessions reveals the importance of rare variants[J]. *Nucleic Acids Research*, 2023, 51(20): 10924-10933.