

基因组测序揭示结构变异对甘蓝多样性的贡献

王静

科技服务技术支持部产品经理

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
Brassica oleracea diversification

- ◆ 发表时间: 2020年10月
- ◆ 发表期刊: bioRxiv
- ◆ 主要单位: 北京农林科学院
- ◆ 关键词: 基因组、结构变异、多样性

背景介绍

甘蓝介绍

十字花科芸薹属的一年生或两年生草本植物，存在很多变种。甘蓝包括几种不同的优势蔬菜作物。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。

花椰菜(*B. oleracea* var. *botrytis*)：花球供食用，由茎轴和很多的肉质花梗及绒球状的花枝顶端包括许多分化到花序阶段的花芽原基聚合而成；

西兰花 (*B. oleracea* var. *italica*)：花球供食用，顶端生紧密群集成花球状的群生花蕾；

结球甘蓝(*B. oleracea* var. *capitata*)：又名卷心菜、洋白菜，顶芽或叶芽形成的叶球供食用；

抱子甘蓝 (*B. oleracea* var. *gemmifera*)：又名小圆白菜、小卷心菜，以腋芽形成的小叶球为产品；

芥蓝(*B. oleracea* var. *alboglabra*)：肥嫩的花薹和嫩叶供食用；

擘蓝(*B. oleracea* var. *gongylodes*)：擘蓝以膨大的肉质球茎和嫩叶为食用。



花椰菜



西兰花



结球甘蓝



抱子甘蓝



芥蓝



擘蓝

背景介绍

相关研究

物种	时间	期刊	主要单位	品系	组装策略	组装指标
甘蓝	2014.05	Nature Communications	中国农科院油料所	02-12	Illumina, Roche 454, Sanger	基因组大小: 540 Mb Contig N50: 26.8 kb Scaffold N50: 1.46 Mb
	2014.06	Genome Biology	加大拿农业与农产品部	TO1000DH	Illumina, Roche 455, Sanger (共94×)	基因组大小: 488.6 Mb Scaffold N50: 850kb
	2018.11	Nature Plants	法国巴黎-萨克雷大学	HDEM	Nanopore 36X,Illumina,BioNano	Contig N50 9.49Mb, Scaffold N50 25.52Mb

时间	期刊	主要单位	材料	测序策略	研究内容
2016.8	Nature Genetics	农业科学院	199个芜菁及119个甘蓝品	二代重测序	亚基因组平行选择与形态多样化及作物驯化的关联
2016.11	Nature Communications	澳大利亚昆士兰大学	9种形态多样的甘蓝品种和野生亲缘芸薹属大白菜	二代迭代组装	PAV对于芸苔属作物的育种改良

背景介绍

文章出发点

- **结构变异很重要：**结构变异(SVs)，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。
- **甘蓝中结构变异被忽略：**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。
- **短读长方法有缺陷：**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。
- **基因组组装+长读长SVs检测准确性高：**而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。

研究内容：甘蓝结构变异对甘蓝多样性的影响。

材料方法

- **基因组denovo**: 花椰菜(Korso)和结球甘蓝 (OX-heart) 。

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), 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。

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**: 163份甘蓝属不同种质(花椰菜89份, 甘蓝65份, 西蓝花9份)。

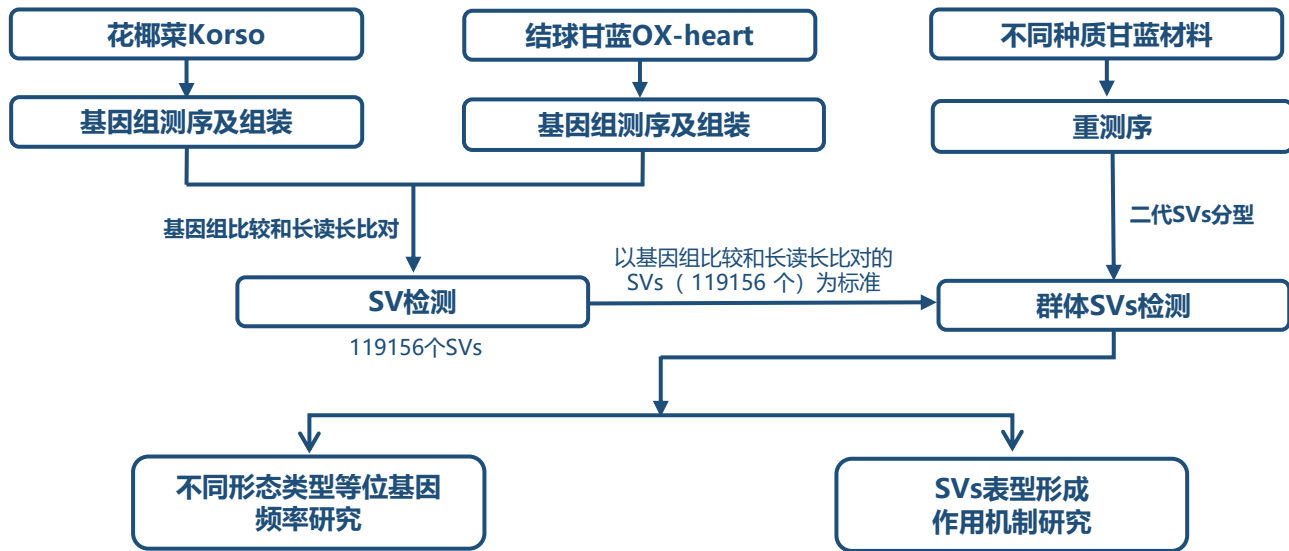
Illumina HiSeq 2500平台, PE150 测序策略(117份)或PE100测序策略 (46份)。

- **转录组**: Korso样本 (根、茎、叶、花球、芽、花、角果及不同时期的顶端分生组织(SAM)) 和OX-heart样本 (根、茎、叶、头状花序、芽、花和角果) 。

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。

研究思路



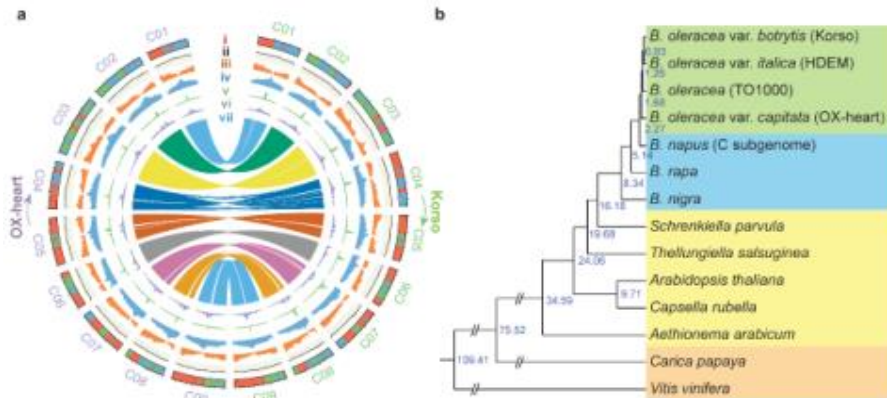
一、花椰菜和结球甘蓝基因组组装与注释

样本	组装策略	基因组大小	Contigs N50	BUSCO
花椰菜 Korso	PacBio+Illumina+ BioNano+Hi-C	549.7 Mb	4.97 Mb	97.2%
结球甘蓝 OX-heart	PacBio+Illumina+ Hi-C	565.4 Mb	3.1Mb	96.5%

1. Korso、OX-heart与HDEM组装结果的Hi-C热图之间有良好的良好一致性。
2. 约60.7%的Korso和62.0%的OX-heart被注释为重复元件，其中 *Gypsy* 和 *Copia* 反转录转座子最为丰富。
3. 对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。
4. 从组装结果能够精确识别着丝粒位置。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。
5. Korso和OX-heart基因组分别预测到60,640和62,232个高可信度蛋白质编码基因，BUSCO显示Korso和OX-heart预测基因的完整性分别为96.1%和94.9%。

二、比较基因组分析

i, 染色体
红色: LF亚基因组
绿色: MF1亚基因组
蓝色: MF2亚基因组
灰色: 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。



甘蓝

油菜

其他
十字
花科

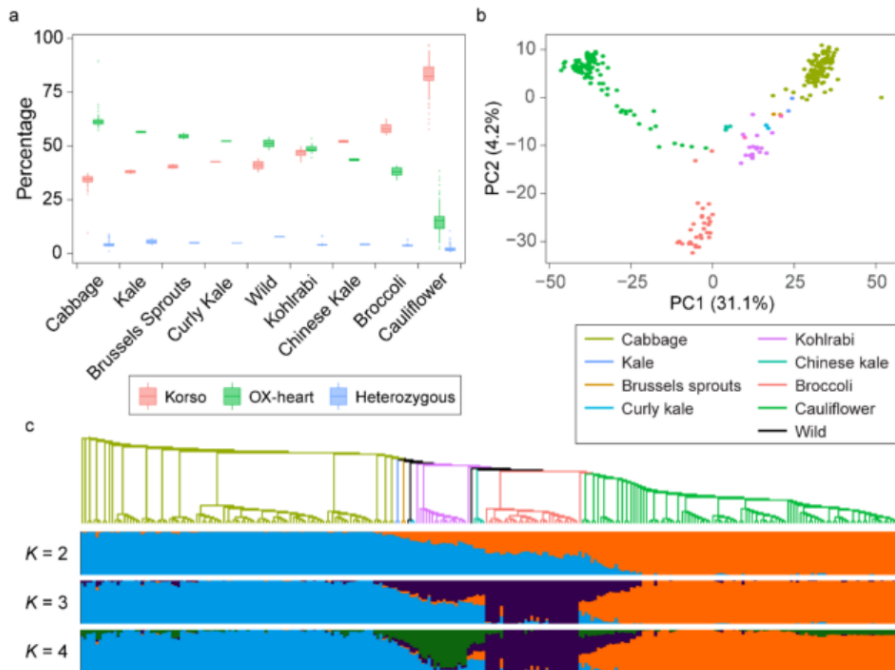
其他科

Korso和OX-heart基因组注释与比较基因组分析

1.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, LF(最不分离的), MF1(中等分离的)和MF2(最分离的)。

2.进化距离:甘蓝与花椰菜和花椰菜的共同祖先进化距离为1.68 百万年, 而芸薹属与其他十字花科物种的进化距离约为16.18 百万年。

三、SVs检测及其种群动态

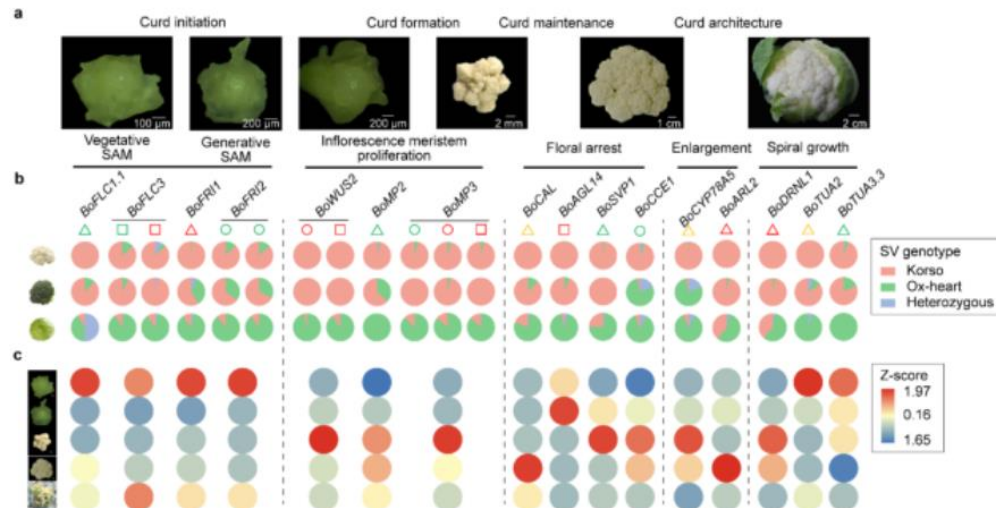


不同形态甘蓝中SVs

群体SVs检测：通过基因组比较与PacBio长读长结合来鉴定高可信度SVs。以鉴定的高可信度SVs为标准，用271份样本二代重测序数据进行SVs分型，得到群体SVs。

1. 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
2. 超过一半的注释基因（Korso (58.5%) 和OX-heart (58.6%)）中在其基因或启动子区域中受到至少一种SV的影响。
3. 基于SVs的Tree、PCA、Sturcture分析可将花椰菜，卷心菜，西兰花和大头菜的种质清楚地分为不同的组。

四、SVs与花椰菜花球形成



SV对花椰菜花球形成的贡献

1. 共检测到294个基因（其启动子或基因区域中带有受选择的SVs）在花椰菜花球的形成中具有潜在作用。
2. 不同阶段的潜在作用：从营养发育向生殖发育的转变阶段-花序分生组织增殖阶段-花停滞阶段-花序膨大和螺旋状生长阶段，SVs可影响不同阶段基因结构及表达情况，从而影响花椰菜花球行成。

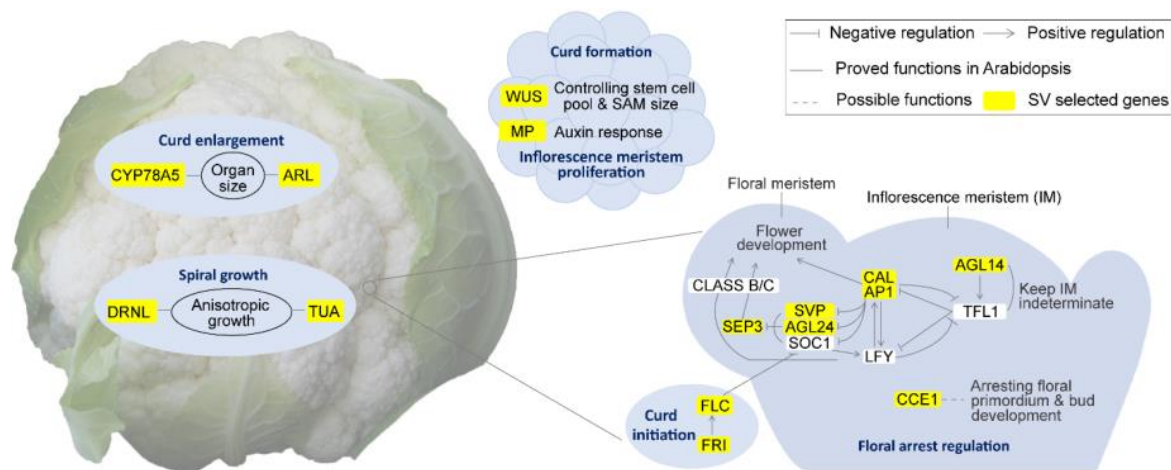
四、SVs与花椰菜花球形成

花球形成阶段	相关基因描述	相关基因及区域	受选择的SVs	等位基因频率	基因表达量
从营养发育向生殖发育的转变	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。	BoFLC1.1	启动子区域3,371 bp的插入	花椰菜99%，西兰花88%，卷心菜9%。	BoFLC1.1及其两个串联旁系同源物（BoFLC1.2和BoFLC1.3）以及BoFLC3在花椰菜过渡阶段均被显著下调
		BoFLC3	第一个内含子中263 bp缺失和49 bp插入	花椰菜（86.7%和86.4%），西兰花（96.9%和92.9%），在卷心菜（9.7%和8.7%）	-
	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。	BoFRI1	启动子区域的743 bp缺失	花椰菜（98.0%）	-
		BoFRI2	编码区中鉴定出两个插入片段（12和21 bp，SV_w_31837和SV_w_31838）	花椰菜和卷心菜的基因型频率有显著差异	-
	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。	BoFES1.1	-	-	在过渡期显著下调
		BoSUF4.2	-	-	在过渡期显著下调
	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。	PRC1和PRC2复杂成分BoVIN3，BoVIL2.3，BoVRN1.1	-	-	-

四、SVs与花椰菜花球形成

花球形成阶段	相关基因描述	相关基因及区域	受选择的SVs	等位基因频率	基因表达量
花序分生组织增殖阶段	干细胞维持和分生组织扩散相关基因。	BoWUS2	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	花椰菜	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
	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。	BoMP2	CDS区11 bp插入和23 bp缺失，内含子区14 bp缺失	花椰菜 (95.6%， 96.1%， 96.2%)，卷心菜 (12%， 14%， 13.3%)	花球形成阶段最高表达
		BoMP3	-	-	花球形成阶段最高表达
花停滞阶段	分生组织决定 (FMI) 相关基因	BoCAL	启动子区域鉴定出大的变异 (OX-heart中约11.4 kb，在Korso中约7.7 kb)	花椰菜 (99.0%) 和大多数西兰花 (87.5%) 品种具有Korso等位基因，大多数卷心菜 (79.2%) 中具有OX-heart等位基因	花停滞期表达最高
	BoTFL1.2正调控子基因	BoAGL14	内含子有13 bp缺失	-	与FMI基因表达模式几乎相反
	SVP是过渡的关键负调控因子	BoSVP1	启动子中的420 bp插入	花椰菜 (98.1%)，西蓝花 (100%)，卷心菜 (25.2%)	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
	花椰菜花球特异性基因	BoCCE1	覆盖整个BoCCE1基因的1,505 bp的插入片段	花椰菜 (97.1%)	-
花序膨大和螺旋状生长阶段	防止增殖停滞并促进器官生长相关基因	BoCYP78A5	启动子中的2775-bp取代	花椰菜 (98.1%)，卷心菜 (8.2%)	花球形成和膨大阶段高表达
	促进花椰菜花球大小相关基因	BoARL2 (或CDAG1)	启动子中269 bp的缺失	花椰菜 (100%)，西兰花 (96.9%)，卷心菜 (41.7%)	花球中高度表达，在花球扩大阶段表达最高
	花序分生组织外围区域的侧方器官开始分化的标记基因	BoDRNL1	启动子258 bp的缺失	花椰菜 (100%)，西兰花 (100%)，卷心菜 (40.6%)	花球形成和膨大阶段表达最高

四、SVs与花椰菜花球形成



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

1. 花球起始阶段，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。
2. 花球形成的关键步骤是花序增殖，其中WUS和MP等基因受到影响。
3. 花分生组织的停滞阶段，几个花序决定基因（例如CAL，AP1和SEP3）以及潜在负调控基因（例如AGL14，SVP和CCE1）受到影响。
4. 花序膨大和螺旋状生长阶段，器官大小控制（例如CYP78A5和ARL）和花序螺旋组织行成（例如DRNL和TUA）中可能发挥作用的基因受到影响。
5. 综上，受选择的SVs通过影响花椰菜开花时间，分生组织维持，器官大小等过程，从而影响其花球的行成。

总结

亮点:

- 材料丰富，思路清晰，跟随研究热点（结构变异）
- 群体SVs检测方法准确性高（相对于纯二代检测），检测成本较低（相对于纯三代），适应性广
- 对SVs在某些表型形成中的作用结合代谢通路进行系统阐述

不足:

- 生物学故事不够，缺乏典型重测序探讨及结果（如群体进化研究，物种形成机制，群体进化动态和环境适应性等）
- 热点未深入探讨，未进行全基因组关联分析（SV-GWAS），未挖掘SVs与其他表型关联信息
- 证据不充分，对结果未进行验证，所有文章结果均基于测序结果及拟南芥中相关的研究情况

Thank You!



官方网站



官方微信