植物遗传资源学报 2021, 22(1): 7-15
Journal of Plant Genetic Resources
DOI: 10.13430/j.cnki.jpgr.20200528001

泛基因组及其在植物功能基因组学研究中的应用

赵均良,张少红,刘　斌
(广东省农业科学院水稻研究所/广东省水稻育种新技术重点实验室,广州 510640)

摘要: 参考基因组是现代功能基因组学的核心框架,以此为基础的现代基因组学技术在过去20年对植物遗传变异发掘、功能基因克隆等研究起了巨大的推动作用。然而,越来越多的研究发现,单一或少数参考基因组不能完整代表和呈现物种或特定群体内的所有基因组变异,因此其在功能基因组学研究中应用存在很大的局限性,甚至会导致错误的结果。泛基因组是指物种或特定群体内全部基因或基因组序列的总和。泛基因组通过完整捕获和呈现群体内全部的基因或基因组序列,代替单一参考基因组应用于功能基因组学研究,可以突破单一参考基因组的局限性。泛基因组在植物功能基因组学研究中有广泛的应用,以泛基因组为基础,结合最新的基因组学技术可以高效、精准鉴定种质资源中的遗传变异。泛基因组研究是目前植物基因组学研究的前沿和热点。本文综述了泛基因组概念的起源和发展,泛基因组组装的技术和策略,以及泛基因组在植物基因组学研究和分子育种方面的应用和最新进展,最后对植物泛基因组研究目前存在的问题和今后研究方向进行了展望。本综述可为植物泛基因组研究和应用提供参考。

关键词: 泛基因组;种质资源;功能基因组学;遗传变异

Research Progress on Pangenome and Its Application in Plant Functional Genetics

ZHAO Jun-liang, ZHANG Shao-hong, LIU Bin
(*Rice research institute, Guangdong Academic of Agricultural Sciences/Guangdong Key Laboratory of New Technology in Rice Breeding, Guangzhou 510640*)

Abstract: The reference genome provides a fundamental framework in the field of modern functional genomics. Modern genomic techniques based on reference genomes is the driving force for identification of genetic variations and functional gene cloning in the past two decades. However, with the progress of plant genome research, it is becoming visible that single reference genome cannot represent all the genomic variances within a species or a population. This consequence has become one of the limits for the functional genomic studies. Pangenome refers to the full complement of genes/genomic sequences within a population. By capturing and presenting all the genes/genomic sequences within a population, pangenome can be used as reference genome in functional genetic studies, to overcome the limit posed by single reference genome. While implementing in the state-of-art genomic techniques, pangenome can greatly improve the efficiency and precision of identification of genetic variations in plant germplasm, thus becoming one of the frontiers and hotspots in genomic research. In this

收稿日期: 2020-05-28　　**修回日期:** 2020-07-06　　**网络出版日期:** 2020-07-31
URL: http://doi.org/10.13430/j.cnki.jpgr.20200528001
第一作者研究方向为水稻功能基因组学,生物信息学, E-mail: zhao_junliang@gdaas.cn
通信作者: 刘斌,水稻功能基因组学, E-mail: lbgz1009@163.com
基金项目: 广东省基础与应用基础研究基金(2020A1515010906);科技创新战略专项资金(高水平农科院建设)(R2019-JX001);广东省重点领域研发计划项目(2018B020202004);2019年省级现代农业科技创新联盟建设共性关键技术创新团队项目(2019KJ106);广州市科技计划(201804020078)

Foundation projects: Guangdong Basic and Applied Basic Research Foundation(2020A1515010906), Special Fund for Scientific Innovation Strategy-construction of High Level Academy of Agriculture Science(R2019-JX001), Key Areas Research Projects of Guangdong Province(2018B020202004), Team Project of Guangdong Agricultural Department(2019KJ106), Scientific and Technological Plan of Guangzhou(201804020078)

review, we summarized the definition and recent progress in studies of pangenome, the techniques and strategies applied in pangenome assembly and construction, as well as the application of pangenome in plant genomic and molecular breeding. The current problems and future perspectives of plant pangenome were outlined. This review is expected to provide a better understanding and reference for future study and application of plant pangenome.

Key words: pangenome; germplasm; functional genomic; genetic variation

控制重要农艺性状的遗传变异和功能基因是植物育种和遗传改良的核心物质基础。如何从广泛的种质资源中高效、精准地发掘并利用优异的遗传变异，是目前植物育种基础理论研究和应用研究最重要的课题之一。高质量参考基因组的构建，以及现代基因组学技术的发展，是近 20 年来植物功能基因组学研究的核心推动力。高质量参考基因组为分子图谱构建，遗传变异发掘以及功能基因克隆提供统一的标准和框架。在此框架上，最新的高通量基因组学和遗传分析技术，如重测序分析、转录组测序、全基因组关联分析（GWAS，genome wide association study）等技术为遗传变异发掘和功能基因克隆提供了高效的技术手段。

然而，植物在自然界中存在大量多样性的资源，随着对这些多样性资源的大规模鉴定和研究，特别是大量资源的比较基因组学研究发现，单一或少数参考基因组不能完整代表和呈现这些资源中所有的遗传变异，以单一参考基因组为基础对这些多样性资源进行基因组学研究，极易出现系统性的偏差，甚至错误[1]。单一参考基因组在功能基因组学研究中最大的局限性在于难以有效鉴定基因组中的结构变异（SV，structure variation），特别是在参考基因组上缺失的结构变异和大结构变异（大于 100 bp 的结构变异）。大量的研究表明，包括拷贝数变异（CNV，copy number variation）、获得与缺失变异（PAV，presence/absence variation）在内的结构变异在作物品种间普遍存在，例如水稻基因组大约 370 Mb，而两个水稻品种之间受结构变异影响的序列平均达到 20~70 Mb[2]。基因组的结构变异可以导致物种个体间基因结构、表达特性、基因剂量等发生实质性地变化，并对植物农艺性状有非常深刻的影响[3-6]。

为了突破单一参考基因组的局限，完整捕获和描述群体中所有的遗传变异，高效鉴定和解析基因组结构变异，科学家提出了泛基因组（pangenome）的概念。通过捕获、呈现群体（物种）中全部的基因组序列，泛基因组为功能基因组学研究提供更完整的基础变异数据和参考框架。泛基因组在植物功能基因组学、分子育种等研究领域都有广泛的应用。近年来，植物泛基因组研究在新一代测序技术和现代生物信息学技术快速发展的背景下得到了飞跃性的发展，已成为植物基因组学研究的前沿和热点。本综述介绍了泛基因组概念及其演变，以及植物泛基因组最新研究进展，重点介绍最新测序技术和生物信息学技术在泛基因组构建和研究中的应用，以及泛基因组在植物基因组学研究和分子育种方面的应用，最后对植物泛基因组研究未来发展方向进行了展望，为植物泛基因组的理论与应用研究提供参考。

1 泛基因组的概念及植物泛基因组研究进展

1.1 泛基因组概念来源与发展

泛基因组的概念最初来源于微生物基因组学研究。早期通过对细菌基因组进行组装和分析研究发现，不同菌株之间存在大量的基因存在 - 缺失变异（PAV）。根据这个结果，Tettelin 等[7]首次提出“泛基因组”的概念，具体是指某种微生物所有菌株的全部基因总和。泛基因组概念很快被拓展并应用于人类、动物、植物等生物的基因组学研究中。目前已经构建包括人类、猪、大豆、木豆等超过 20 种真核生物的泛基因组[8-14]。

根据研究对象和重点的不同，泛基因组有两种定义。一种是功能性的定义，以功能基因为核心研究对象[15]。根据此定义，泛基因组是指群体内所有功能基因的总和。这个定义以功能基因为泛基因组的基本元素，通过泛基因组研究群体内功能基因的分布和结构变异，捕获和呈现群体内全部的功能基因。这个定义还可以进一步从基因拓展至基因家族，以基因家族为元素进行泛基因组构建和分析[16]。第二种是结构性的定义，是通过对群体内的基因组序列进行比较分析，收集和呈现群体内全部基因组序列，并以序列变异作为泛基因组分析的基本元素。以基因为核心的泛基因组定义在微生物泛基因组研究中应用较多。然而真核生物基因组存在大量基因间的序列变异，这些变异很大部分是具有重要生物学功能的，如基因启动子等。因此以序列为基础的

泛基因组定义更适合于真核生物泛基因组研究[5]。

无论以基因为核心还是以序列为核心的定义，泛基因组都包括核心基因组（core genome）和非必需基因组（dispensable genome）。核心基因组是指存在于所有个体中的基因（序列），非必需基因组是指只存在于部分个体中的基因（序列）[15]。从进化的角度看，核心基因组的基因（序列）主要包括维持该物种存活所必需的最基本生理功能的基因，这些基因在细胞内行使极其关键的功能，如维持 DNA 复制、翻译和细胞稳态等功能，因此这些基因在群体中具有较大的保守性[17]。非必需基因组则包含了群体内全部的可变基因（序列）。研究发现，很多物种都存在大量非必需基因组，而且非必需基因（组）相对核心基因（组），其进化时间更短，拥有更少的外显子和更丰富的多样性。例如非必需基因（组）的非同义突变与同义突变的比例显著高于核心基因（组）。微生物泛基因组研究发现大部分非必需基因组与微生物的多样性和适应性有关，对植物的非必需基因功能研究发现，这些基因的功能主要富集在生物胁迫和非生物胁迫抗性、植物发育等生物学功能中。因此，非必需基因（组）很可能与生物对环境的适应性有关。

1.2　植物泛基因组研究进展

在植物中，泛基因组最早应用于对转座子的研究[18]。此后出现越来越多的植物泛基因组构建和研究的报道。截至目前，已有超过 16 个植物物种构建了泛基因组，其中包括水稻、玉米、小麦等重要的农作物（表 1）。植物泛基因组研究发现，植物群体中非必需基因数量非常庞大，非必需基因占泛基因组的比例范围在 33%~80% 之间[15]。例如油菜泛基因组中，有 38% 的基因属于非必需基因[21]；水稻的泛基因组鉴定到 10872 个在原来日本晴参考基因组中缺失的基因，其中包括控制水稻耐淹的功能基因 *Sub1A*，控制低磷胁迫的基因 *Pstol1* 等重要农艺性状的功能基因[28]。笔者前期组装的木豆泛基因组，一共组装得到 1900 个在参考基因组上不存在的新基因，基因功能注释显示，这些基因的功能主要富集在防御反应等生物学功能中[8]，可能与不同来源的材料的抗病性有关。可见，非必需基因是控制植物重要农艺性状和表型变异的一类功能基因，泛基因组对这些基因进行捕获和呈现，有利于这些功能基因的鉴定和克隆。

表 1　植物泛基因组研究汇总

Table 1　The pangenomes accomplished in plant species

物种 Species	基因组大小（Mb） Genome size	构建泛基因组材料数量 Number of individuals	泛基因组大小（Mb） Pangenome size	构建策略 Construction strategy	参考文献 References
拟南芥 *Arabidopsis* Heynh.	120	7	135	从头组装	[19]
二穗短柄草 *Brachypodium distachyon*（L.）P. Beauv.	272	54	430	从头组装	[20]
甘蓝 *Brassica oleracea* L.	488	10	587	迭代组装	[21]
甘蓝型油菜 *Brassica napus* L.	1000	8	1800	从头组装	[9]
甘蓝型油菜 *Brassica napus* L.	1000	53	1044	迭代组装	[22]
芜菁 *Brassica rapa* L.	490	3	NA	map-to-pan	[23]
辣椒 *Capsicum* L.	3360	383	4316	迭代组装	[24]
野生大豆 *Glycine soja* Siebold & Zucc.	1000	7	NA	从头组装	[13]
大豆 *Glycine max*（L.）Merr.	1000	26	NA	从头组装	[10]
向日葵 *Helianthus annuus* L.	3000	493	NA	map-to-pan	[25]
蒺藜苜蓿 *Medicago Truncatula* Gaertn.	400	15	463	从头组装	[26]
水稻 *Oryza sativa* L.	400	3	NA	从头组装	[27]
水稻 *Oryza sativa* L.	400	3010	638	迭代组装	[2]
水稻 *Oryza sativa* L.	400	67	NA	从头组装	[28]
杨树 *Populus* L.	500	3	563	重测序单独比较	[29]
芝麻 *Sesamum indicum* L.	350	5	554	迭代组装	[30]
番茄 *Solanum Lycopersicum* L.	950	725	1300	从头组装	[31]
小麦 *Triticum aestivum* L.	17000	19	17350	迭代组装	[32]
玉米 *Zea mays* L.	2400	503	NA	转录组组装比较	[33]
木豆 *Cajanus cajan*（L.）Huth	606	89	622	迭代组装	[8]

2 泛基因组组装方法与策略

2.1 泛基因组组装方法与策略概述

目前泛基因组主要是以新一代测序数据为基础，运用相应的生物信息学技术进行组装和构建。从整体策略上，主要分为以下 3 种，第 1 种是对所有材料进行基因组从头组装，通过从头组装的基因组进行相互比较来构建泛基因组；第 2 种是以已有的参考基因组为基础，提取基因组非冗余序列并进行组装，通过“比对 - 组装”的迭代过程或者通过类似“宏基因组”的混池组装进行构建；第 3 种是基于变异图（variation graph）的组装方法[34]（图 1）。

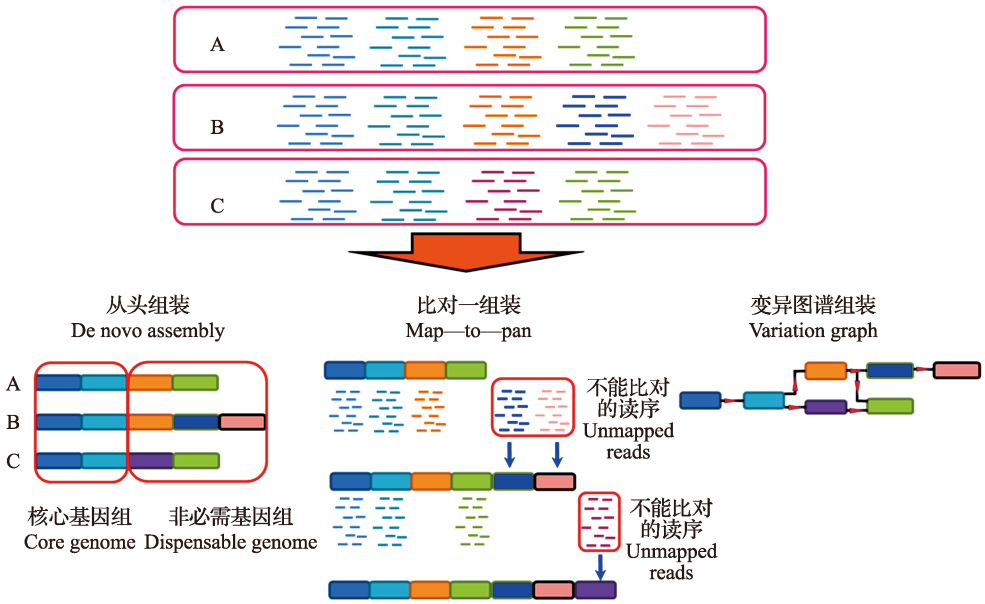

图 1 泛基因组构建方法示意图

Fig.1 Schematic diagram of pangenome assembly approaches

不同的组装策略有不同的优缺点。目前应用较多的是第一和第二种策略。无论应用哪种策略，植物泛基因组构建最大的挑战都是基因组内大量重复序列和结构变异序列的正确组装和捕获。第一种策略，即对每个材料进行基因组从头组装，再进行比较分析的方法，可以较好解决重复序列的组装及其在基因组上物理位置的正确放置的问题，减少组装错误，也有利于结构变异的发掘和鉴定，适用于构建高质量的泛基因组。但缺点是基因组从头组装费用较高，而且对计算要求较高，难以大规模实施。另外，很多基于二代测序数据的从头组装基因组质量较差，并不能很好解决重复序列的组装和大结构变异的解析问题。

第 2 种策略是以参考基因组为基础进行迭代组装或者 map-to-pan 的组装策略。通过测序数据与参考基因组进行比对，把不能比对的测序数据进行组装，把组装序列放回参考基因组作为下一轮比对的参考基因组，如此迭代进行，这就是迭代组装法。如果不采用迭代方法，则是 map-to-pan 的策略。还可以运用类似宏基因组组装的策略，把不能比对的测序数据进行混池后再组装。通过这种策略组装得到的泛基因组，再利用原始测序数据进行比对，实现基因的 PAV 分析。这种策略优点是成本低，技术相对简单，易于实现，缺点是难以解决重复序列的组装，以及由于大结构变异存在而导致的组装错误，也不能很好地对大结构变异进行解析。

第 3 种策略基于图论的组装方法是利用 de Bruijn 有向图把基因组分成若干部分，每部分在泛基因组中的关系和位置可以通过变异图谱进行分析和追踪，从而组成泛基因组图（pangenome graph）。泛基因组图为泛基因组提供了全新的概念和数据结构，然而相对于常规的线性化基因组概念，泛基因组图的概念较难理解，且相应技术和理论尚不成熟，因此目前应用还较少。但是相对于线性的基因组概念和变异展示形式，泛基因组图更适合于变异数据的组织和展示，以及下游的遗传变异分析，因此具有非常广阔的应用前景[35]。

2.2 新测序技术和组装技术在泛基因组组装中的应用

目前绝大部分植物泛基因组的构建是基于二代

测序技术进行组装的。由于二代测序读长短的天然缺陷，无论是应用基于参考基因组的策略还是从头组装策略，都难以解决因重复序列和大结构变异造成的组装错误。同时，以二代测序为基础组装的泛基因组碎片化问题非常严重。这些问题直接影响到后续泛基因组在结构变异鉴定和分析中应用的效果。

随着基因组技术的快速发展和成本大幅下降，PacBio、Nanopore 等长读长的基于单分子测序的三代测序技术以及基因组从头组装技术在泛基因组组装中得到广泛应用，极大地提高了泛基因组组装质量。具体而言，由于三代测序具有长读长的优点，可以跨越大范围的结构变异和重复序列，因此可以有效解决这些序列的组装和解析问题。另一方面。在最新的从头组装辅助技术，如 Hi-C，光学图谱等技术的支持下，极高质量的基因组组装成为可能[36]。以高质量的基因组为基础，可以组装得到更为完整，质量更好的泛基因组，最大限度的捕获和解析群体内的基因组结构变异[37]。基于三代测序和从头组装技术进行泛基因组组装的方法已成功在油菜、拟南芥、大豆等植物的泛基因组构建中应用[9-10, 19]，并组装到质量极高的泛基因组。

除了新测序技术的应用，组装策略的选择也是泛基因组组装的重要问题。基于从头组装策略组装的泛基因组完整性和质量都很高，但较高的费用使其难以在大规模群体中实现，因此可能导致所构建的泛基因组多样性不足、稀有变异缺失等问题。以参考基因组为基础的迭代组装或 map-to-pan 组装，主要利用二代测序数据进行组装，费用低，可以通过对大量个体分析来提高泛基因组的多样性，充分捕获群体内的稀有变异，但泛基因组的完整性和质量不高。这两种策略在某程度上是互补的。选择群体中少数代表性材料进行基因组从头组装，以此为基础构建泛基因组框架，结合大规模多样性材料的二代测序数据进行比对和迭代组装，提高稀有变异的捕获效率，这是目前构建高质量泛基因组的一种经济、有效且可行的策略[6]。

3　泛基因组的应用

3.1　泛基因组在植物进化与驯化研究中的应用

植物进化与驯化研究是植物基因组学研究的重要课题之一。解析植物在进化以及驯化过程中基因组结构变异的演变，有助于深入理解和剖析植物进化和驯化的过程，为植物育种和遗传变异发掘提供线索和理论指导。泛基因组在群体基因组结构变异的捕获和分析方面具有单一参考基因组不能比拟的优势，可以为植物进化与驯化研究提供更完整、更宽泛的基础数据[10]。

以大豆为例，大豆的种皮颜色是大豆驯化的一个重要性状，研究表明从野生大豆的黑色种皮演变成大多数栽培大豆的黄色种皮，是由于一个查尔酮合成酶基因发生结构变异，导致该基因沉默造成的。Liu 等[10]通过对大豆野生资源和现代品种构建的泛基因组进行研究，发现了一种全新的导致黄色种皮的基因结构变异，进一步通过泛基因组进行单倍体型研究和进化分析，对控制黄色种皮的不同单倍体型驯化时间和顺序进行了研究，剖析了大豆种皮颜色驯化的历史。另一个例子来自于向日葵的泛基因组研究。Sariel 等通过构建和分析向日葵泛基因组，发现现代栽培品种的基因组中约有 10% 的基因组来源于野生资源的基因组渗入，进一步通过泛基因组进行同源基因比较分析发现，大量用于改良现代品种抗病性的功能基因来源于野生资源的抗病基因渗入，从而印证了现代向日葵品种抗病性改良的基因来源和历史[25]。

3.2　泛基因组在遗传变异发掘与功能基因克隆中的应用

从种质资源中鉴定和利用控制重要农艺性状的优异遗传变异，是植物遗传改良的基础。常规的遗传变异鉴定方法主要通过比较具有不同表型的材料的基因组结构，鉴定控制相关表型的遗传变异。新一代测序技术以及 GWAS 等遗传分析技术的出现为遗传变异的分析提供了高效的技术手段。然而，这些技术目前大多以单一参考基因组为基础。由于其呈现的基因组变异和多样性有限，单一参考基因组不适用于由基因组结构变异或参考基因组上不存在的功能基因引起的表型的遗传分析。如水稻的抗病基因 *Xa21* 只存在于野生稻基因组中[38]，基于单一的栽培稻参考基因组的重测序、同源基因克隆、GWAS 等技术对这类基因进行克隆几乎不可行，只能回归传统的图位克隆方法。泛基因组可以通过完整捕获和呈现群体内全部的遗传变异，为这些远源种质资源的遗传变异发掘提供更完整的基因组参考消息，为重测序、GWAS 等技术应用于这类遗传变异的分析提供基础数据和框架（图 2）。

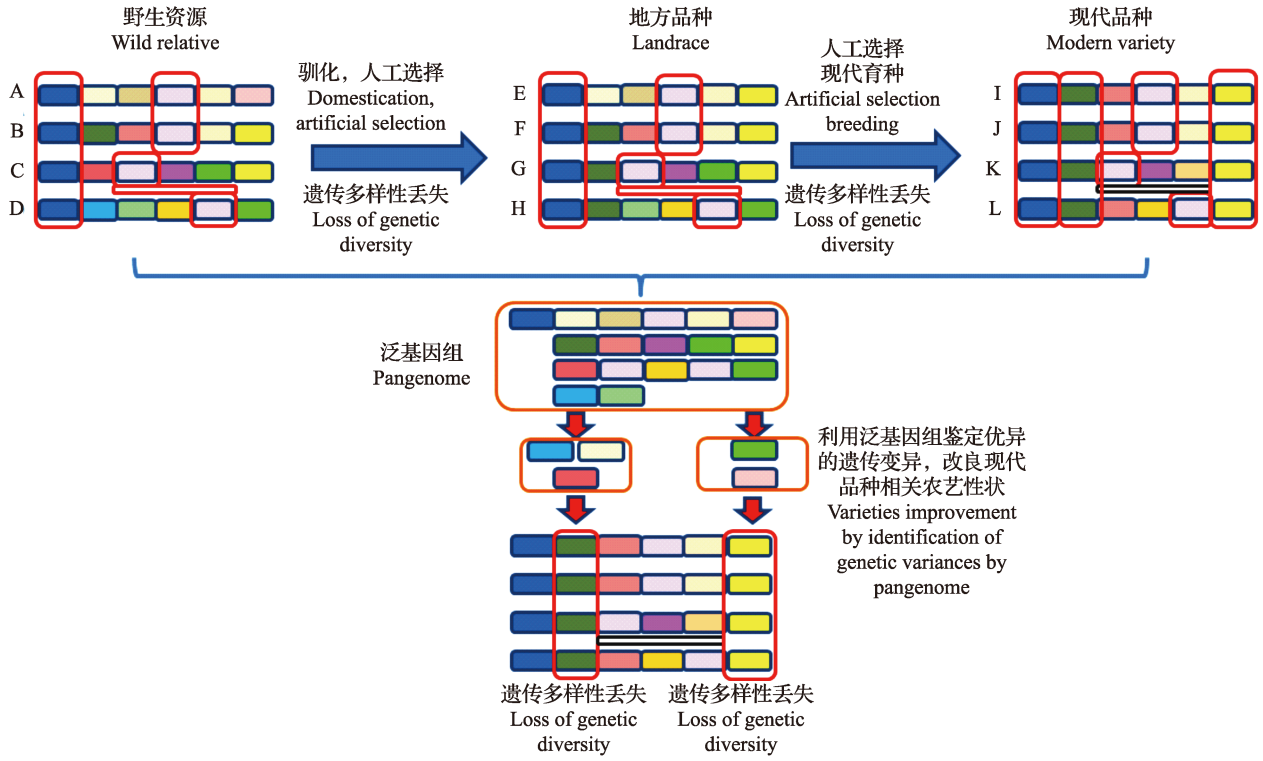

图 2　泛基因组在遗传变异发掘和植物分子育种中应用

Fig.2　Application of pangenome in identification of genetic variations and plant molecular breeding

在具体应用过程中，泛基因组主要作为参考基因组，矫正单一参考基因组在基因组分析过程中导致的偏差。在高通量测序数据分析过程中，多样性或者远源资源的测序数据比对上单一参考基因组会出现系统性的偏差，大量与参考基因组相比具有高度多态性或者参考基因组缺失的序列的读序数据会被丢弃，导致相应的变异区域分析结果出现偏差。从群体角度看，由于泛基因组保留了群体完整的基因组多样性，在测序数据分析中作为参考基因组，可以有效避免高度多态性的读序数据丢失问题，鉴定到单一参考基因组无法鉴定的基因组变异。目前已开发了多个基于泛基因组的新一代测序数据比对方法和工具[39]，而随着更多这样的算法和工具的开发，基于泛基因组的测序数据分析将成为未来功能基因组学研究的重要方法之一。

在 GWAS 分析方面，泛基因组可以有效提高 GWAS 的准确性[40]。常规 GWAS 是把个体材料与参考基因组比较得到的 SNP 与表型进行关联。然而当 GWAS 分析那些来源于参考基因组不存在的功能基因导致的表型的时候，会出现 GWAS 定位区间与实际功能基因之间偏差较大甚至检测不到的情况。例如用 GWAS 鉴定玉米抗甘蔗花叶病毒功能基因的时候，如果以玉米 B73 参考基因组鉴定的 SNP 做关联分析，检测到的最显著的 SNP 与实际功能基因相差达到 8.5 Mb[41]。以泛基因组为参考基因组，把结构变异作为一种变异类型，与 SNP 一起作为基因型数据用于 GWAS 分析，可以有效解决这类因单一参考基因组而导致的 GWAS 结果偏差或错误问题[42]。例如通过以油菜泛基因组鉴定到的 PAV 为基因型进行 GWAS 分析，可以直接鉴定到导致表型的基因组上一段 3.9 kb 的插入变异[9]。

3.3　泛基因组在植物育种中的应用

泛基因组中的非必需基因组与很多重要的农艺性状相关联，特别是对非生物胁迫和生物胁迫的抗性。因此通过泛基因组对非必需基因进行分析，对鉴定抗性功能基因用于植物遗传改良具有重要的应用价值[6]。

另一方面，大量研究表明，大结构变异常常会阻止重组的发生，从而导致从野生资源中引入某些优异功能基因的同时，功能基因附近的野生资源基因组片段会保留下来，且难以通过杂交和遗传重组清除。这种连锁累赘大大增加了育种家选择远缘种质资源作为亲本，通过杂交进行遗传改良的难度[6]。在广泛多样性种质资源中进行基因组分析，选择具有合适基因型背景的亲本作为育种材料，可以避免或打破这种连锁累赘[43]。泛基因组适用于这类广

泛多样性群体的基因组结构变异分析，通过对种质资源群体的泛基因组进行分析，育种家可以更高效和准确的选择合适的育种亲本和育种方法来解决上述连锁累赘问题。

4　泛基因组研究前瞻

泛基因组研究进展非常迅速，然而目前泛基因组领域还亟需解决以下两方面的问题：首先，如何更好地把最新的测序技术、生物信息学技术乃至数据技术应用于泛基因组组装，更高效地构建高质量且具有广泛代表性的泛基因组。三代测序技术和基因组组装技术在泛基因组组装中的应用，提高了泛基因组的质量。然而如何解决大规模群体的稀有遗传变异捕获问题，提高泛基因组的代表性，还需要综合运用新的组装技术和策略，开发相应的技术流程。另一方面，人工智能技术在泛基因组组装和研究中的应用，可以自主识别泛基因组数据中的一些模式，有可能有助于解决泛基因组组装和功能研究中的某些挑战[44]。

第二，如何更好地解析泛基因组的变异数据，并进行重要农艺性状相关遗传变异的鉴定，是泛基因组应用研究的最大挑战。构建合适的数据框架，更有效且更有逻辑地通过泛基因组呈现群体内的结构变异，并使这些变异信息更适合生物信息学算法处理，是泛基因组应用研究的核心框架。其中基于图论（graph）的泛基因组数据结构化和可视化技术是目前这方面研究的热点[36,45]。最新的大豆泛基因组的构建和分析已经利用相关技术构建了基于图（graph-based）的泛基因组和变异图谱[10]。然而这个领域目前还尚在发展初期，相应的技术和算法尚未成熟，另外，利用泛基因组图（pangenome graph）作为参考基因组应用于生物信息学分析，相对于线性参考基因组，需要更复杂的计算和更多的计算资源，这是目前泛基因组图在应用上的主要困难之一[35]。开发完整的泛基因组数据分析流程和工具，使泛基因组数据与表型数据能有效整合，是泛基因组应用于遗传变异发掘研究的技术支撑，然而目前这部分研究还很少，相应的理论和流程都不成熟。因此，泛基因组变异数据结构以及相适应的遗传变异发掘流程和工具的开发是未来泛基因组应用研究的重点之一。

综上所述，随着过去 10 年基因组学技术飞速发展和基因组数据大量涌现，单一参考基因组已经越来越不适合现代功能基因组学研究的需要，从基于单一参考基因组向基于泛基因组的研究范式转变，是功能基因组学未来发展的必然趋势。泛基因组作为一种全新的基因组学概念，可以为功能基因组学研究提供更完整的基础变异数据和更完善的分析框架，构建和利用高质量、多样性的泛基因组将会推动植物功能基因组学研究的进一步发展。

参考文献

[1] Tao Y F, Jordan D, Mace E S. Crop genomics goes beyond a single reference genome. Trends in Plant Science, 2019, 24(12): 1072-1074

[2] Wang W S, Mauleon R, Hu Z Q, Chebotarov D, Tai S, Wu Z C, Li M, Zheng T Q, Rommel F R, Zhang F, Mansueto L, Copetti D, Sanciangco M, Christian P K, Xu J L, Sun C, Fu B, 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 X, Li J, Zhang D B, Lee S H, Hu W S, Poliakov A, Dubchak I, Victor J U, Frances N B, John R M, 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 W, Li J Y, Sackville H R, Wing R, Ruan J, Zhang G Y, Wei C C, Alexandrov N, McNally K, Li Z K, Leung H. Genomic variation in 3, 010 diverse accessions of Asian cultivated rice. Nature, 2018, 557: 43-49

[3] Alonge M, Wang X G, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, Levy Y, Harel T H, Shalev-Schlosser G, Amsellem Z, Razifard H, Caicedo A L, Tieman D M, Klee H, Kirsche M, Aganezov S, Ranallo-Benavidez T R, Lemmon Z H, Kim J, Robitaille G, Kramer M, Goodwin S, McCombie W R, Hutton S, Van Eck J, Gillis J, Eshed Y, Sedlazeck F J, van der Knaap E, Schatz M C, Lippman Z B. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell, 2020, 182(1): 145-161

[4] 李娜，尚建立，周丹，李楠楠，王吉明，马双武．一个获得与缺失变异（PAV）调控甜瓜果实苦味．植物遗传资源学报，2020，21(2)：377-385
Li N, Shang J L, Zhou D, Li N N, Wang J M, Ma S W. A presence-absence variation regulates fruit bitterness in melon (*Cucumis melo* L.). Journal of Plant Genetic Resources, 2020, 21(2): 377-385

[5] Tranchant-Dubreuil C, Rouard M, Sabot F. Plant pangenome: impacts on phenotypes and evolution. Annual Plant Reviews, 2019, 2: 1-25

[6] Tao Y F, Zhao X R, Mace E, Henry R, Jordan D. Exploring and exploiting pan-genomics for crop improvement. Molecular Plant, 2019, 12: 156-169

[7] Tettelin H, Masignani V, Cieslewicz M J, Donati C, Medini D, Ward N L, Angiuoli S V, Crabtree J, Jones A L, Durkin A S, DeBoy R T, Davidsen T M, Mora M, Scarselli M, Ros I M Y, Peterson J D, Hauser C R, Sundaram J P, Nelson W C, Madupu R, Brinkac L M, Dodson R J, Rosovitz M J, Sullivan S A, Daugherty S C, Haft D H, Selengut J, Gwinn M L, Zhou L W, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K,

O'Connor K J B, Smith S, Utterback T R, White O, Rubens C E, Grandi G, Madoff L C, Kasper D L, Telford J L, Wessels M R, Rappuoli R, Fraser C M. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences, 2005, 102(39): 13950-13955

[8] Zhao J L, Bayer P E, Ruperao P, Saxena R K, Khan A W, Golicz A A, Nguyen H T, Batley J, Edwards D, Varshney R K. Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). Plant Biotechnology Journal, 2020, 18(9): 1946-1954

[9] Song J M, Guan Z L, Hu J L, Guo C C, Yang Z Q, Wang S, Liu D X, Wang B, Lu S P, Zhou R, Xie W Z, Cheng Y F, Zhang Yu T, Liu K D, Yang Q Y, Chen L L, Guo L. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nature Plants, 2020, 6: 34-45

[10] Liu Y C, Du H L, Li P C, Shen Y T, Peng H, Liu S L, Zhou G A, Zhang H K, Liu Z, Shi M, Huang X H, Li Y, Zhang M, Wang Z, Zhu B G, Han B, Liang C Z, Tian Z X. Pan-genome of wild and cultivated soybeans. Cell, 2020, 182(1): 162-176

[11] Tian X M, Li R, Fu W W, Li Y, Wang X H, Li M, Du D, Tang Q Z, Cai Y D, Long Y M, Zhao Y, Li M Z, Jiang Y. Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data. Science China Life Sciences, 2020, 63(5): 750-763

[12] Sherman R M, Forman J, Antonescu V, Puiu D, Daya M, Rafaels N, Boorgula M P, Chavan S, Vergara C, Ortega V E, Levin A M, Eng C, Yazdanbakhsh M, Wilson J G, Marrugo J, Lange L A, Williams L K, Watson H, Ware L B, Olopade C O, Olopade O, Oliveira R R, Ober C, Nicolae D L, Meyers D A, Mayorga A, Knight-Madden J, Hartert T, Hansel N N, Foreman M G, Ford J G, Faruque M U, Dunston G M, Caraballo L, Burchard E G, Bleecker E R, Araujo M I, Herrera-Paz E F, Campbell M, Foster C, Taub M A, Beaty T H, Ruczinski I, Mathias R A, Barnes K C, Salzberg S L. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nature Genetics, 2019, 51: 30-35

[13] Li Y H, Zhou G Y, Ma J X, Jiang W K, Jin L G, Zhang Z H, Guo Y, Zhang J B, Sui Y, Zheng L T, Zhang S S, Zuo Q Y, Shi X H, Li Y F, Zhang W K, Hu Y Y, Kong G Y, Hong H L, Tan B, Song J, Liu Z X, Wang Y S, Ruan H, Yeung C K L, Liu J, Wang H L, Zhang L J, Guan R X, Wang K J, Li W B, Chen S Y, Chang R Z, Jiang Z, Jackson S A, Li R Q, Qiu L J. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nature Biotechnology, 2014, 32: 1045-1052

[14] Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. Nature Methods, 2012, 9(4): 357-360

[15] Golicz A A, Bayer P E, Bhalla P L, Batley J, Edwards D. Pangenomics comes of age: from bacteria to plant and animal applications. Trends in Genetic, 2019, 36(2): 132-145

[16] Carlos G L, Benevides J L, Vinicius C V M, Silva A, Thiago J R R, de Castro S S, Azevedo V. Inside the pan-genome-methods and software overview. Current Genomics, 2015, 16(4): 245-252

[17] Vernikos G, Medini D, Riley D R, Tettelin H. Ten years of pan-genome analyses. Current Opinion in Microbiology, 2015, 23: 148-154

[18] Morgante M, De Paoli E, Radovic S. Transposable elements and the plant pan-genomes. Current Opinion in Plant Biology, 2007, 10(2): 149-155

[19] Jiao W B, Schneeberger K. Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. Nature Communications, 2020, 11: 989

[20] Gordon S P, Contreras-Moreira B, Woods D P, Des Marais D L, Burgess D, Shu S Q, Stritt C, Roulin A C, Schackwitz W, Tyler L, Martin J, Lipzen A, Dochy N, Phillips J, Barry K, Geuten K, Budak H, Juenger T E, Amasino R, Caicedo A L, Goodstein D, Davidson P, Mur L A J, Figueroa M, Freeling M, Catalan P, Vogel J P. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nature Communications, 2017, 8: 2184

[21] Golicz A A, Bayer P E, Barker G C, Edger P P, Kim H R, Martinez P A, Chan C K K, Severn-Ellis A, McCombie W R, Parkin I A, Paterson A H, Pires J C, Sharpe A G, Tang H B, Teakle G R, Town C D, Batley J, Edwards D. The pangenome of an agronomically important crop plant Brassica oleracea. Nature Communications, 2016, 7: 13390

[22] Hurgobin B, Golicz A A, Bayer P E, Chan C K K, Tirnaz S, Dolatabadian A, Schiessl S V, Samans B, Montenegro J D, Parkin I A P, Pires J C, Chalhoub B, King G J, Snowdon R, Batley J, Edwards D. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. Plant Biotechnology Journal, 2018, 16(7): 1265-1274

[23] Lin K, Zhang N W, Severing E I, Nijveen H, Cheng F, Visser R G F, Wang X W, de Ridder D, Bonnema G. Beyond genomic variation-comparison and functional annotation of three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics, 2014, 15: 250

[24] Ou L J, Li D, Lv J H, Chen W C, Zhang Z Q, Li X F, Yang B Z, Zhou S D, Yang S, Li W G, Gao H Z, Zeng Q, Yu H Y, Ouyang B, Li F, Liu F, Zheng J Y, Liu Y H, Wang J, Wang B B, Dai X Z, Ma Y Q, Zou X X. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses. The New Phytologist, 2018, 220(2): 360-363

[25] Hübner S, Bercovich N, Todesco M, Mandel J R, Odenheimer J, Ziegler E, Lee J S, Baute G J, Owens G L, Grassa C, Ebert D P, Ostevik K L, Moyers B T, Yakimowski S, Masalia R R, Gao L, Ćalić I, Bowers J E, Kane N C, Swanevelder D Z H, Kubach T, Muños S, Langlade N B, Burke J M, Rieseberg L H. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nature Plant, 2019, 5: 54-62

[26] Zhou P, Silverstein K A T, Ramaraj T, Guhlin J, Denny R, Liu J Q, Farmer A D, Steele K P, Stupar R M, Miller J R, Tiffin P, Mudge J, Young N D. Exploring structural variation and gene family architecture with De Novo assemblies of 15 Medicago genomes. BMC Genomics, 2017, 18: 261

[27] Schatz M C, Maron L G, Stein J C, Wences A H, Gurtowski J, Biggers E, Lee H, Kramer M, Antoniou E, Ghiban E, Wright M H, Chia J M, Ware D, McCouch S R, McCombie W R. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. Genome Biology, 2014, 15: 506

[28] Zhao Q, Feng Q, Lu H Y, Li Y, Wang A H, Tian Q L, Zhan Q L, Lu Y Q, Zhang L, Huang T, Wang Y C, Fan D L, Zhao Y, Wang Z Q, Zhou C C, Chen J Y, Zhu C R, Li W J, Weng Q J, Xu Q, Wang Z X, Wei X H, Han B, Huang X H. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nature Genetics, 2018, 50: 278-284

[29] Pinosio S, Giacomello S, Faivre-Rampant P, Taylor G, Jorge V, Le Paslier M C, Zaina G, Bastien C, Cattonaro F, Marroni F, Morgante M. Characterization of the poplar pan-genome by genome-wide identification of structural variation. Molecular Biology and Evolution, 2016, 33(10): 2706-2719

[30] Yu J Y, Golicz A A, Lu K, Dossa K, Zhang Y X, Chen J F, Wang L H, You J, Fan D D, Edwards D, Zhang X R. Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. Plant Biotechnology Journal, 2019, 17(5): 881-892

[31] Gao L, Gonda I, Sun H H, Ma Q Y, Bao K, Tieman D M, Burzynski-Chang E A, Fish T L, Stromberg K A, Sacks G L, Thannhauser T W, Foolad M R, Diez M J, Blanca J, Canizares J, Xu Y M, van der Knaap E, Huang S W, Klee H J, Giovannoni J J, Fei Z J. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics, 2019, 51: 1044-1051

[32] Montenegro J D, Golicz A A, Bayer P E, Hurgobin B, Lee H T, Chan C K K, Visendi P, Lai K, Doležel J, Batley J, Edwards D. The pangenome of hexaploid bread wheat. The Plant Journal, 2017, 90(5): 1007-1013

[33] Hirsch C N, Foerster J M, Johnson J M, Sekhon R S, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza M A, Barry K, de Leon N, Kaeppler S M, Buell C R. Insights into the maize pan-genome and pan-transcriptome. The Plant Cell, 2014, 26: 121-135

[34] Danilevicz M F, Fernandez C, Geraldine T, Marsh J I, Bayer P E, Edwards D. Plant pangenomics: approaches, applications and advancements. Current Opinion in Plant Biology, 2020, 54: 18-25

[35] Eizenga J M, Novak A M, Sibbesen J A, Heumos S, Ghaffaari A, Hickey G, Chang X, Seaman J D, Rounthwaite R, Ebler J, Mikko R, Shilpa G, Benedict P, Tobias M, Jouni S, Erik G. Pangenome graphs. Annual Review of Genomics and Human Genetics, 2020, 21: 139-162

[36] Michael T P, VanBuren R. Building near-complete plant genomes. Current Opinion in Plant Biology, 2020, 54: 26-33

[37] Jiao W B, Schneeberger K. The impact of third generation genomic technologies on plant genome assembly. Current Opinion in Plant Biology, 2017, 36: 64-70

[38] Song W Y, Wang G L, Chen L L, Kim H S, Pi L Y, Holsten T, Gardner J, Wang B, Zhai W X, Zhu L H, Fauquet C, Ronald P. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science, 1995, 270(5243): 1804-1806

[39] Anari S S, de Ridder D, Schranz M E, Smit S. Pangenomic read mapping. bioRxiv, DOI: https: //doi.org/10.1101/813634

[40] Peter J, De Chiara M, Friedrich A, Yue J X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury J M, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J. Genome evolution across 1, 011 *Saccharomyces cerevisiae* isolates. Nature, 2018, 556: 339-344

[41] Gage J L, Vaillancourt B, Hamilton J P, Manrique-Carpintero N C, Gustafson T J, Barry K, Lipzen A, Tracy W F, Mikel M A, Kaeppler S M, Buell C R, de Leon N. Multiple maize reference genomes impact the identification of variants by genome-wide association study in a diverse inbred panel. The Plant Genome, 2019, 12: 1-12

[42] Hurgobin B, Edwards D. SNP discovery using a pangenome: has the single reference approach become obsolete? Biology, 2017, 6(1): 21

[43] Liu W Q, Fan Y Y, Jie C, Shi Y F, Wu J L. Avoidance of linkage drag between blast resistance gene and the QTL conditioning spikelet fertility based on genotype selection against heading date in rice. Rice Science, 2009, 16(1): 21-26

[44] Gao S, Wu J R, Stiller J, Zheng Z, Zhou M X, Wang Y G, Liu C J. Identifying barley pan-genome sequence anchors using genetic mapping and machine learning. Theoretical and Applied Genetics, 2020, 133: 2535-2544

[45] Marcus S, Lee H, Schatz M C. SplitMEM: a graphical algorithm for pan-genome analysis with suffix skips. Bioinformatics, 2014, 30(24): 3476-3483