

· 综述 ·

植物遗传资源核心种质研究现状与展望^①

李自超^② 张洪亮 孙传清 王象坤

(中国农业大学作物学院)

摘要 近年来急剧增加的遗传资源数量给种质资源的保存、研究与利用带来了很大困难。为解决这一矛盾, Frankel(1984)和 Brown(1989)提出并发展了核心种质(Core Collection)的概念,即以最少的遗传资源份数最大限度的代表该物种的遗传多样性;至今全世界在硬粒小麦、花生和一年生苜蓿等 30 余种植物上已经进行了核心种质研究;本文对有关核心种质构建的理论依据、开展核心种质研究的步骤、方法和内容以及核心种质的检验指标作了全面系统的阐述;此外,作者在 Frankel 的核心种质的概念的基础上,提出核心种质还应包含有实际应用价值的优异基因,并强调核心种质应该是一个动态变化的群体。

关键词 植物遗传资源; 核心种质; 遗传多样性

分类号 S5-02

Status and Prospects of Core Collection in Plant Germplasm Resource

Li Zichao Zhang Hongliang Sun Chuanqing Wang Xiangkun

(College of Crop Sciences, CAU)

Abstract Recent years more and more difficult was met in preserving, reseaching and utilizing plant germplasm resources because of rapid increase of volume of plant germplasm resource. In order to solve the problems Frankel (1984) and Brown(1989) proposed and developed concept of core collection, i. e. a minimum of accessions may represent a maximum of the genetic diversity contained in the whole collection; More than 30 species, including peanut, durum wheat, annual *medicago* and so on, have been established core collection all over the world so far; Theoretical basis of establishing core collection, method, procedure and content of core collection research, as well as the test parameters of core collection were reviewed and discussed in this paper. Based on concept of Frankel, the author proposed that core collection should include elite genes, and core collection has to be a population with dynamics of variation.

Key words plant germplasm resource; core collection; genetic diversity

人类对植物遗传资源的认识及考察、收集、利用、创造等有关研究经过了漫长的历史过程,至今已进入到第三个阶段。第一阶段,随着作物起源中心学说的提出,人们对植物遗传资源的生物地理、分类、进化等问题进行了研究。第二阶段,二战后,人们面临植物遗传资源大量流失的严峻问题,加强了对遗传资源的收集与保存,在此期间,各国都收集并保存了大量的遗传材料,使保存的遗传资源增加到 450 万份^[1]。那么,怎样对如此众多的遗传资源进行保存、评价、创新和及有效利用呢? 这是第三阶段也是目前植物遗传资源研究的主要内容。

收稿日期: 1998-02-24

①为国家自然科学基金 39770450 和云南省省院省校合作项目及“973”项目资助课题

②李自超,北京圆明园西路 2 号中国农业大学(西校区),100094

核心种质的概念由 Frankel(1984)与 Brown(1989a)提出并发展^[2~4],根据国际植物遗传资源委员会的调查发现^[4],国际上一些农业研究中心和一些发达国家很重视对核心种质的研究。1992年在巴西召开了关于核心种质的国际会议,会上就核心种质的概念,建立核心种质的步骤以及今后核心种质研究的方向进行了讨论。中国于1994年在杭州召开了核心种质专题会议。至今,全世界在包括野生大豆^[4]、花生^[6]、芝麻^[7]等30多种植物上已经或正在建立核心种质(表1)。

表1 研究较深入的植物核心种质一览表

植物种类	国家	作者	时间	原始群体	核心种质	总体比例/%	备注
多年生大豆	澳大利亚	Brown, et al.	1987	1 400	111	稍低于10	12个种取样的比例不同
秋葵	西非	Hamon, et al.	1989	2 283	189	8	包括一些珍稀类型
巴西木薯	巴西	Cordeiro	1990	4 132	1 200	29	
咖啡		Hamon, et al.	1993	338	88	26	运用了主成分分析
多年生苜蓿	美国	Basigalup, et al.	1995	1 100	200	18	
一年生苜蓿	美国	Diwan, et al.	1995	3 159	209~218	6.6~7	进行了取样方法探索
一年生苜蓿	美国	Diwan, et al.	1994	3 159	211	6.7	用占原始群体40%的初级群体代表原始群体
野生大豆	哥伦比亚	Tohme	1996	975	114	11.7	应用了AFLP分子标记
多年生黑麦草		Casler	1995		375		
多年生黑麦草	法国	Charmet, et al.	1993	550	112	20.4	
花生	美国	Corley Holbrook, et al.	1995	7 432	831	稍大于10	
硬粒小麦	美国	Spagnoletti, et al.	1993	3 038	约500	16	对取样策略有较为详细的论述
大麦	荷兰	Hintum, et al.	1994	153	10	6.5	应用系谱分析对现代品种进行筛选
栽培大麦	荷兰	Hintum	1995	96			对据收集地点与数量和质量性状聚类作了比较
扁豆	叙利亚、美国	Erskine, et al.	1991	109	10	10	进行形态、同工酶位点多态性分析
奎奴亚藜	丹麦	Ortiz, et al.	1998	1 029	103	10	
墨西哥菜豆	美国	Paul, et al.	1998	24 000	90		
多年生黑麦草	法国	Charmet	1995	547		5或10	
芝麻	中国	沈金雄等	1995	4 251	884	约20	仅获得核心种质待选品
小麦特殊遗传资源	中国	范传珠等	1994	993	500	约50	

1 核心种质概念与研究内容

1.1 核心种质概念的提出

丰富的遗传资源为遗传研究及育种工作提供了大量的材料,然而,如此众多的资源给保存、评价、鉴定及利用带来了困难^[7],人们开始寻求解决这一矛盾的方法。Harlan(1972)曾提出为方便资源的保存、评价与利用而建立整个资源库的亚库,称为活动收集品^[8](active collection)。澳大利亚 Frankel 于1984年提出了核心种质(core collection)的概念,并与 Brown(1989a)将其进一步发展。它是用一定的方法选择整个种质资源的一部分,以最小的资源数量

和遗传重复最大程度地代表整个遗传资源的多样性,未包含于核心种质中的种质材料并不被遗弃,而是作为保留种质(reserve collection),从而方便于种质的保存、评价与利用。他同时指出了建立核心种质最重要和一般的标准,其组成应表现和包括当前多样性的主要和大部分类型。概括起来核心种质应具有以下特征:

①异质性:核心种质是从现有遗传资源中选出的数量有限的一部分材料^[9],彼此间在生态和遗传上的相似性尽可能小,最大限度地去除遗传上的重复。

②多样性和代表性:核心种质应代表本物种及其近缘种主要的遗传组成和生态类型,包括了本物种的尽可能多的生态和遗传多样性^[10],而不是全部收集品的简单代表和压缩^[11]。

③实用性:由于核心种质的规模急剧减小,与备份的保留种质间存在着极为密切的联系,因此,极大地方便了对种质资源的保存、评价与创新利用,更容易找到所需特性或特性组合的特殊资源材料^[4]。

④动态性:核心种质是满足当前及未来遗传研究和育种目标需要的重要的材料来源,因此,应该在核心种质与保留种质之间保持材料上的动态交流与调整。

1.2 核心种质研究的步骤及内容

Hodgkin(1993)提出建立核心种质基本包含4个步骤^[12]:

①数据收集整理:建立核心种质的第一步是收集整个种质中现有的数据,包括种质库现有的基本数据,尽可能多的性状评价鉴定数据,特征数据,如同工酶、种子蛋白等生物化学、分子标记数据等。

②收集数据分组:根据现有的数据,把具有相似特点的种质材料分组,比如可以根据分类学、地理起源、生态分布、遗传标记和农艺性状等数据来进行。

③样品选择:把整个种质科学地分组后,以合理的取样方法及取样比例选取核心种质。

④核心种质的管理:选出一套核心种质后,要建立完善的繁种、供种及管理体制,以保证核心种质的有效利用。

李自超等通过对中国稻种资源核心种质的研究提出,首先利用现有的基本数据、特征数据和鉴定评价数据等进行取样策略研究,选择合理的取样策略提取核心种质初级样品;然后对初级样品进行种植及详尽的室内外考种鉴定等,并从形态、生化及DNA分子标记等水平上进行聚类压缩,最终成为中国稻种资源的核心种质。

2 构建核心种质的理论依据

有三方面的论点支持核心种质的建立并对核心种质的构建有指导意义^[4]。首先从统计取样上考虑,如果核心种质中包括了整个资源中等位基因丰度的绝大部分,而仅有较少数的变异丢失,那么核心种质就会对整个遗传资源具有令人信服的代表性。其次,从整个植物群体尤其种质资源的遗传结构上考虑,核心种质可以代表整个遗传资源的大多数遗传多样性。假设育种者在需要时,可以通过杂交选择(现在可以通过基因工程)来恢复所需要的等位基因,那么,从理论上讲人们只需保存等位基因的一个拷贝就可以了。另外,核心种质的构建有助于对种质资源的有效管理与应用。

种质资源中等位基因分为一般广泛性(common, widespread)、一般局域性(common, localised)、稀有广泛性(rare, widespread)和稀有局域性(rare, localised)4种类型。一般广泛性基

因在取样时几乎确定包括在核心种质中,不用进一步考虑。稀有局域性基因要包括于核心种质中有一定的偶然性。对于稀有广泛性应放入中性理论中考虑。而一般局域性基因则放入另一模型中考虑,这种模型考虑了植物遗传多样性的非均等分布,即遗传多样性理论模型^[13]。

2.1 中性突变理论模型

根据中性突变理论,在一个无限大的群体(N_e)中,若基因的突变率为 u ,则预期基因频率位于 p 与 q ($0 < p < q < 1$)之间的等位基因数(n_s)为

$$n_s = \theta \int_p^q (1-x)^{\theta-1} x^{-1} dx$$

据公式,在 θ ($\theta = 4N_e u$)一定的情况下,当核心种质样品数为1 000~3 000时所包含的不同等位基因的数目的增加便产生了边缘效应。在一个含有3 000份样品的样本中,含有基因频率在 $p=0.0001$ 内的等位基因数目已经与整个资源在同样水平上的等位基因数目非常接近。在大约95%置信区间内,在 $\theta=1.0$ 的变异水平下,在一个占原始群体10%的核心种质中,对于任何单一基因位点,我们有95%的把握在样本中至少包含原始群体所有等位基因数的20%,即与原始群体相比其等位基因丰度增加1倍;而对大多数位点来说则会包含的更多,如对于100个相等的多态位点来说,至少包含了其中的70%,其丰度增加了6~7倍。当样本规模小于10%时,随规模的减小等位基因丢失的速率明显增加;当规模再大后,则随规模的减小等位基因的丢失速率趋缓。因此可以得出结论,对于广泛稀有基因来说,当我们预期要在样本中保留一定数目的等位基因 n_s 时,可以通过选择合适的样本数 N_r 来实现。

2.2 遗传多样性的分层结构模型

Brown的中性突变理论模型仅用于广泛分布的等位基因。事实上,资源中变异的分布是不均匀的,往往某些变异局限于资源中的一小部分,然而,这些广泛而局域性基因是应用上非常重要的一类,它包括某些病虫抗源和特殊生态类型。如果按随机取样,由这种某些变异局部集中分布的资源中所得的核心种质多样性水平会降低。并且有人证明,某些同工酶位点也不是中性的(Morishima, 1991)。因为种质中的材料间并非随机杂交,因此,平衡状态的假定也是不成立的。高度局域化的基因包括于核心种质中的可能性较小,可以在设计核心种质时将整体数量保持在可接受的范围内而将这一部分基因成分最大化。取样时,通过保证将主要的地理、生态区及形态不连续性包括在内,可以在核心种质中将这一部分基因包括进来。当然,某些在生产上非常需要的优良基因可能仅存在于一二个样品中,在核心种质中难免漏掉,人们仍不得不到保留种质中去寻找。然而,在我们为增加资源在育种上的利用而构建核心种质的策略中,这种极端的情况不能作为考虑的主要问题^[4]。因此对于局域化分布的等位基因应放于Yonezawa等所建议的模型中讨论。

在遗传多样性分层结构理论模型中,主要考虑了资源群体的遗传结构,同时考虑了种质的复壮(rejuvenation)扩繁中遗传结构的改变和各种投入的多少,认为核心种质的有效性(efficiency of the core collection)与遗传冗余度(degree of genetic redundancy)、等位基因分布、投入参数及种质保留时间有关,对于遗传冗余度在0.2~0.9之间的群体,其最佳取样比例为0.2~0.3。

3 核心种质的构建

在核心种质的研究中,最为关键的是怎样构建满足要求的核心种质。包括对构建核心种质

所用数据的选择,数据的处理,材料的分组及组内取样方法的确定等。

3.1 遗传多样性的评估方法

核心种质要求将遗传多样性最大化,因此多样性的评估对于核心种质的取样策略及检验都很关键。遗传多样性和遗传分化是遗传多样性的不同表达方式,遗传多样性可以理解为组内遗传物质的差异程度;遗传分化则表示组间遗传物质的差异程度。另外,遗传相似性也是遗传多样性研究的基础。植物的遗传多样性表现在多个方面(如系谱关系、基因多样性和杂合度等)并由多个水平(如形态多样性、地理生态多样性、系统分化与发育及酶与分子水平的多样性等)反映出来。根据所用数据类型不同,提出了不同的植物多样性的评估方法^[14]。

①系谱分析:系谱关系表现了材料间的亲缘关系。从而,可以作为评估遗传多样性的指标,常用个体间的共祖度(r)表示。系谱分析已经用于描述作物的遗传基础、时间发展及预测杂种的表现。

②遗传标记与质量性状:用于多样性研究的遗传标记包括形态标记、储藏蛋白、酶标记和各种分子标记(如 RAPD, RFLP, AFLP, SSR 等)。这些遗传标记及质量性状大多受环境影响较小,有明显的等位基因或性状类别。对这类数据多样性的估计有多种方法。常用量度值有等位基因频率及其分布、Nei's 多样性指数(H)、Shannon and Weaver's 信息指数(I)。常用表征遗传分化的量为 GT_{ST} 。

③数量性状:数量性状的遗传基础复杂,环境因子影响显著。因此数量性状不适用于多样性研究,而应用于鉴定相似的适应性。群体内多样性的量度值通常有标准差(σ)和变异系数(CV)。群体间的分化常用平均数和方差成分分析量度^[15]。

3.2 构建核心种质时对不同数据的利用

用于建立核心种质的数据有3种类型^[4]:基本数据(passport data)、特征数据(characterisation data)和评价鉴定数据(evaluation data)。

①基本数据是指有关材料收集地、起源地的生态地理状况或育种体系、分类体系等有关信息。在核心种质的构建中对基本数据的应用最为广泛,并证明与其他数据结合更为有效^[16~21]。

②特征数据是指包括形态、生化、分子标记在内的表征某材料特征的数据。有些作物种质资源中,部分特征数据是比较齐全的,在构建核心种质的研究中应用也较多^[17~22]。Hamon 等在构建秋葵核心种质时除利用了已有的基本数据外,用了包括多个数量性状(如茎色、叶形、果位等)和质量性状(如株高、开花期、节数等)在内的形态特征数据。生化水平上的同工酶、储藏蛋白质及 DNA 分子标记等一类的特征数据因受环境影响较小,多态性高,检验迅速,而且评价遗传多样性时更直接,更易理解生物的进化^[23],但是由于其化费较高等一些原因,目前在核心种质的研究中应用较少。Perry 等在对大豆核心种质的研究中就采用了正规的同工酶方法进行了多态性分析^[24]。Diwan 在对美国一年生苜蓿核心种质的研究中指出分子标记在评价遗传多样性以及理解物种进化上可能更为直接^[20]。Singh 认为多数品种同工酶和分子标记能够表现一致^[25],1991 年 Perry^[26]又指出某些情况下大豆在形态上表现分化明显的基因型在遗传上非常接近,而形态上表现相似的性状可能同工酶上表现出远的进化关系。贾继增(1996)也指出 AFLP, RAPD, AFLP, SSR 等分子标记技术是检验种质资源遗传多样性的有效工具^[27]。

③评价鉴定数据包括一些产量、品质及抗耐性等在内的农艺性状。在一些鉴定工作做的比较深入的作物上^[3,17~22],建立核心种质时运用了诸如农艺性状、抗性鉴定评价数据,其中产量性状应用较多。

以上对数据的分类仅仅是为了研究上的方便,在实际应用中往往是对种质资源中的数据进行综合利用。在构建一年生苜蓿时,Diwan(1995)比较了11种不同的取样方法^[22],有7种方法综合利用了起源分类等基本数据与特征及评价鉴定数据,另外4种方法则仅仅基于基本数据。研究认为,综合运用特征评价鉴定与基本数据所建立的核心种质比仅仅利用基本数据对整个种质资源具有更大的代表性。Singh et al. (1991)也认为应对种质资源的形态、农艺、生化、分子等性状进行综合评价,这样不同类型数据间可以提供互补的信息。

3.3 数据分析方法

在核心种质的构建中要对大量数据进行分析,如数据标准化、分类、相关分析,检验数据在多样性分析中的有效性等。在数据分析中除一般的平均、方差分析及卡方检验外,需要对大量多维数据采用多变量分析,如聚类分析^[19~22,28]、主成分分析^[21,28]、因子回归^[29]、典型及逐步判别分析^[28,30]和典型变量分析^[30]等。其中常用的有聚类分析、主成分分析和判别分析等。

①聚类分析可有效的将据相似性状的材料归类。Diwan等(1995)在对一年生苜蓿进行核心种质的研究时,用SAS软件包计算了个体间的欧氏距离并做了类平均距离聚类,为产生一具有200~250份占有原始材料10%左右的核心种质,在欧氏距离3.0处将聚类图断开,然后在每个类内采用一定的方法选择所需样品。沈金雄等进行芝麻核心种质构建时主要用聚类分析的方法。李自超(1999)等研究通过聚类分析的聚类取样效果明显优于随机取样^[11]。

②主成分分析对于象自交、无融合生殖及无性繁殖等具有高冗余的群体尤其有效,可以扩大多样性和减少由于性状间相关而造成的遗传冗余。Basigalup等(1995)在构建多年生苜蓿时,用主成分分析与其他方法的比较发现,用全部可供利用的数据所选核心种质反而不如用主成分所选核心种质更有效。

③线性判别分析是应用最广的分类方法之一^[28], Spagoletti等(1993)用典型判别分析的方法根据8个性状将来自美国小谷类作物种质库的约3,000份硬粒小麦分为5个组^[31],这些分组被证明与先前根据多样性中心所分组一致。

根据所研究资源的性质及数据情况可以将3种方法及其他方法结合起来分析,这样各种方法间可以相互弥补或验证,效果会更好。聚类分析用于对材料分类,主成分分析对聚类结果进行补充和验证,判别分析用于验证核心种质分组与原分组的符合情况,三者结合对检验自然界中存在的遗传多样性起到了相辅相成的作用。

3.4 对材料的分组方法

在构建核心种质时必须充分考虑生物多样性的遗传层次结构,将整个材料分为互不重叠的小组,然后在组内选择有代表性的材料。具体分组标准与方法因具体种质特征及数据的有无而异。常见分组标准及方法有按地理及农业生态分组、按分类体系分组、按育种体系分组和大数据组合分组等。

3.4.1 分类体系 植物的分类体系是植物学家在长期对某个植物类群的观察及其进化史研究的基础上得出的,比较科学的反映了植物在自然条件下的遗传结构。因此,对一些由多种组成的遗传资源一般首先按分类体系进行分类,然后再在组内选择。Diwan在研究一年生苜蓿时首先按种将材料分组^[20,22],然后再在每个组内或按来源地、表现型聚类或随机取样模拟出不同方法的核心种质11个,后在3 159份材料中选出211份作为核心种质,经检验在两年间保持稳定并代表了整个资源的变异性。李自超等在研究云南稻种资源取样方案时提出以丁颖分类体系和程王分类体系分组效果较好^[11]。

3.4.2 地理及生态 地理起源可以为多样性提供间接根据,由于对不同环境条件的适应,可以认为来自相同地方的材料与不同地方的材料相比在一定程度上有共同的遗传背景^[31]。来自相似的生态环境的材料同样反映了相似的遗传背景,从而对取样有指导性作用。Erskine等(1991)对来自3个国家的扁豆材料的形态及同工酶位点与地理来源做了比较分析认为,同一来源地内平均等位基因、多态性比例及多样性指数无显著差异,然而形态及多数酶的多态性因地理来源而异,来源地的等位基因频率、异交率显著不同。

3.4.3 植物育种 植物育种使得自然进化及驯化的一些过程处于人为控制下,改变了分化的速率。有些育种技术增加了作物的多样性,如突变、杂交及基因转入等;另外有些方法则减少了多样性,如自交系及杂交种的生产、大多数适应高投入农业的育种等。鉴于这种情况,对于不同材料应根据其特点给予不同的处理方法,即对由不同类型材料组成的资源建议首先按育种体系分类。

3.4.4 单一性状 根据入库资源的编目性状分组,然后对各性状的各相对性状按比例取样,该分组方法对于编目性状较齐全的资源效果较好^[11]。

3.5 取样比例的确定

对材料分析后,取样前首先要确定总体取样比例,对分层取样策略要确定每个组内的取样比例。

3.5.1 总体取样比例 根据中性理论模型,Brown(1989a)指出核心种质一般占整个种质资源的5%~10%,或总量不超过3000份。Diwan和Spafnoletti指出10%以内并不可靠。Yonezawa等(1995)则在多样性分层模型的基础上,指出对于Dr值在0.2~0.9之间的群体,其最佳取样比例为20%~30%。由于生物进化及人类对驯化作物干预的复杂性,对于总体取样比例的确定不能简单化,具体到某种作物应视其遗传多样性而定。李自超等在云南稻种资源核心种质的研究中提出16%和10%的总体取样比例可达到97%以上的保留比例,5%的总体取样比例可达到96%以上的保留比例(待发表)。

3.5.2 分组组内取样比例 组内取样比例应视组内数据情况具体而定,下面将详细阐述有关策略。

3.6 核心种质构建中的取样策略

取样策略是指从资源中提取部分样品作为初级核心种质的取样方法,要达到核心种质的要求,取样策略是关键环节之一。取样方法总体来说分为两种,即随机取样与系统取样,在实际工作中常把二者结合起来。在有些作物核心种质的构建中在随机取样的基础上最后又人为进行了补充^[6,32]。当然在数据信息不完全的情况下,运用人工定向取样可以利用所有可以利用的信息,因此有利于核心种质的构建。但是在构建核心种质时一般种质量都很大,完全人工取样是不可能的。

3.6.1 随机取样 R策略:R策略即完全随机取样策略(completely random sampling strategy),是在整个资源的基础上对所有材料同等对待,完全在整个资源中随机取样。Brown(1989)曾指出,为获得种质资源所期望的最基本的遗传变异,随机取样可以考虑。然而,在实际中很少有人用完全大随机的方法,仅仅在几个研究中作为方法比较时有人用过^[11,19,22,31]。Spagnoletti(1993)在对硬粒小麦核心种质的构建过程中,提出用随机取样可以获得对整个资源的无偏样本。但是,核心种质构建强调的是它对整个遗传资源多样性的代表性及在生产育种中的利用,显然随机的方法对一些在整个资源中占较小比例而变异性较大的材料则缺乏足够的有效性。

Diwan(1995)构建一年生苜蓿核心种质时通过对11种方法的比较认为,与用分组的方法所构建的核心种质相比较,大随机的方法所构建的核心种质对整个种质资源的代表性较差。

3.6.2 系统取样 实际上资源的遗传多样性分布并非是均匀的,不能对所有资源同等对待,应给予不同权重。不同等位基因在总资源中的重复次数也是不象同的,所以对核心种质的选择应采取系统的方法,使得在所构建的核心种质中减少重复高等位基因数,而增加稀有等位基因的比例。因此,系统取样在分组的基础上关键是确定在分组水平上的取样方法,根据种质资源的特性及所拥有的数据不同概括为C,P,L,S,G,H及M共7种策略。

C策略:C策略(constand strategy)指不管每个组有多少材料,在每个组中随机取同样多的材料,作为本组的代表。显然这种策略的适用范围非常小,只有当各个组的材料数量相近,而且其多样性也相近时才能得到比较满意的结果。

P策略:P策略(proportional strategy)指各个组在核心种质中的取样比例与组内的材料总数占整个资源总数的比例一致。当各组材料数量相差很多,且多样性与资源量一致时这一策略比较有效。

L策略:L策略(logarithmic strategy)指组内取样比例由整个组内资源份数的对数值占各组对数值之和的比例来决定。Brown(1989b)认为当原始种质资源分组内遗传变异未知时对数取样的方法是有效的。经对数处理,使得占比例大的组取样比例变小了,而占比例较小的组的取样比例变大了,从而,可以部分的修正核心种质中多样性的偏离。

S策略:S策略(square root strategy)指组内取样比例由整个组内资源份数的平方根值占各组平方根之和的比例来决定。该策略的效果与L策略基本相同^[11]。

G策略:G策略(genetic diversity-dependent strategy)指分组的取样量由组内多样性占整个资源多样性的比例确定。当可以获得种质资源中每个分组的遗传变异或形态多样性的信息时,根据相对多样性来确定整个种质资源及各个分组中的取样比例是最为可靠的方法。

H策略:H策略(H strategy)也是G策略的一种具体形式,它是指每个分组的取样量与组内 $\theta_{ij}=4N_{e(ij)} \cdot v_i$ 的估计值总和的比率成正比。这种方法是由分子标记方法发展来的,但是在理论上可以推广应用由数量遗传变异得到的信息。

M策略:以上策略根据一定标准确定在某组内的取样比例,然后在组内聚类或随机选择。M策略目的在于用一系列遗传标记鉴定既具有高丰度等位基因又成对差异(低冗余)的材料并确定在组内如何合理分配这些材料。这种策略建立于这样一种假设上,将核心种质中的标记等位基因最大化就等于将目标等位基因最大化了。因此称这种策略为M策略(maximisation strategy)。满足条件的组合通常几个,可根据在几个组合中目标等位基因平均保留度(target allele retention averaged)考查其有效性。很显然这种策略需要很大的计算量,需有方便的计算程序。

4 对核心种质有效性的检验

根据核心种质种质的概念^[1~3],核心种质要求以最少的遗传重复和种质数量代表整个遗传资源最大的多样性,并应满足目前与将来生产与育种的需要。因此,在构建核心种质中这两个方面都应考虑。检验核心种质的有效性一方面要检验核心种质对整个种质资源遗传多样性的代表性,即对其遗传多样性进行评价;另一方面要对其在生产上实用性进行评估。

4.1 对核心种质遗传多样性的符合性检验

4.1.1 离散指标多样性 $Dv = \sum p_i \exp(-p_i)$, 此处 p_i 表示材料含有第 i 个指标状态的比例。这种指数具有可以量度多样性的合理特征: 当存在越多指标时其值越大, 指标状态近似时较指标状态差异较大时的值大, 然而此值受稀有基因的有无影响较小。在某些情况下这是一个好的性质, 但是在核心种质中稀有指标状态相当重要。在强烈受到偶然性影响的情况下, 它不是多样性理想的量度值。一个更为广泛应用的多样性量度方法是 Shannon 多样性信息指数: $H' = -\sum p_i \log p_i$, 其特征与 Dv 相似。第三种方法是 $1/\sum p_i$, 当在一个无限大的群体中随机选择时, Shannon 指数特别适合。在研究一个完全的有限的群体时, 其多样性用 Brillouin's 指数表示: $H = c/N(\log N! - (\log N!))$, 这里 $N!$ 为材料具有第 i 个指标状态的数量, N 指材料总数, c 为以 10 为底的对数与其他底间换算的常数。

4.1.2 连续性指标多样性 对连续性指标多样性常用方差、标准离差、变幅、平均数及高四分值与低四分值间的距离等表示。就目前的情况来说, 多数作物的核心种质还处于刚刚建立的阶段, 对核心种质进行检验, 只能凭借现有的数据资料。目前所建立起的核心种质基本上代表了整个种质资源的多样性^[5, 6, 19, 22, 23, 32~35]。沈金雄等通过聚类分析对中国芝麻核心种质的研究表明, 核心种质中参聚性状符合度为 90% 以上, 非参聚性状的符合度也均达到了 70% 以上。Basigalup(1995) 在建立多年生苜蓿引进种的核心种质时, 通过包括简单分组随机、系统随机以及简单分组人工、系统分组人工等 8 种方法的比较, 发现所有方法都没有改变核心种质的平均值, 系统方法和人工方法增加了核心种质中的方差。李自超等研究表明, 表型保留比例、表型方差、表型频率方差、遗传多样性指数和变异系数等 5 个参数作为核心种质检验指标比较有效^[11]。

4.2 对核心种质实用性的检验

对核心种质的实用性的检验有二方面的内容, 一方面是检验所构建的核心种质库中是否保留已知农艺性状及其他性状(基因); 另一方面是未来的实际应用中能否找到目的性状或基因, 要满足该方面的要求, 所构建的核心种质必须包括总资源中所有类型的基因。通常植物育种学家参与核心种质实用性的评价和检验是必要的。

在对核心种质构建后的利用上报道较少, 现主要是在几个构建较早的核心种质上进行了检验^[5, 29]。Charmet 等通过对法国多年生黑麦草群体的核心种质的基因型与环境互作的研究, 认为当将来来自于干旱和温暖地带的材料在具有同样特征地点鉴定时, 具有正向互作, 因此可以为生产与育种研究提供参考。Corley Holbrook 等在通过花生核心种质对后期叶斑病抗性进行鉴定时, 对仅相当于整个种质资源 27% 的材料进行鉴定, 鉴定出了相当于对整个种质资源进行鉴定时的 54% 的抗性材料, 即其效率提高了一倍。

5 问题与展望

5.1 遗传资源数据的缺乏

Plucknelt 等(1987)认为^[36], 在全世界的种质中有 65% 的缺乏基本数据, 80%~95% 缺乏特征及鉴定数据。因此有人认为种质库中的材料缺乏基本信息是构建核心种质的主要限制。当然有关种质的完整记录与评价是构建完整核心种质的理想基础, 然而, 没有必要因缺乏完整的数据而拖延构建核心种质^[37]。Brown(1995)同时指出, 构建核心种质应在对植物深入评价之

前,而不是之后,即构建核心种质的目的是为了对种质资源进行更深入的评价。

笔者在进行中国稻种资源核心种质的研究中发现,已编目入库的数万份资源中所记录的基本数据、特征数据和评价鉴定数据很少,而且已记录性状还很不完全,无法有效的进行系统的取样方法的研究。这可能是资源数量大,缺乏足够的人力和物力进行详细的考查和鉴定所致,这也是目前遗传资源研究中普遍存在的问题。核心种质的构建只能在这些并不完善的数据的基础上提取尽可能有代表性的样本,这可能是影响核心种质质量的限制因素之一。不过对初选样品再经过种植和详细的考查和鉴定可弥补一些数据,然后进行聚类压缩,也可构建成比较理想的核心种质。

5.2 核心种质的动态变化

核心种质有效性的一个重要指标是指它对所代表的基因库中多样性的代表程度,然而,基因库中的材料往往不能代表一个种或作物的整个多样性^[38]。为最恰当地代表一个种所具有的遗传多样性,核心种质的大小及内容应随时间而演变。当然,核心种质的大小及内容的变化应该是一个相当慢的过程^[37]。然而,在短时间内某些变化是必要的。

促使核心种质结构和大小改变的原因有:①相对于保留种质来说,使核心种质变异最大化(等位基因的频率及质量);②提供考虑到生物及非生物胁迫,满足育种者要求和新的耕作制度的额外遗传材料;③鉴定有用变异在原种质中有限的性状,并尽可能将稀有而有价值的等位基因包括进核心种质;④挖掘在未鉴定的野生及栽培材料中的潜在有效性;⑤除去因变异、外来花粉或种子等污染或混杂造成的变化,或通过保证足够大的样本和减小自然选择以将遗传漂移最小化;⑥阐明种内的遗传多样性结构,优化核心种质的结构等;⑦收到来自新地区或代表新类型的材料。

笔者提出,为便于种质资源的保存与研究利用,种质资源应具有动态的层次结构,核心种质是这种动态结构的一个活跃的层次。构建种质资源的四级结构:保留种质(reserve collection)、初级核心种质(initial collection)、核心种质(core collection)、核心应用种质(working collection)。初级核心种质可望能够代表整个种质资源95%以上的多样性^[11],从而保证为今后研究与利用提供丰富的遗传变异类型;核心种质将以占原种质资源5%~10%的种质量代表其70~80%以上的多样性,并含有主要的生态与变异类型,较容易在研究与生产中利用;核心应用种质则保存有更少的种质量,但含有目前急需和广泛利用的优异和特殊材料,因而可以很容易的为生产和研究所利用。这种四级结构相互之间可以在内容上动态变换,在信息上互相补充,既考虑到了其多样性又满足了种质资源目前和长远利用的需要。

5.3 核心种质概念的完善

随着核心种质研究的进展,Brown提出的核心种质似乎已不够完善,笔者提出核心种质的概念应该包含两方面的内容,一是Frankel等提出的以最少的资源份数最大限度的代表其遗传多样性,二是包含生产实践中所需要的优异农艺性状或基因。

核心种质并不是全部收集品的简单代表,而是以最少的份数最大限度地代表其遗传多样性。到目前为止,核心种质的研究多局限于形态或农艺性状上,少数研究利用了同工酶、储藏蛋白的信息等,由于实验条件和经费的原因,还没有从引起生物遗传变异的本质DNA上进行植物核心种质的构建。笔者认为,利用形态、生化及分子等水平的遗传信息所构建的核心种质才能真正的最大限度的代表该物种的遗传多样性。据报道,近年来诞生的SSR和AFLP等分子标记技术可有效地评价生物的遗传多样性,因此,在今后的植物核心种质构建中利用分子标记

技术将可能是最佳的研究手段,从而提高所构建的核心种质的有效性。

此外,核心种质的在遗传和育种实践中的可利用性已愈来愈受到资源研究者和遗传育种学家的重视。资源利用者期待着从核心种质中得到所需要的资源或基因,遗传资源研究者可以通过全面、详细的考查、分析和鉴定核心种质,以此全面认识所有遗传资源。核心种质的可利用性主要体现在包含当今已知的优异基因和未来所需要的优异基因,当今已知的优异基因可通过鉴定来发现,未来所需要的优异基因主要通过核心种质概念的第一个内容来实现。

参 考 文 献

- 1 FAO. Report on the state of the world's plant genetic resources for food and agriculture, prepared for the International Technical Conference on Plant Genetic Resources. Leipzig, Germany, June 1996, 17~23
- 2 Frankel O H. Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock W J, et al, eds. Genetic Manipulation: Impact on Man and Society. Cambridge, UK: Cambridge University Press, 1984
- 3 Frankel O H, Brown A H D. Current plant genetic resources—a critical appraisal. In: Genetics: New Frontiers (vol. N). New Delhi, India: Oxford and IBH Publishing, 1984
- 4 Brown A H D. The case for core collections. In: Brown A H D, Frankel O H, Marshall R D, et al. eds. The Use of Plant Genetic Resources. Cambridge, England: Cambridge Univ Press, 1989a, 136~156
- 5 IBPGR. Annual Report. Rome, Italy: IBPGR, 1990
- 6 Corley Holbrook C, William F Anderson. Evaluation of a core collection to identify resistance to late leafspot in peanut. *Crop Sci*, 1995, 35: 1700~1702
- 7 沈金雄,郭庆元等. 中国芝麻种质资源的聚类分析. *华中农业大学学报*. 1995,14(6):532~536
- 8 Holden J H W. The second ten years. In: Holden J H W, eds. Crop Genetic Resources: Conservation and Evaluation. George Allen and London: Unwin, 1984
- 9 周明德. 核心收集品的研究及其发展. *作物品种资源*,1994(增刊):3~6
- 10 Brown A H D. The core collection at the crossroad. In: Hodgkin T, Brown A H D, Hintum van T H L, Morales E A V, eds. Core Collections of Plant Genetic Resources. International plant genetic resources institute (IPGRI): A Wiley-Sayce Publication, 1995, 3~19
- 11 李自超,张洪亮,孙传清等. 云南地方稻种资源核心种质取样方法研究. 水稻遗传育种国际学术讨论会,杭州,1999
- 12 Hodgkin T. Core collection and conservation of genetic resources. In: Arora R K, Riley K W, eds. Sesame Bio-diversity in Asia, Conservation, Evaluation and Improvement. New Delhi, India, 1993
- 13 Yonezawa K, Nomura T, Morishima H. Sampling strategies for use in stratified germplasm collections. In: Hodgkin T, Brown A H D, Hintum van T H L, Morales E A V, eds. Core Collections of Plant Genetic Resources. International Plant Genetic Resources Institute (IPGRI): A Wiley—Sayce Publication, 1995, 35~54
- 14 Hintum van T J L. Hierarchical approaches to the analysis of genetic diversity in crop plants. In: Hodgkin T, Brown A H D, Hintum van T H L, Morales E A V, eds. Core Collections of Plant Genetic Resources. International Plant Genetic Resources Institute (IPGRI): A Wiley-Sayce Publication, 1995, 23~34
- 15 Spagnoletti Zeul P L, Qualset C O. Geographical diversity for quantitative spike characters in a world collection of durum wheat. *Crop Sci*, 1987, 27: 235~241
- 16 Brown A H D, Grace J P, Speer S S. Designation of a “core” collection of perennial Glycine. *Soybean-Genetics-Newsletter*, 1987,14: 59~70

- 17 Cordeiro C M T, Morales E, Ferreira F, Rocha D M S, et al. Towards a Brazilian core collection for cassava. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August 1992
- 18 Hamon S, Van Sloten D H. Characterization and evaluation of Okra. In: Brown A H D, Frankel O H, Marshall R D, et al. eds. The Use of Plant Genetic Resources. Cambridge, England: Cambridge Univ Press, 1989
- 19 Basigalup D H, Barnes D K, Stucker R E. Development of a core collection for perennial *Medicago* plant introductions. *Crop Sci*, 1995, 35: 1163~1168
- 20 Diwan Noa, Cary R Bauchan, Marla S McIntosh. A core collection for the US annual *Medicago* germplasm collection. *Crop Sci*, 1994, 34: 279~285
- 21 Casler M D. Patterns of variation in a collection of perennial ryegrass accessions. *Crop Sci*, 1995, 35: 1169~1177
- 22 Diwan N, McIntosh M S, Bauchan G R. Methods of developing a core collection of annual *Medicago* species. *Theor Appl Genet*, 1995, 90: 755~761
- 23 Joe Tohme, Orlando Gonzalez D, et al. AFLP analysis of gene pools of a wild bean core collection. *Crop Sci*, 1996, 36: 1375~1384
- 24 Perry M C, Mcintosh M S, Stoner A K. Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. *Crop Sci*, 1991a, 31: 1350~1355
- 26 Perry M C, Mcintosh M S, Stoner A K. Geographical patterns of variation in the USDA soybean germplasm collection: II. Allozyme frequencies. *Crop Sci*, 1991b, 31: 1356~1360
- 25 Singh P S, Gutierrez J A, Molina A, Urrea C, Gepts P. Genetic diversity in cultivated common bean: I. Marker-based analysis of morphological and agronomic traits. *Crop Sci*, 1991, 31: 23~29
- 27 贾继增. 分子标记种质资源鉴定和分子标记育种. *中国农业科学*, 1996, 29(4): 1~10
- 28 Crossa J, Delacy I H, Taba S. The use of multivariate methods in developing a core collection. In: Hodgkin T, Brown A H D, Hintum van T H L, Morales E A V, eds. Core Collections of Plant Genetic Resources. International Plant Genetic Resources Institute (IPGRI): A Wiley-Sayce Publication, 1995, 77~93
- 29 Charmet G, Balfourier F, Ravel C, Benis J B. Genotype \times environment interactions in a core collection of French perennial ryegrass populations. *Theor Appl Genet*, 1993, 86: 731~736
- 30 Erskine W, Muehlbauer F J. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theor Appl Genet*, 1991, 83: 119~125
- 31 Spagnoletti Zeul P L, Qualset C O. Evaluation of five strategies for obtaining a core subset from a large genetic resources collection of durum wheat. *Theor Appl Genet*, 1993, 87: 295~304
- 32 范传珠, 刘旭, 马缘生等. 小麦特殊遗传材料核心样品的建立. *作物品种资源*, 1994(增刊): 7~10
- 33 Gilles Charmet, Francois Balfourier. The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Genetic Resources and Crop Evolution*, 1995, 42: 203~309
- 34 Skroch Paul W, James Nienhuis, Steve Beebe, et al. Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. *Crop Sci*, 1998, 38: 488~496
- 35 Hintum van T J L. Comparison of marker systems and construction of a core collection in a pedigree of European spring barley. *Theor Appl Genet*, 1994, 89: 991~997
- 36 Plucknett D L, et al. Gene Banks and the World's Food. Princeton. New Jersey, USA: Princeton University Press, 1987
- 37 Brown A H D. Core collection: a practical approach to genetic resources management. *Genome*, 1989b, 31: 818~824
- 38 ICARDA. Report of the First Meeting on the Barley Core Collection Committee, May 1992. Mimeograph. Aleppo, Syria: ICARDA, 1992