基因组测序揭示结构变异对甘蓝多样性的贡献

Genome sequencing sheds light on the contribution of structural variants to B*rassica oleracea* diversification

发表杂志：bioRxiv

发表单位：北京农林科学院

发表时间：2020 年 10 月

摘要

甘蓝包括几种形态多样、具有重要经济价值的蔬菜作物。我们展示了花椰菜和卷心菜（结球甘蓝）两种形态的甘蓝的高质量染色体规模的基因组组装结果，通过两者比较确定了约 120K 的高置信度结构变异(SVs)。利用这些 SVs 对 271 份甘蓝种质进行群体分析，可以明显区分出不同的形态类型，表明 SVs 与甘蓝种群分化有关。研究表明，在花椰菜和甘蓝之间受选择 SVs 影响的基因在胁迫和刺激反应、分生组织及花发育相关功能富集。此外，还鉴定了受选择的 SVs 影响并参与从营养生长到生殖性生长的转换的基因，这些基因定义了花球的起始、花序分生组织增殖和花球的形成、维持和扩大，为了解花球发育的调控网络奠定了基础。本研究揭示了 SVs 在甘蓝不同形态类型多样性中的重要作用，新组装的基因组和 SVs 为今后的研究和育种提供了丰富的资源。

研究背景

甘蓝包括几种不同的优势蔬菜作物，2018 年全球总产量近 1 亿吨(http://www.fao.org/faostat).)。该物种的极端多样性是独一无二的，它的形态类型是为了扩大代表收获产品的不同器官而选择的，例如，花椰菜(*B. oleracea var. botrytis*)和西兰花（*B. oleracea var. italica*）的花序、结球甘蓝(*B. oleracea var. capitata*)的头状花序(顶芽)、抱子甘蓝（*B. oleracea var. gemmifera*）的侧生叶芽、芥蓝(*B. oleracea var. alboglabra*)的叶子、擘蓝(*B. oleracea var. gongylodes*)的块茎。在过去的几年中，已经为不同形态的甘蓝属植物生成了参考基因组序列，包括芥蓝、结球甘蓝、花椰菜和西兰花等，这些基因组序列极大地方便了

遗传变异分析，从而更好地了解甘蓝的遗传多样性、种群结构以及进化和驯化。

结构变异(SVs)，包括插入、缺失、重复和易位，在植物基因组中大量存在，比单核苷酸多态性(SNPs)更有可能引起表型变化。许多 SVs 已被鉴定为不同作物重要农艺性状的因果遗传变异，例如花椰菜中 *Or* 基因第三个外显子中 4.7kb 插入导致橙色花球；甘蓝型油菜中 *BnaA9.CYP78A9* 上游区域的 3.7kb 插入导致的长角果和大种子；油菜中 *BnaFLC.A10* 启动子区域 621bp 插入有助于油菜对冬季栽培环境的适应。 以前对甘蓝属植物的全基因组变异分析主要集中在 SNPs 和小 Indels，而基因组 SVs 在很大程度上被忽视，这主要是由于在遗传变异鉴定中使用短序列读数的限制。通过将短测序读数映射到参考基因组来进行 SV 检测容易导致高水平的假阴性和假阳性，特别是对于高度重复的植物基因组，如甘蓝的基因组。因此，到目前为止，甘蓝不同形态类型中 SVs 的种群动态仍然很大程度上没有被研究。

最近，通过直接比较高质量的染色体水平的基因组组装结果和/或将使用 PacBio 或 Nanopore 测序技术产生的长读数映射到参考基因组的方法已被证明对于大型和复杂植物基因组中的 SV 检测是非常准确的。在这项研究中，我们使用 PacBio 长读取技术和高通量染色体构象捕获(Hi-C)技术，为花椰菜和结球甘蓝产生了高质量的染色体规模的基因组组装。通过基因组直接比对和长读长比对，在这两个基因组之间共鉴定出 119,156 个高置信度 SVs。我们进一步收集 271 份（本研究中产生的 163 份甘蓝种质和先前报道的 108 份甘蓝种质）不同形态类型甘蓝种质的基因组重测序数据，并利用这些数据对这些种质的 119,156 个高置信度 SVs 进行基因分型。对甘蓝不同形态类型的等位基因频率进行研究，主要比较花椰菜和结球甘蓝群体间的等位基因频率，结合基因表达分析，我们证实了 SVs 对花椰菜花球形成的调控作用。

研究方法

基因组：

样本情况：

花椰菜(*B. oleracea var. botrytis*)样本 Korso_1401 是从 IPK Gatersleben 基因库获得的来自 Korso 的高度自交系，花球白色致密，成熟时间长(>95d)。

结球甘蓝（*B. oleracea var. capitate*）样本 OX-heart_923，来自江苏省农业科学院蔬菜研

究所的绿色尖头晚抽薹自交系。

测序策略：

PacBio 测序，Korso 样本共 24 个 cell，9 个 PacBio RSII 平台，15 个 PacBio Sequel 平台）；OX-heart 样本 15 个 cell（PacBio Sequel 平台）。

Illumina 测序，HiSeq 2500 平台，PE150 测序策略。

Hi-C 测序，Illumina HiSeq X Ten 平台，PE150 测序策略。

BioNano 光学图谱，Saphyr 平台。

重测序：

样本情况：

163 份甘蓝属不同种质(花椰菜 89 份，甘蓝 65 份，花椰菜 9 份)

测序策略：

Illumina HiSeq 2500 平台，PE150 测序策略(117 份)或 PE100 测序策略 (46 份)。

转录组：

样本情况：

采集 Korso 样本（根、茎、叶、球、芽、花和角果）和 OX-heart 样本（根、茎、叶、头状花序、芽、花和角果）7 个不同组织，进行了转录组测序。此外，还采集了 Korso 样本处于以下发展阶段的顶端分生组织(SAM)：营养、过渡(花球起始)、花球形成(花球直径~1 厘米)、早熟(花球直径 10 厘米)和枝条伸长(成熟)。每个样本进行两到三个生物学重复。

测序策略：

二代转录组， Korso 样本和 OX-heart 样本，Illumina HiSeq X Ten；

三代转录组， Korso 样本，PacBio Sequel 平台。

研究结果

1.花椰菜和结球甘蓝基因组的从头组装

对花椰菜自交系样本 Korso_1401（以下简称 Korso）和结球甘蓝样本 OX-heart_923（以下简称 OX-heart）进行基因组测序。每个样本均产生约 70.0 Gb 的 PacBio 序列，覆盖约 120 倍的 Korso 和 OX-heart 基因组，其估计大小分别为 566.9 Mb 和 587.7 Mb。用 PacBio 序从头组装成 Contigs，使用 PacBio 长序列和 Illumina 短序列(每个样本约 100 Gb)校正组装的 Contigs 中的错误。用 242.2 GB 的 BioNano 光学图谱数据辅助 Korso 基因组组装。用 Hi-C 数据分别对 Korso 和 OX-heart 进行辅助组装。最终组装得到的 Korso 和 OX-heart 的基因组 Contig 个数分别为 615 个和 973 个，累积长度分别为 549.7 Mb 和 565.4 Mb，N50 大小分别为 4.97Mb 和 3.10Mb。结果表明，Korso、OX-heart 与西兰花 HDEM 组装结果的 Hi-C 热图之间有良好一致性。

对组装质量进行评估，结果发现，约 99.8%的 Illumina 基因组数据可以被映射回组装的基因组上，98.0%的 RNA-Seq 读数可以被映射回组装的基因组上。BUSCO 结果表明，97.2%和 96.5%的核心保守植物基因在 Korso 和 OX-heart 的基因组中被完整组装。总而言之，这些结果证明了 Korso 和 OX-heart 基因组组装的质量较高。

2.基因组注释与比较基因组

约 60.7%的 Korso 和 62.0%的 OX-heart 被注释为重复元件，其中 *Gypsy* 和 *Copia* 反转录转座子代表了这两个基因组中最丰富的家族。从 Korso 和 OX-heart 和 *B.rapa* (V3.0)(基因组中提取全长长末端重复反转录转座子(LTR-RTS)，对这些完整 LTR-RT 的插入时间估计揭开了分别发生在大约 20 万年和 150 万年前的 Korso 和 OX-heart 两个 LTR-RT 的爆发。相反，在 *B. rapa* 中，大多数 LRT-RT 是最近形成的，超过 30%的完整 LTR-RT 小于 20 万年，而在 Korso 和 OX-heart 中，这一比例分别为 16.3%和 15.9%。

高质量的 Korso 和 OX-heart 组装结果使我们能够精确识别着丝粒位置。在两个基因组中确定的着丝粒在每条染色体上的位置与之前用荧光原位杂交(FISH)分析确定的着丝粒位置一致。不出所料，重复序列在着丝粒区域得到了丰富。不同的重复元件家族在染色体上表现出明显不同的模式，如 *Copia* 类型的 LTR 主要位于着丝粒上，而 *Gypsy* 类型的 LTR 主要

位于着丝粒周围区域(图 1A)。

使用结合了从头预测，基于转录本和基于同源性的预测的综合策略，分别从 Korso 和 OX-heart 基因组预测了总共 60,640 和 62,232 个高可信度蛋白质编码基因。在这些预测基因中，转录组证据支持了 70.9%和 76.4%，其他植物中有 91.0%和 90.0%具有同源物。BUSCO 显示 Korso 和 OX-heart 预测基因的完整性分别为 96.1%和 94.9%。

对 Korso、OX-heart、*B. rapa* 和 *A. thaliana* 基因组的共线性分析证实了芸苔属植物的全基因组三倍化事件(WGT)和随后的亚基因组分化事件。基于这些共线关系，我们确定了 Korso 和 OX-heart 基因组中的三倍型区域，并根据其保留的基因密度将它们划分为三个亚基因组(图 1)。正如先前研究中报道的一致，Korso 和 OX-heart 的分为三个亚基因组，LF(最不分离的)，MF1(中等分离的)和 MF2(最分离的)，在二倍化过程中显示出相同的重复基因的偏向保留模式。不同亚基因组中的重复基因拷贝显示出不同的基因表达模式，位于 LF 的拷贝通常比 MF1 和 MF2 中的拷贝有更高的表达水平。

我们比较了四个甘蓝种质(花椰菜 Korso、结球甘蓝 OX-heart、西兰花 HDEM 和类似甘蓝的快速循环 TO1000)、另外三个油菜(*B.rapa*、*B.ngra*)和甘蓝 C 亚组的预测基因的蛋白质序列，以及其他五个十字花科植物(*Aethionema arabicum*、*Arabiopsis thaliana*、*Capsella Rubella*、*Thellungiella salsugina* 和 *Schrenkiella parual*)的预测基因的蛋白质序列。利用 1,638 个单拷贝同源基因构建了系统发育树，结果表明甘蓝与花椰菜和花椰菜的共同祖先的进化距离约为 1.68 百万年，甘蓝与甘蓝 C 亚组供体的进化距离约为 2.27 百万年，而芸薹属（*Brassica*）与其他十字花科植物的进化距离约为 16.18 百万年(图 1B)，表明甘蓝与花椰菜和花椰菜的共同祖先存在约 1.68 百万年的差异，甘蓝与甘蓝 C 亚组供体的差异约为 2.27 百万年，而 Brassica 与其他十字花科物种的差异约为 16.18 百万年(图 1B)

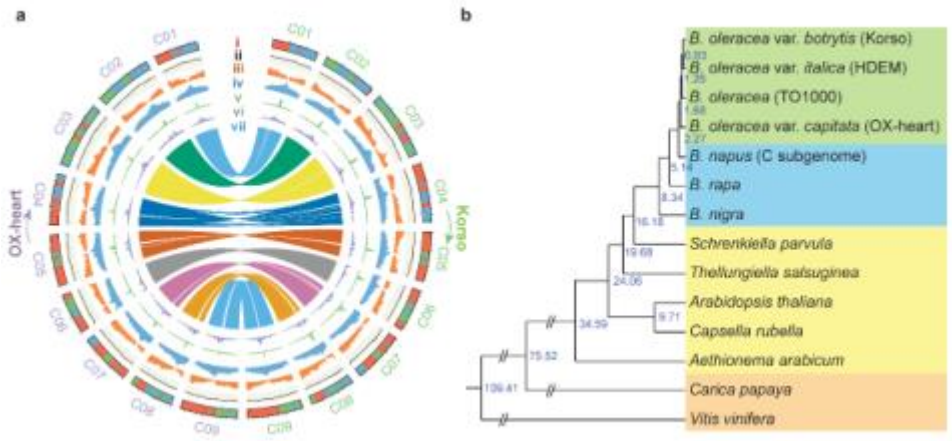

图 1 Korso 和 OX-heart 基因组注释与比较基因组分析

a，Korso 和 OX-heart 基因组的特征。i，染色体表意文字。红色，绿色，蓝色和灰色分别表示 LF，MF1，MF2 亚基因组和着丝粒区域。 ii。 GC 含量。 iii，基因密度。 iv，重复密度。 v，*Copia* 型 LTR 密度。 vi，*Gypsy* 型 LTR 密度。 vii，Korso 和 OX-heart 基因组之间的共线性分析。 b，基于 1638 个单拷贝直系同源基因的 14 种植物/变种的系统发育树及其估计的发散时间（百万年前）。

3. Korso 和 OX-heart 基因组间的 SVs

通过利用 Korso 和 OX-heart 的高质量基因组组装结果，能够通过直接基因组比较与 PacBio 长读长结合来鉴定高可信度 SVs。Korso 和 OX-heart 组装结果显示出很高的共线性，这表明它们之间的平衡重排（倒位和易位）并不深刻。因此，在本研究中，我们重点研究了不平衡的 SVs，主要是 indel。在 Korso 和 OX-heart 的基因组之间共鉴定出 119,156 个 SVs，大小在 10 bp 至 667 kb 之间，明显偏向相对较短的 SVs。

基因和启动子区域中的 SVs 可以影响相应基因的功能或表达。 SV 区分别占 Korso 和 OX-heart 总基因组大小的 14.5%和 15.0%，分别占基因区的 10.0%和 11.3%，以及编码序列的 5.9%和 6.6%，提示在基因中，特别是在编码区域中，功能限制了 SVs 的出现，而在启动子区域中没有发现 SV 的明显限制。 Korso（58.5%）和 OX-heart（58.6%）中超过一半的注释基因在其基因或启动子区域中受到至少一种 SV 的影响，并在多种生物学过程（例如细胞成分组织）中功能丰富，对压力和刺激的反应，信号转导，细胞分化，胚胎发育，基因表达和表观遗传调控以及花和分生组织发育。

我们在油菜中检测到多个先前描述的 SVs，包括 BoFLC3 中两个与亚热带花椰菜适应有关的 indel，以及 BoFRIa 的两个 indels 与花椰菜和卷心菜的冬季一年生或二年生的习性有关。

4.不同形态甘蓝中 SVs 的种群动态

花椰菜和结球甘蓝是甘蓝属植物的两种极端形态类型，识别它们独特表型(如花序和叶状头)形成背后的基因组变异，将为了解这些重要性状的分子调控提供新的见解，并为促进育种提供重要信息。我们在 Korso 和 OX-heart 中鉴定的高品质 SVs，为研究不同形态分化的甘蓝样本中 SVs 的动态变化提供了有价值的参考。为此，我们对 163 份甘蓝进行了基因组重测序，其中包括 89 份花椰菜、65 份卷心菜和 9 份西兰花。我们还收集了 Cheng 等人(2016)报道的另外 108 个甘蓝的重测序数据，包括 15 个花椰菜，39 个卷心菜，24 个西兰花，18 个大头菜，4 个中国甘蓝，2 个卷甘蓝，2 个羽衣甘蓝，2 个球芽甘蓝和 2 个野生甘蓝的重测序数据。在 271 份中，211 份测序深度在 10 倍以上。根据基因组测序序列与 Korso 和 OX-heart 基因组的比对，在 271 份样本中对 119156 份高质量参考 SVs 进行了基因分型。为了评估我们的 SV 基因分型的准确性，我们分别对 Korso 和 OX-heart 的参考 SVs 进行了基因分型，方法是将它们的 Illumina 短读分别映射到这两个基因组上。超过 86%的 SVs 可以进行基因分型，只有 0.1%的 SVs 进行了错误基因分型，表明我们的基因分型具有较高的敏感性和准确性。每个品系的 SVs 基因分型率在 41.3%至 80.2%之间，其中 187 份(69.0%)和 254 份(93.7%)品系的基因分型率分别大于 70%和 60%。总的来说，在 271 份材料中，有 89，882 的 SVs (75.4%)成功地进行了基因分型。

甘蓝型油菜不同组之间的 SV 等位基因频率变化主要是由于不同的理想性状驯化和适应不同环境的结果。不出所料，带有纯合子 Korso 等位基因的 SV 基因座在花椰菜种中普遍存在，平均占每个样本的 82.3%，而在卷心菜中，纯合的 OX-heart 等位基因则普遍存在，平均频率为 61.7%（图 2a）。使用 SV 的系统发育和主成分分析（PCA）分析将花椰菜，卷心菜，西兰花和大头菜的种质清楚地分为不同的组（图 2b 和 2c），这与我们基于相同 271 种种质的 SNP 数据揭示的模式一致，并且之前的报告基于 119 个种质一致（Cheng 等，2016），进一步支持了我们的 SV 检测和基因分型非常可靠。

为了确定可能与花椰菜或甘蓝的特定性状相关的 SVs，我们提取了总共 49,904 个 SVs，它们在花椰菜和甘蓝的种群之间具有等位基因频率显着不同的频率（图 3a）。在这些 SV 中，

49,285（98.8%）基因型的等位基因频率花椰菜显著高于卷心菜，只有 550 个基因型的等位基因频率卷心菜高于花椰菜。这些可能选择的 SVs 分布在整个染色体上而没有明显的热点。受选择的 SV 在整个基因组中的普遍存在与两个高度专业化的油菜形态型之间相对较大的发散时间（∽1.68 百万年）以及它们的独立进化和驯化历史相符（图 1b）。

在 Korso 和 OX-heart 基因组中，分别有 21,111 和 21,400 个基因在其基因或启动子区域中与至少一个受选择的 SV 重叠，在 CDS 区域中有 6,059 和 6,344 个与受选择的 SV 重叠。使用 GO 富集分析显示受选择的 SV 相关的基因在信号转导，刺激反应，细胞分化，细胞周期，胚胎发育，细胞生长和细胞死亡以及花发育有关的通路富集（图 3b），显示出花椰菜和卷心菜不同表型的潜在关联。

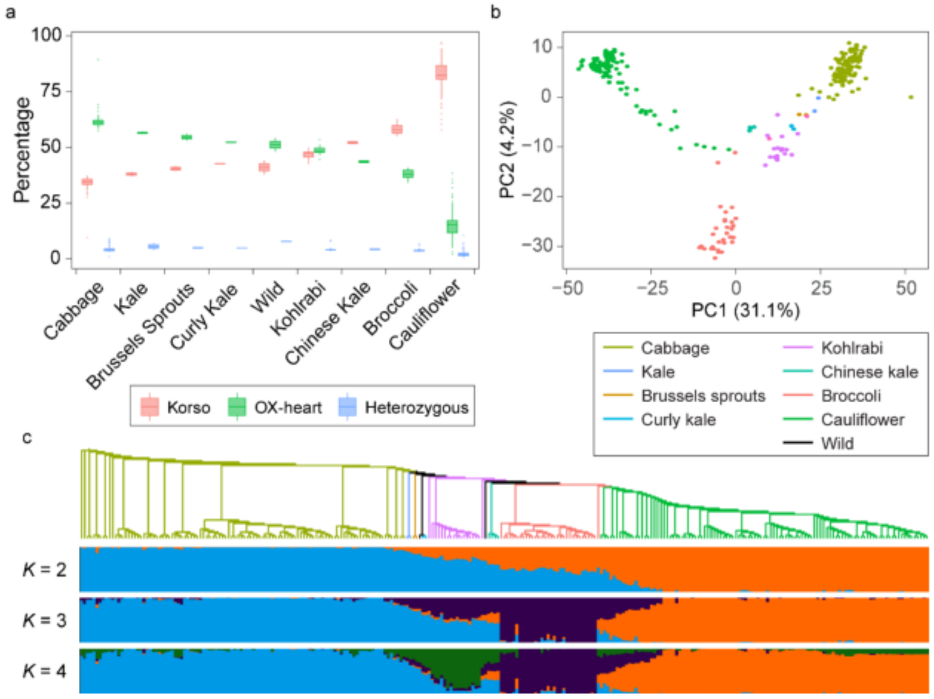

图 2 不同形态甘蓝中 SVs

a， 具有不同基因型的 SVs 在不同形态型的种质中的百分比。 b，基于 SVs 的甘蓝种质的主成分分析。 c，使用 SVs 对 271 个甘蓝种的最大似然树和基于模型的聚类。树的分支颜色表示与 b 中不同的形态颜色一致。 K，祖宗血统的数量。

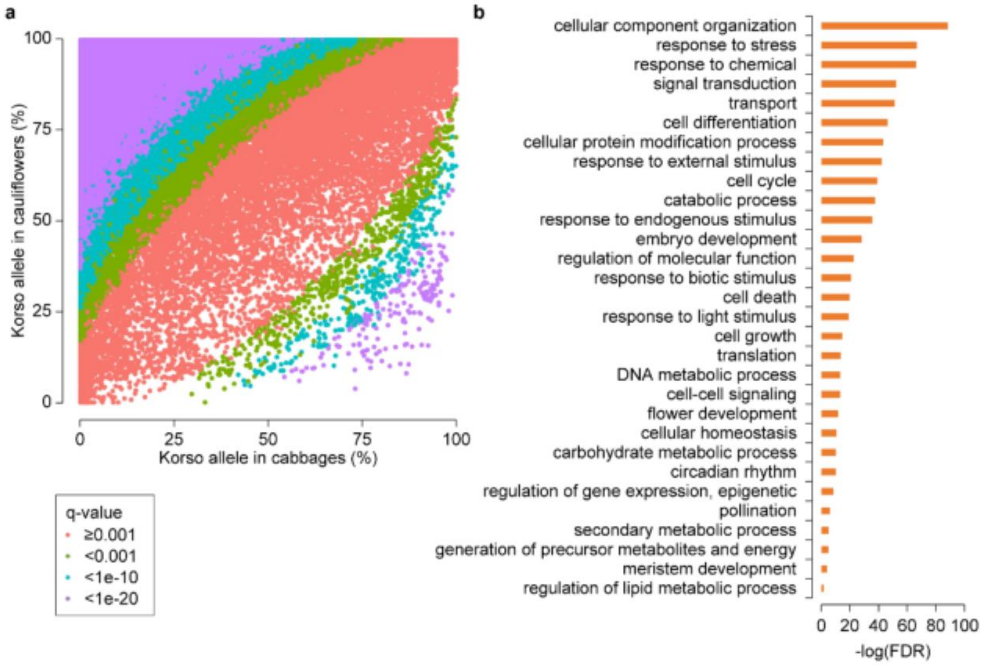

图 3 花椰菜和卷心菜之间的 SVs 差异

a，散点图，显示卷心菜和花椰菜中 SV 等位基因频率。b，与 SV 重叠的在甘蓝和花椰菜组之间具有明显的等位基因频率差异的基因 GO 富集分析。

5.受选择的 SVs 为花椰菜花球形成的进化提供了洞察力

花椰菜的花球由螺旋叠代模式的初生花序分生组织组成，在这里，我们检索到总共 294 个基因，这些基因在其启动子或基因区域中带有受选择的 SV，并且据报道，其拟南芥中的同源物在开花时间和花卉发育，分生组织维持和测定，器官大小控制以及芽或花序结构中起作用。此外，从营养枝顶端分生组织（SAM）到扩大的花球的五个阶段进行了 RNA-Seq 分析，以揭示 SV 在花球形成和发育中的潜在作用。

6.从营养发育向生殖发育的转变

花球起始的第一阶段与从营养发育到生殖发育的转变相对应（图 4a）。花椰菜及时地过渡到花蕾的形成阶段对于花球的形成是必不可少的，而对于卷心菜，需要较长的营养阶段才能使叶头正常发育。 MADS 盒转录因子 FLC 是春化和自主途径中的开花时间整合子，起着抑制开花的作用。多项研究证明了 FLC 旁系同源物在不同的油菜形态型中的开花时间中的作用。在 Korso 的 *BoFLC1.1* 启动子中发现了 3,371 bp 的插入片段（SV_b_92666a），该序列在强烈的差异选择下出现，分别存在于 99%和 88%的花椰菜和西兰花中，而仅在 9%

的卷心菜中（图 4b）。*BoFLC1.1* 及其两个串联旁系同源物（BoFLC1.2 和 BoFLC1.3）以及 *BoFLC3* 在过渡阶段均被显着下调（图 4c）。Korso 的等位基因 *BoFLC3* 基因在第一个内含子中包含 263 bp 缺失（SV_w_24534）和 49 bp 插入（SV_w_24533）。在拟南芥和十字花科作物中已经报道了 FLC 第一内含子的结构对开花时间的影响。我们发现在这两个 SV 基因座上，Korso 等位基因在花椰菜（86.7%和 86.4%）和西兰花（96.9%和 92.9%）中占优势，而在卷心菜（9.7%和 8.7%）中则很少见（图 4b）。

FLC 功能由 FRI 激活，已被确定为花椰菜花球诱导的温度依赖性时机 QTL 区域中的候选基因。在 Korso 和 OX-heart 基因组中都鉴定出两个 FRI 同源物 *BoFRI1* 和 *BoFRI2*。*BoFRI1* 启动子区域的 743 bp 缺失（SV_b_96002）是 Korso 等位基因的特征。大部分花椰菜（98.0%）具有纯合子 Korso 基因型，而大多数卷心菜（87.0%）具有纯合子 OX-heart 基因型（图 4b）。对于 *BoFRI2*，在其编码区中鉴定出两个插入片段（12 和 21 bp，SV_w_31837 和 SV_w_31838），都显示出花椰菜和卷心菜的基因型频率有显着差异（图 4b）。已经发现这两个插入缺失与冬季一年或两年一次的花椰菜和卷心菜习性有关。FES 和 SUF 可与 FRI 形成推定的转录激活物复合物，以促进 FLC 表达。与卷心菜相比，*BoFES1.1* 和 *BoSUF4.2* 在花椰菜中具有受选择的 SV，其表达在过渡期显着下调，与 *BoFLC1s* 和 *BoFLC3* 相似。参与调控 FLC 表达的其他基因，包括参与表观遗传修饰的基因，例如 PRC1 和 PRC2 复杂成分 *BoVIN3*，*BoVIL2.3*，*BoVRN1.1*，也包含受选择的 SV。这些结果共同表明，FLC 相关的自主途径和春化途径可能受到花椰菜和卷心菜之间差异 SV 的影响，从而导致它们转换到生殖阶段的时机不同。

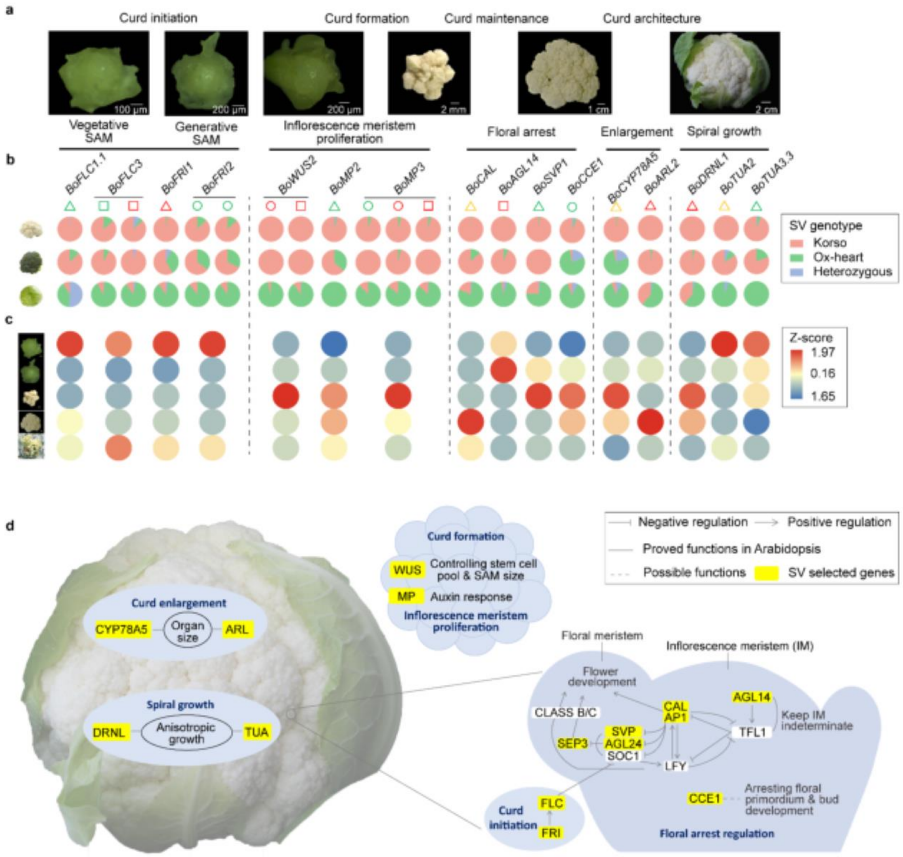

图 4 SV 对花椰菜花球形成的贡献

a，顶端分生组织（SAM）和花球在花球发展的不同阶段的示例图。b，花椰菜，西兰花和卷心菜中 SV 与候选基因重叠的等位基因频率。三角形，圆形和正方形分别表示启动子，CDS 和内含子区域中的 SV，它们的不同颜色（绿色，红色和黄色）分别表示 Korso 中与 OX-heart 相比不同类型的 SV：插入，缺失和替换。c，显示在花球发育不同阶段的候选基因表达的热图。d，花球形成和发育的调控网络。黄色背景的基因为花椰菜和卷心菜之间受选择的 SV 基因。

7. 花序分生组织增殖

花球起始后的主要过程是形成花球的未确定的花序分生组织的连续规则螺旋增殖。干细胞维持和分生组织扩散在这一过程中起着关键作用。WUSCHEL 充当生长素反应变阻器，以维持拟南芥中的顶端干细胞。在 Korso 的 *BoWUS2* 的第一个内含子中鉴定出 12 bp 的读框内缺失（SV_w_83072）和 21 bp 的插入（SV_w_83073）。所有采样的花椰菜和花椰菜种质均具有两个 SV 的纯合子 Korso 基因型，而甘蓝中的 Korso 等位基因很少（4%）（图 4b）。*BoWUS2* 的表达从营养形成到花球形成显着上调，在花球形成阶段的表达最高（图 4c），

这意味着这两个 SV 可以在花球形成中发挥作用。

MP / ARF5 与 ANT 和 AIL 一起在依赖植物生长素的器官启动和叶序模式中发挥关键作用。我们鉴定了在 *BoMP2*，*BoMP3*，*BoANT*，*BoAIL5*，*BoAIL6* 和 *BoAIL7* 的启动子和其同源基因的基因区域中选择的 SV。花椰菜 *BoMP2* 的启动子中 23 bp 的插入（SV_w_71238）在花椰菜中处于强选择状态（花椰菜和卷心菜种分别为 96.1%和 0%）。花椰菜中 *BoMP3* 基因 CDS 区中 11 bp 插入和 23 bp 缺失（SV_w_92482 和 SV_w_92481）和内含子区 14 bp 缺失（SV_w_92433）受到强烈选择（花椰菜分别为 95.6%，96.1%，96.2%和卷心菜分别为 12%，14%，13.3%（图 4b）。与 *BoWUS2* 一样，在花球形成阶段也观察到了 *BoMP2* 和 *BoMP3* 的最高表达（图 4c）。

8. 花序维持和花停滞

花椰菜的花球由成千上万的花序分生组织组成，花序分生组织在发育中被阻滞。在分生组织决定（FMI）基因 *BoCAL* 的启动子区域鉴定出大的变异（SV_b_70950）（在 OX-heart 中约 11.4 kb，在 Korso 中约 7.7 kb），受到强烈的选择。几乎所有花椰菜（99.0%）和大多数西兰花（87.5%）品种都具有 Korso 等位基因，而大多数卷心菜品种（79.2%）在此基因座上都具有 OX-heart 等位基因（图 4b），表明其花球形成中的潜在作用。

其他一些 FMI 基因，包括 *BoAP1.2*，*BoFUL1*，*BoFUL3* 和 *BoSEP3*，也受到受选择的 SVs 的影响，并且在营养，过渡和花球形成阶段均表达相对较低，但在花球膨大时表达显着提高。拟南芥的研究表明，花分生组织决定（IMI）基因 *TFL1* 和 *FMI* 基因之间的拮抗相互作用调节了发育命运的转变。与 FMI 基因相比，*BoTFL1.2* 的表达模式几乎相反，表明其抑制作用。虽然在 *BoTFL1.2* 中未鉴定出受选择的 SV，但在强选择下发现其正调控子 *BoAGL14* 的内含子有 13 bp 缺失（SV_w_84836）（图 4b），而 *BoAGL14* 表现出与 *BoTFL1.2* 相同的表达模式，表明 *BoTFL1.2* 和 *BoAGL14* 在阻止花序形成和维持，从而形成花球和维持花球中潜在重要的作用。SVP 是花卉过渡的关键负调控因子。在 Korso、98.1%的其他花椰菜和所有西兰花种质中发现了 *BoSVP1* 启动子中的 420 bp（SV_w_74120）插入，而在卷心菜种质中仅 25.2%（图 4b）。*BoSVP1* 从营养期到过渡期显着上调，并在整个花球形成过程中保持高表达水平（图 4c），表明其在花芽发育中的阻遏作用，如拟南芥中报道。

据报道，花椰菜花球特异性基因 *BoCCE1* 在分生组织发育/停止过程中具有潜在的作用。

在这里，我们确定了一个 1,505 bp 的插入片段（SV_b_67089a），覆盖了整个 Korso 中的 *BoCCE1* 基因体。该插入片段的基因分型显示，*BoCCE1* 基因存在于大多数花椰菜种质中（97.1%），但在大多数卷心菜（86.5%）和西兰花（78.1%）种质中均不存在，这表明 *BoCCE1* 在花椰菜芽中可能具有花停滞作用。与花椰菜芽相比，它们在发育后期被阻滞（图 4b，c）。

9. 花序膨大和螺旋状生长

和器官大小调节、细胞分裂和扩增、胞周期等相关基因可以调节花球的重量。*CYP78A5*（KLU）已在拟南芥中鉴定，可防止增殖停滞并促进器官生长。花椰菜 *BoCYP78A5* 的高表达仅在花球中检测到，尤其是在花球形成和膨大阶段（图 4c）。*BoCYP78A5* 启动子中的 2775-bp 取代（SV_b_76292）存在于 98.1%的花椰菜种中，而仅存在于 8.2%的卷心菜种中（图 4b），可能有助于 *BoCYP78A5* 的花球特异性表达。*BoARL2*（或 *CDAG1*）已被证明在促进花椰菜花球大小中起作用。*BoARL2* 在花球中高度表达，在 Korso 的花球扩大阶段表达最高（图 4c）。在 *BoARL2* 的启动子中检测到一个 269 bp 的缺失（SV_w_38468），并存在于所有花椰菜和大多数西兰花（96.9%）的种中，而只有 41.7%的卷心菜种（图 4b）。

花序的螺旋排列是花椰菜花球的典型特征。*DRNL* 基因的转录标志着花序分生组织外围区域的侧方器官开始分化。在 *BoDRNL1* 的启动子中发现了 258 bp 的缺失（SV_w_30645），该启动子存在于所有花椰菜和西兰花品种中，而仅在卷心菜品种中占 40.6%（图 4b）。*BoDRNL1* 特别在花球中表达，在花球形成和膨大阶段表达最高（图 4c），这表明其在决定花球结构中的潜在作用。在 alpha-微管蛋白基因 *BoTUA2* 和四个 *BoTUA3* 基因中也鉴定出了受选择的 SVs（图 4b），其在在拟南芥中的同源基因导致螺旋生长。

总结

油菜 *B. oleracea* 物种包括许多重要的蔬菜作物，它们表现出异常高的形态多样性，其中花椰菜和卷心菜代表两种极端形态。在这项研究中，我们通过整合 PacBio 的长读长和 Hi-C 数据，为花椰菜和结球甘蓝组装了高质量的染色体级基因组序列，这为油菜作物的未来研究和改良提供了重要资源，为全面探索油菜表型多样性提供基础。

SVs 在植物表型变化的遗传调控中起着至关重要的作用，并且通常是许多重要性状的致病性遗传变异，这些重要性状是作物驯化和育种的目标。但是，农作物中 SVs 的种群分析

远远落后于 SNP 的种群，这主要是由于准确识别 SVs 的技术困难。当前广泛使用的 SVs 检测方法依赖于短测序 reads 和参考基因组的比对，这容易导致高假阳性率和高假阴性率。PacBio 和 Nanopore 等长读长测序技术的最新进展有助于 SVs 的检测。但是，由于读取长度受限制，因此无法检测到一些较大的 SV（例如插入）。在本研究中，通过直接比较花椰菜和卷心菜的高质量参考基因组，结合长读比对，识别出约 120K 个高置信度 SV，其中许多大于 100 kb。在包含 271 个代表不同 *B. oleracea* 形态的种质的群体中，此 SVs 参考集的基因分型以及对这些 SVs 在不同形态类型群体中（主要是花椰菜，西兰花和卷心菜）的等位基因频率差异的研究揭示了许多在某些形态型下处于选择状态的 SVs，其中许多影响基因与其相应的独特表型的行成。

花椰菜的花球由成千上万的花序分生组织组成，这些分生组织螺旋状排列在较短的扩大的花序分支上。这使得花椰菜成为分析花序发育和极端器官发生的遗传机制的理想模型。在花椰菜中受选择的 SVs 影响许多基因。结合对花球发育过程中表达谱的分析，我们鉴定了数十种关键的 SVs 和相关基因，它们与花椰菜的独特花球表型具有潜在的关联。这些包括在花球发育的不同发育阶段中起作用的基因。第一阶段是花球起始，涉及从营养阶段到生殖阶段的转变，影响开花时间调节的基因（例如 *FLC* 和 *FRI*）受到影响。花球形成的关键步骤是花序增殖，其中 *WUS* 和 *MP* 等基因具有花椰菜和西兰花特定的 SVs。花椰菜花球的特征还在于花分生组织的停滞，与几个花序决定基因（例如 *CAL*，*AP1* 和 *SEP3*）以及受选择的 SVs 影响的潜在负调控基因（例如 *AGL14*，*SVP* 和 *CCE1*）相匹配。在花椰菜花球结构中起作用的几个基因也受到所选 SVs 的影响，这些包括在器官大小控制（例如 *CYP78A5* 和 *ARL*）和花序螺旋组织（例如 *DRNL* 和 *TUA*）中可能发挥作用的基因（图 4d）。我们的分析表明，SVs 对花椰菜独特的花球表型有重要贡献，并阐明了花椰菜花球形成的调控网络。

参考文献

1. Ning Guo, Shenyun Wang, Lei Gao,et al. Genome sequencing sheds light on the contribution of structural variants to *Brassica oleracea* diversification.*bioRxiv*,2020,doi: https://doi.org/10.1101/2020.10.15.340224.