植物遗传资源学报 2020，21（3）：605-615
Journal of Plant Genetic Resources
DOI：10.13430/j.cnki.jpgr.20190618003

基于 SNP 标记的玉米自交系类群划分方法和分群功效评估指标的比较

李念念[1]，王义波[2]，徐国平[2]，易　黎[2]，王爱方[2]，李　婷[2]，曹刚强[1]
（[1]郑州大学农学院，郑州 450001；[2]北京联创种业有限公司，北京 100081）

摘要：随着高通量测序技术的不断进步带来的 SNP 标记成本的持续下降，可用于种质资源分群和分子育种应用中基因型鉴定的 SNP 位点数越来越多，亟需系统地比较不同分群方法以便找到最合适的分群法和最可靠的分群功效评估指标。本研究比较了 4 个分群方法（包括目前常用的邻接算法（NJ 法）、SNPhylo 法、ADMIXTURE + SNPs 和在 ADMIXTURE +SNPs 基础上开发的 ADMIXTURE+TagSNPs 分群法），以及 4 个分群功效评估指标（PCA 散点图、群体遗传指标 GD 和 PIC 及贝叶斯统计指标 BIC）在利用 GBS 简化基因组测序产生的 525141 个 SNPs 位点数据将 490 份玉米自交系划分成 3 个和 6 个亚群时的表现。结果表明：4 个评估指标中的 PCA 散点图和 BIC 指标（BIC_{BW}，S_{BIC}）探测亚群间变异的能力强，是评估不同分群方法分群功效的可靠指标，而 GD 和 PIC 探测亚群间变异的能力差，不适合用作分群功效的评估。结果还表明，4 个分群法均为有效分群法，所以都可用于种质资源分群，但 ADMIXTURE+TagSNPs 分群法划分的亚群边界清晰，亚群间个体混杂少，相对群间变异度大，综合表现最好，而 SNPhylo 法的综合表现最差。考虑到 ADMIXTURE+TagSNPs 需要输入的 SNP 标记数显著少于其他 3 种分群法，因而实际应用中基因型鉴定的成本最低，所以建议在遗传资源研究和分子育种应用中首选该分群法。

关键词：玉米；聚类分析；标签 SNP；主成分分析；遗传多样性；多态信息量；贝叶斯信息度

Comparison of Different Grouping Procedures and Evaluation Criteria for Grouping Maize Inbreds Using SNP Data

LI Nian-nian[1], WANG Yi-bo[2], SHU Guo-ping[2], YI Li[2], WANG Ai-fang[2], LI Ting[2], CAO Gang-qiang[1]
（[1]*School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001*; [2]*Beijing Lantron Seed Co., Ltd, Beijing 100081*）

Abstract: Grouping germplasm lines and assisting plant breeding using large number of SNP marker have become well accepted due to constant price drops of SNP markers brought about by the advance at high-throughput sequencing technology. How to handle the large SNP datasets becomes an increasing interest, and the user-friendly statistical methodologies are in demand. In this study, four grouping procedures (NJ, SNPhylo, ADMIXTURE + SNPs, and the ADMIXTURE + TagSNPs which we modified from ADMIXTURE + SNPs), were deployed to group 490 corn inbreds into 3 and 6 subgroups using 525141 SNP markers and their performance were evaluated with four criteria (PCA Scatter Plot, GD, PIC and BIC). The result showed that PCA Scatter Plot and BIC (BIC_{BW}, S_{BIC}) among the four criteria are more powerful in revealing between-subgroup variation, whereas GD and PIC showed less powerful. All four grouping procedures were effective and could be adopted in grouping germplasm. Particularly, ADMIXTURE+TagSNPs was the most effective in delineating subgroups with clear boundary and very little between-group mixing, while SNPhylo was the least effective. ADMIXTURE + TagSNPs required fewer SNP markers thus would cost less than other three procedures, and therefore was highly

收稿日期：2019-06-18　**修回日期：**2019-08-08　**网络出版日期：**2019-11-07
URL：http://doi.org/10.13430/j.cnki.jpgr.20190618003
第一作者研究方向为作物遗传育种与生物信息，E-mail：1511225891@qq.com；王义波，徐国平为共同第一作者
通信作者：曹刚强，研究方向为生物信息和植物遗传学，E-mail：caogq@zzu.edu.cn
基金项目：北京市科技计划重大项目（D171105007700003）；河南省高等学校重点科研项目（13A180687）
Foundation project: Major Projects of Beijing Science and Technology Plan (D171105007700003), Key Scientific Project for Universities of Henan Province (13A180687)

recommended for germplasm study and marker-assisted breeding.

Key words: maize; clustering analysis; tagSNP; principal component analysis; genetic diversity; polymorphism information content; Bayesian information criterion

农作物种质材料的聚类分群在作物育种研究和种质资源研究中有非常重要的作用。例如利用分群技术发现种质材料之间的亲缘关系，根据分群结果发现重要性状的基因来源和遗传材料来源，在玉米育种中参考玉米自交系的分群结果划分自交系的杂种优势群，选择合适的自交、回交亲本和测定配合力用的测验系等[1-4]。以前研究者使用形态特征、亲本的遗传系谱记录和大量实际田间测交结果来完成上述工作，工作量大且不准确，特别是对于系谱来源复杂但缺少准确系谱记录的种质资源。利用 DNA 分子标记的遗传多态信息研究种质资源和育种群体的遗传多样性，划分种质类群，开发利用新种质一直是近年来植物遗传资源研究的一个重要方向[5-8]。目前，DNA 分子标记经历了从 RFLP、SSR 到 SNP 的 3 个发展阶段，分子标记的数目也由限制性片段长度多态性标记（RFLPs，restriction fragment length polymorphism）的数十个，到简单重复序列标记（SSRs，simple sequence repeat）的数百个，再到目前的单核苷酸多态性标记（SNP，single nucleotide polymorphism）的数十万到数百万个[9-13]。随着高通量 DNA 测序技术进步带来的测序成本的降低，各种简化基因组测序技术如 GBS、GBTS、SLAF-seq、SNP 芯片技术，和全基因组重测序技术（Whole Genome Resequencing）[14-16]产生数量巨大的 SNP 数据。适用于 SNP 大数据的聚类和分群的数学模型、分群流程、计算机算法和分群软件，如 STRUCTURE[17]和 ADMIXTURE[18]不断被开发出来，一些过去常用的经典分群方法如邻接法（NJ，Neighbor-Joining）[19]仍在广泛使用。科研工作者对采用哪种分群方法和分群软件通常感到困惑，采用何种指标评估众多分群方法和分群软件的分群功效的系统研究尚不多见，对正确地选用合适的分群方法和解释分群结果造成困难。亟需系统地比较不同分群方法以便找到最合适的分群法和最可靠的分群功效评估指标。

本研究将邻接法、SNPhylo、ADMIXTURE+SNPs 和 ADMIXTURE+TagSNPs 4 种分群方法用于分析高通量 GBS 测序产生的 525141 个 SNP 位点的多态性数据，把 490 份玉米自交系分为 3 个和 6 个亚群，利用 4 个评估指标比较评估每种分群方法的分群功效。同时，利用实际分群结果比较了 4 个评估指标的适用性，以期为研究者在选择合适的植物遗传资源分群方法和分群功效的评估指标方面提供参考依据。

1 材料与方法

1.1 试验材料和 SNP 数据

490 个玉米自交系均来自联创种业种质资源库，包括近年黄淮海地区育种用的骨干种质，重要的中国地方种质和 1970-2010 年多次从北美、欧洲、南美和东南亚引进的国外种质。利用 CTAB 方法提取 490 份玉米自交系幼苗的 DNA，SNP 数据来自简化基因组测序[14]：用限制性内切酶 *Ape*KI 处理 DNA，构建 DNA 文库，利用 Illumina HiSeq2000 进行测序，测序深度为 5X，以玉米自交系 B73-V3.0 作为参考基因组，利用 Tassel-GBS 流程得到 876305 个 SNP 位点多态性变异[14-16]。本研究用 Tassel V5.2 软件将最小等位基因频率（MAF，minimum allele frequency）小于 5% 的 SNP 过滤掉后得到 525141 个 SNP 位点数据，这些 SNP 数据被用作输入分群软件的原始数据。

1.2 分群方法

本试验利用 NJ、SNPhylo、ADMIXTURE+SNPs 和 ADMIXTURE+TagSNPs 4 种分群方法对 490 份玉米自交系进行亚群划分。

1.2.1 NJ 分群法 NJ 分群法以构造聚类树状图时采用的计算机算法“邻接算法（Neighbor Joining Algorithm）”命名；本研究采用 TASSEL V5.2.3 软件中的 NJ 分群法，首先将 490 个自交系的 525141 个 SNP 位点数据导入到 TASSEL V5.2 软件中，软件根据 SNP 数据计算任意 2 个自交系之间的修正欧几里德距离（Modified Euclidean Distance），然后从产生的距离矩阵用邻接算法产生自交系的树状图[19]。

1.2.2 SNPhylo 分群法 采用 Lee 等[20]2014 年开发的适用于处理大型 SNP 数据的系统树构建和分群软件 SNPhylo，该软件整合了系统学和进化生物学中多个常用软件的功能，包括用 PHYLIP 软件包[21]中的最大似然法子程序 DNAML 进行树构建和分群，用 PHANGORN 中的自助重抽样法（bootstrapping resampling）估算进化树上遗传

材料之间的进化距离和最优进化关系树的置信度[22-23]，该软件根据 SNP 位点的多态信息和连锁不平衡度删除信息量低的 SNP 位点，实际使用 8249 个 SNP 位点数据，极大地减少了软件运行时间。

1.2.3　ADMIXTURE+SNPs 分群法　本研究利用 Kenneth 等[24]开发的软件 ADMIXTURE V1.3.0 对 490 份玉米自交系进行分群和群体结构分析。与分群软件 STRUCTURE[12] 相比，ADMIXTURE 也是基于模型分群（Model-based Clustering），具有相似的推断模型，但使用速度更快的数值优化算法“拟牛顿加速度算法”，更适合处理 SNP 大数据[17-18, 24-25]。ADMIXTURE 软件根据群体遗传学的群体混合原理（admixture theory）汰选 SNP，实际使用了 511 个 SNP 位点数据。

1.2.4　ADMIXTURE+TagSNPs 分群法　该方法是本试验室开发的分群流程。它利用 Haploview V4.2 软件先对 525141 个 SNP 位点进行单倍型块（haplotype block）分析和单倍型提取，在每个单倍型块中寻找标签 SNPs[26-28]，即可以代表该单倍型块的行为的 SNPs，将标签 SNP 作为 ADMIXTURE 软件的输入数据。用 Haploview V4.2 软件对 10 条染色体上的 SNPs 进行单倍型块分析的结果见表 1。从 3618 个单倍型块中挑选出 4849 个 TagSNP 位点，将其多态性信息输入 ADMIXTURE V1.3.0 软件中，该软件根据群体遗传学中的群体混合原理汰选 SNP，最终实际使用 929 个 TagSNP 位点信息。

表 1　10 条玉米染色体的单倍型块数目以及标签 SNPs 数目
Table 1　Number of haplotype blocks and number of TagSNPs on 10 maize chromosomes

染色体 Chromosome	单倍型块数目 No. of haplotype blocks	标签 SNP 数目 No. of TagSNPs	染色体 Chromosome	单倍型块数目 No. of haplotype blocks	标签 SNP 数目 No. of TagSNPs
1	533	708	7	291	397
2	448	594	8	304	405
3	411	554	9	260	356
4	478	644	10	293	383
5	321	427	总计 Total	3618	4849
6	279	381			

1.3　分群功效评估指标

1.3.1　PCA 散点图法　主成分分析（PCA，principal component analysis）是考察多个个体或变量间相关关系，并将它们的相关关系在降维（通常是二维）几何空间表达出来的一种多元统计方法。用线性代数学的术语，就是通过正交变换将原始的 n 维数据集变换到一个新的被称作主成分的降维数据集中。变换后的结果中，第一个主成分具有最大的方差值，每个后续的主成分在满足与前述主成分正交的限制条件下具有最大方差。降维时仅保存前 m（m<n）个主成分即可保持最大的数据信息量。PCA 散点图利用第一和第二主成分值作为 X 和 Y 坐标值，把待分类的个体之间的距离和相关关系表达在二维平面图上（m=2）。图上相距较近的个体可以被归为一个亚群，从图中亚群内个体的相对集中度和亚群间边缘的清晰度可以直观地判断一个分群方法的分群功效的高低。PCA 分析利用 R 软件包 SNPRelate V2.14.0 中的 PCA 分析和可视化工具[29-31]。

1.3.2　亚群间／亚群内相对变异度指标　评估分群或聚类方法分群功效高低的各种不同方法的基本原理为：用该分群或聚类方法把一群未分类个体分为 K 个亚群，比较亚群间变异度与亚群内变异度的相对大小，即亚群间相对变异度（E）= 亚群间变异度 / 亚群内变异度，亚群间相对变异度越大，说明该分群方法的分群功效越高。本研究采用 3 个度量亚群间和亚群内变异度的指标：2 个是群体遗传学中通常用来度量遗传多态性的指标：遗传多样性（GD，genetic diversity）和多态信息量（PIC，polymorphism information content），1 个是来自贝叶斯统计的信息度指标：贝叶斯信息量（BIC，Bayesian information criterion）[32-35]。以上 3 个指标值可以通过 Excel、Tassel、PowerMarker、

SAS 等软件计算获得。以下是 3 个指标的计算公式：

$$GD_u = \left(1 - \sum_{i=1}^{k} P_i^2\right) \Big/ \left(1 - \frac{1+f}{n}\right) \quad (1)$$

$$PIC_u = 1 - \sum_{i=1}^{k} P_i^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2P_i^2 P_j^2 \quad (2)$$

$$BIC_u = 2\ln(L) - k\ln(n) \quad (3)$$

GD 值可以由公式（1）计算得到。u 表示标记位点，i、j 表示等位基因，k 表示标记位点 u 的等位基因数目，f 表示近交系数，由于所用材料均为纯合自交系，故 f=1，n 表示被分类的个体数，P_i 是标记位点 u 的第 i 个等位基因的频率。PIC 值可以由公式（2）计算得到，P_i 和 P_j 是标记位点 u 的第 i 个和第 j 个等位基因的频率；BIC 值可以由公式（3）计算得到，L 是似然函数，n 是总样本数，k 是亚群数目。

GD、PIC、BIC 亚群间相对变异度（亚群间变异度 / 亚群内变异度）计算公式见（6）、（9）和（12），其计算过程见公式（4）~（12），其中 W_i 是第 i 个亚群的权重，GD_i、PIC_i、BIC_i 是第 i 个亚群的 GD、PIC 和 BIC 值，首先得到亚群内和亚群间 GD、PIC、BIC，进而计算出 GD、PIC、BIC 亚群间相对变异度。

$$GD_W = W_1GD_1 + W_2GD_2 + \ldots\ldots + W_iGD_i \quad (4)$$

$$GD_B = GD_T - GD_W \quad (5)$$

$$GD_{BW} = GD_B / GD_W \quad (6)$$

$$PIC_W = W_1PIC_1 + W_2PIC_2 + \ldots\ldots + W_iPIC_i \quad (7)$$

$$PIC_B = PIC_T - PIC_W \quad (8)$$

$$PIC_{BW} = PIC_B / PIC_W \quad (9)$$

$$BIC_W = W_1BIC_1 + W_2BIC_2 + \ldots\ldots + W_iBIC_i \quad (10)$$

$$BIC_B = BIC_T - BIC_W \quad (11)$$

$$BIC_{BW} = BIC_B / BIC_W \quad (12)$$

公式（4）~（12）中，GD、PIC、BIC 的下标 T、W、B、BW 分别表示总体、亚群内、亚群间和亚群间相对变异度。

同时，本研究采用灵敏度（S，sensitivity）来比较每种分群方法在分群数目增加时，亚群间变异的增加和亚群内变异减少的相对速度。在本研究中，其定义为当亚群数增加时（在这里即由 K=3 增加到 K=6），亚群间相对变异度值的增加幅度，其计算方法为 K=6 时 GD_{BW}、PIC_{BW} 和 BIC_{BW} 与 K=3 时 GD_{BW}、PIC_{BW} 和 BIC_{BW} 的比值。例如，用 BIC 度量 NJ 法的灵敏度可以表示为 $S_{BIC} = BIC_{BW.6} / BIC_{BW.3}$，同理，$S_{GD}$ 和 S_{PIC} 简化表达为 S=K6/K3。

1.4 不同分群方法功效比较

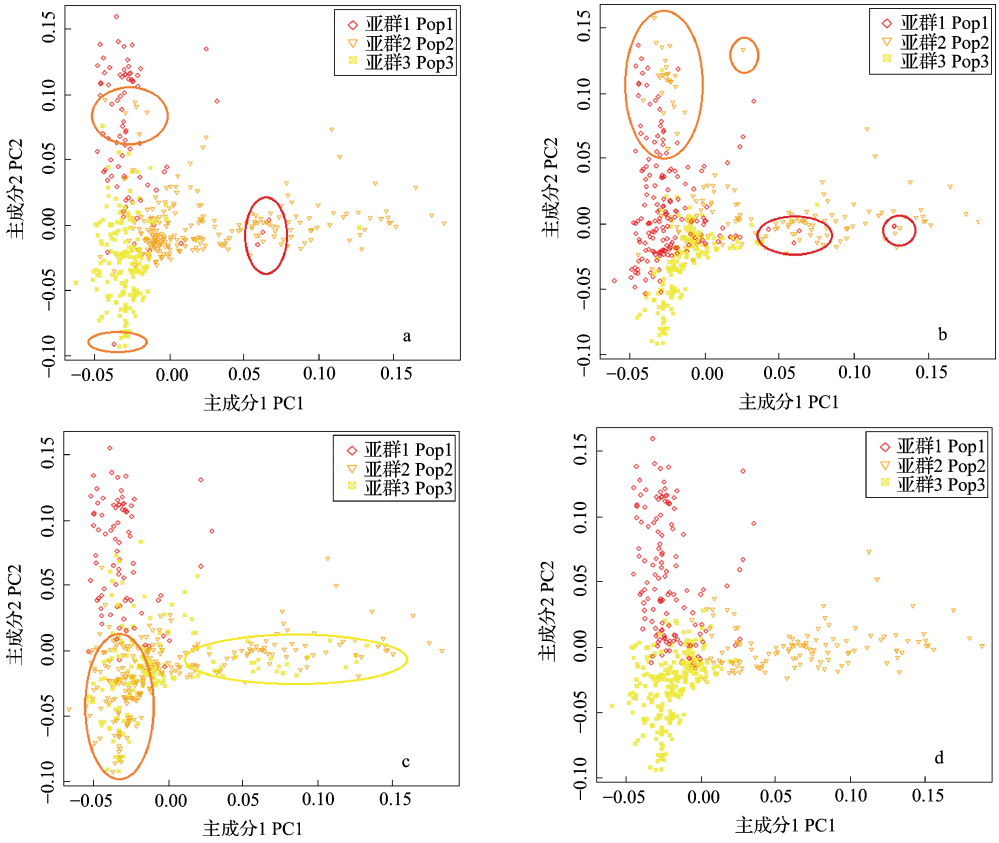
NJ、SNPhylo、ADMIXTURE+SNPs 和 ADMIXTURE+TagSNPs 分群法划分亚群时实际利用的 SNP 位点数目分别是 525141 个、8249 个、511 个和 929 个。由于所用的 SNP 位点数不同，无法直接比较 4 种分群方法的分群功效，但可以通过把 3 种基于最大似然法的分群方法的每一种和经典分群方法 NJ 法做两两比较，间接发现它们在分群功效方面的表现。在比较 SNPhylo 法与 NJ 法时，使用了二者共有的 8249 个 SNPs 位点，由此计算出将 490 份自交系分为 3 个亚群（K=3）和 6 个亚群（K=6）时各个亚群的玉米自交系数目、GD 值和 PIC 值（详见 http://doi.org/10.13430/j.cnki.jpgr.20190618003，附表 1、附表 2）以及 BIC 值（详见 http://doi.org/10.13430/j.cnki.jpgr.20190618003，附表 3）；在比较 ADMIXTURE+SNPs 法和 NJ 法时使用了 511 个共有 SNPs，在比较 ADMIXTURE+TagSNPs 法与 NJ 法时使用 929 个共有 SNPs（附表 1、附表 2、附表 3）。

2 结果与分析

2.1 不同分群方法分群功效的主成分分析评估

深入的分群研究结果表明将 490 个自交系分为 3 个和 6 个亚群最接近真实情况[36]，所以在评估分群方法时仅研究将 490 份玉米自交系分成 K=3 和 K=6 个亚群时这些分群方法的表现。通过 PCA 图上个体的相对集中度和亚群间边缘的清晰度（两个亚群的空间重叠程度）对 4 种分群方法进行评估。利用 NJ 法划分为 3 个亚群的 PCA 展示，亚群 1 有 3 个玉米自交系延伸到亚群 2 的边界内，有 2 个玉米自交系延伸到亚群 3 的边界内，亚群 2 有 6 个玉米自交系延伸到亚群 1 的边界内，有 3 个玉米自交系延伸到亚群 3 的边界内；亚群 3 有 3 个玉米自交系延伸到亚群 1 的边界内，有 1 个玉米自交系延伸到亚群 2 的边界内（图 1a）；由 SNPhylo 法可知，亚群 2 被亚群 1 隔开，被分成了两部分，且相距较远（图 1b）；由 ADMIXTURE+SNPs 法可知，亚群 2 和亚群 3 有较多个体几乎混合在一起，无法清晰分开（图 1c）；由 ADMIXTURE+TagSNPs 法可知，可以看到各亚群之间边界清晰没有明显的重叠现象（图 1d）。从 PCA 分析结果可以看出在分为 3 个亚群时，ADMIXTURE+TagSNPs 法和 NJ 法的分群功效相对较高。

NJ 法、SNPhylo 法和 ADMIXTURE+SNPs 法划分的 6 个亚群间有少量混合，各个亚群之间边界有重叠现象（图 2a、2b、2c）。而 ADMIXTURE+TagSNPs 法



a：NJ 法；b：SNPhylo 法；c：ADMIXTURE+SNPs 法；d：ADMIXTURE+TagSNPs 法。下同

a：NJ，b：SNPhylo，c：ADMIXTURE+SNPs，d：ADMIXTURE+TagSNPs.The same as below

图 1　4 种分群方法划分群体结构的主成分分析图（3 个亚群）

Fig.1　Population structure PCA plots by four grouping procedures（3 populations）

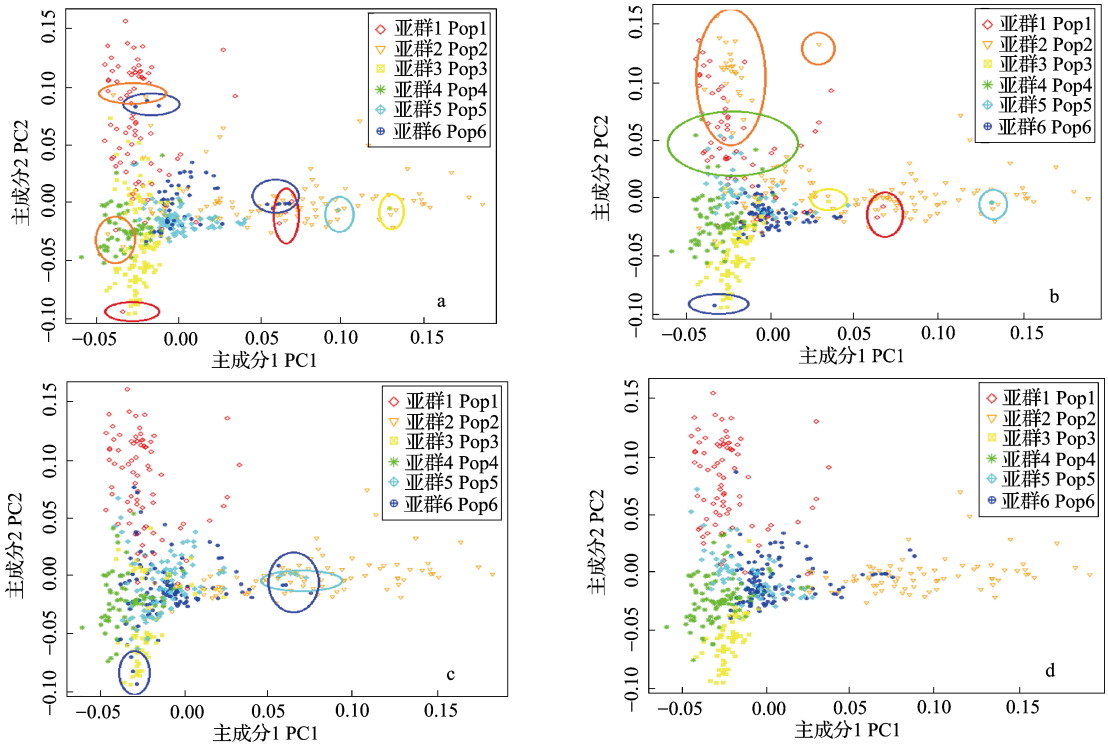


图 2　4 种分群方法划分群体结构的主成分分析图（6 个亚群）

Fig. 2　Population structure PCA plots by four grouping procedures（6 populations）

划分的 6 个亚群边界较清晰，只有第 6 亚群有个别个体混入第 1 和第 3 亚群（图 2d）。从主成分分析结果可以看出在分为 6 个亚群时，ADMIXTURE+TagSNPs 法的分群功效相对较高。

2.2 评估不同分群方法分群功效的群体遗传指标

将 490 份玉米自交系划分为 3 个亚群时，根据 NJ 法和 SNPhylo 法（8249 个 SNP）、NJ 法和 ADMIXTURE+SNPs 法（511 个 SNP）以及 NJ 法和 ADMIXTURE+TagSNPs 法（929 个 SNP）的分群结果计算 GD 值和 PIC 值，获得 GD 值总分分别为 0.215、0.185 和 0.344，PIC 值总分分别为 0.185、0.164 和 0.277（附表 1）。根据公式（4）、公式（7），当 NJ 法和 SNPhylo 法比较时，NJ 法的亚群内 GD（GD_W）为（80 × 0.176+228 × 0.221+182 × 0.202）/490=0.207，亚群内 PIC（PIC_W）为 0.176，根据公式（5）、公式（8），计算出亚群间合并 GD（GD_B）和亚群间合并 PIC（PIC_B）分别是 0.009 和 0.009；其他计算结果见附表 1。利用公式（6）和公式（9）得到 GD 相对变异度（GD_{BW}）和 PIC 相对变异度（PIC_{BW}）详见附表 1。

同理计算将 490 份玉米自交系划分为 6 个亚群时的 GD 和 PIC 值，获得 GD 总分分别为 0.215、0.185 和 0.344，PIC 总分分别为 0.185、0.164 和 0.277。与计算附表 1 亚群内和亚群间合并变异的方式相同，根据公式（4）、公式（5）、公式（7）、公式（8），计算出 NJ 和 SNPhylo、ADMIXTURE+SNPs、ADMIXTURE+TagSNPs 分群法进行两两比较时的亚群内和亚群间合并 GD 和 PIC 值详见附表 2，利用公式（6）和公式（9）得出 GD 相对变异度和 PIC 相对变异度详见附表 2。从附表 1 和附表 2 可以看出亚群间合并 GD 和 PIC 值均很小，亚群间 / 亚群内的比值（GD_B/GD_W、PIC_W/PIC_B）均接近于 0，表明用 GD 和 PIC 这 2 个指标探测群间变异效果不好。

2.3 评估不同分群方法分群功效的统计指标

将 490 个玉米自交系分别划分 3 个亚群和 6 个亚群时，根据 NJ 法和 SNPhylo 法（8249 个 SNP）、NJ 法和 ADMIXTURE+SNPs 法（511 个 SNP）以及 NJ 法和 ADMIXTURE+TagSNPs 法（929 个 SNP）的分群结果，根据公式（10）、公式（11）计算亚群内和亚群间合并变异的 BIC 值（表 2，附表 3）。利用公式（12）计算 BIC 相对变异度（表 2，附表 3）。从表 2 可以看出亚群间 / 亚群内的比值（BIC_B/BIC_W）显著大于 1，表明用 BIC 指标探测群间变异是有效的。

2.4 不同评估指标探测亚群间变异的能力比较

从表 2、附表 1、附表 2、附表 3 中给出的 GD、PIC 和 BIC 3 个评估指标的亚群内和亚群间变异估计值，根据公式（6）、公式（9）和公式（12）计算出每个分群方法将 490 个自交系分成 3 个亚群（K=3）和 6 个亚群（K=6）时，3 个评估指标度量的亚群间相对变异度值 GD_{BW}、PIC_{BW}、BIC_{BW}（表 3）。可以看到使用 NJ 分群法（8249 个 SNP），K=6 时，由 PIC 估计的亚群内和亚群间相对变异度值为 0.168 和 0.017（附表 2），此时由 PIC 估计出的亚群间相对变异度 PIC_{BW} = 0.017/0.168 =10.2%（表 3）。同理，可以看到使用 NJ 分群法（8249 个 SNP），在 K=3 时，由 BIC 估计出的亚群内和亚群间变异度值为 –270.835 和 –325.171（附表 3），此时由 BIC 估计出的亚群间相对变异度 BIC_{BW} =（–325.171）/（–270.835）=120.1%（表 3）。

从表 3 给出的 3 个评估指标估算出的亚群间相对变异度值 GD_{BW}、PIC_{BW}、BIC_{BW}，可以看出当 K=3 和 K=6 时，从 NJ 分群结果中得到的平均 GD 亚群间相对变异度（GD_{BW}）分别为 3.4% 和 6.3%，即 GD 提取的亚群间变异分别相当于亚群内变异的 3.4% 和 6.3%；同理，PIC 提取的亚群间变异相当于亚群内变异的 3.4% 和 7.2%；而 BIC 提取的亚群间变异相当于亚群内变异的 142.5% 和 432.5%，远大于 GD 和 PIC。以上结果表明 3 个评估指标中 BIC 探测亚群间变异的能力强，是评估不同分群方法分群功效的有效指标，而 GD 和 PIC 探测亚群间变异的能力差，不适合用作分群功效的评估。

2.5 不同分群方法的灵敏度

当亚群数由 K=3 增加到 K=6 时，用 NJ 分群法（8249 个 SNP），用 BIC 估计的亚群间相对变异度 BIC_{BW} 由 120.1% 增加到 389.3%，增加为 3.241 倍；用 SNPhylo 分群法（8249 个 SNP），亚群间相对变异度 BIC_{BW} 由 160.7% 增加到 345.8%，增加为 2.152 倍（表 3），表明 NJ 分群法比 SNPhylo 分群法对由亚群数增加引起的亚群间相对变异度的增加的反应更灵敏。同理可以算出由 BIC 估算的 ADMIXTURE+TagSNPs 分群法（929 个 SNP）的灵敏度为 2.963，ADMIXTURE+SNPs 分群法（511 个 SNP）的灵敏度为 2.682。1 个分群方法的灵敏度因输入数据的不同而不同，如 NJ 法在输入 SNP 数目为 8249 个、511 个和 929 个时，灵敏度的估值为 3.241、2.757、3.181，由 BIC 估算出的 NJ 分群法的平均灵敏度为 3.060（表 3）。结果表明 NJ 法和 ADMIXTURE+TagSNPs 的灵敏度明显高于 ADMIXTURE+SNPs 和 SNPhylo。

表 2　NJ 法和其他 3 种分群法两两比较的 BIC 值（6 个亚群）

Table 2　The BIC values of pair-wise comparison between NJ and other three grouping methods（K=6）

群组 Group	参数 Parameter	比较Ⅰ Comparison Ⅰ（8249 SNPs）		比较Ⅱ Comparison Ⅱ（511 SNPs）		比较Ⅲ Comparison Ⅲ（929 SNPs）	
		邻接法 NJ	SNPhylo 法 SNPhylo	邻接法 NJ	ADMIXTURE+SNPs 法 ADMIXTURE+SNPs	邻接法 NJ	ADMIXTURE+TagSNPs 法 ADMIXTURE+TagSNPs
全部自交系 Entire panel	自交系个数	490	490	490	490	490	490
	BIC_T	−614.464	−614.464	−607.951	−607.951	−759.961	−759.961
亚群 1 Pop1	自交系个数	80	61	80	76	80	76
	BIC_1	−107.001	−87.851	−102.441	−97.820	−132.804	−125.010
亚群 2 Pop2	自交系个数	96	147	96	83	96	74
	BIC_2	136.566	−214.758	−115.768	−97.806	−159.513	−120.510
亚群 3 Pop3	自交系个数	117	89	117	61	117	67
	BIC_3	171.722	−128.770	−148.134	−78.715	−180.715	−97.329
亚群 4 Pop4	自交系个数	65	70	65	63	65	91
	BIC_4	−87.675	−94.294	−79.069	−75.183	−89.860	−128.204
亚群 5 Pop5	自交系个数	78	47	78	102	78	63
	BIC_5	−123.822	−71.151	−107.265	−107.340	−121.515	−99.316
亚群 6 Pop6	自交系个数	54	76	54	105	54	119
	BIC_6	−81.7727	−121.128	−69.460	−162.396	−85.769	−195.501
群内 Within-Group	BIC_W	−125.581	−137.835	−109.995	−497.956	−136.800	−134.956
群间 Between-Group	BIC_B	−488.883	−476.629	−162.396	−499.603	−623.161	−625.005

BIC_T、BIC_W、BIC_B、BIC_1~BIC_6 分别表示全体材料、亚群内、亚群间、亚群 1~ 亚群 6 的贝叶斯信息量

BIC_T, BIC_W, BIC_B and BIC_1-BIC_6 represent Bayesian information criterion of entire panel, within- and between-group and Pop_1-Pop_6

表 3　4 个分群方法的群间相对变异度（GD_{BW}、PIC_{BW}、BIC_{BW}）和灵敏度比较

Table 3　The effectiveness（GD_{BW}, PIC_{BW}, BIC_{BW}）and sensitivity of four grouping procedures

SNP 数目 No.of SNPs	方法 Methods	GD_{BW}（%）		GD_{BW} 灵敏度 Sensitivity of GD_{BW}（S=K6/K3）	PIC_{BW}（%）		PIC_{BW} 灵敏度 Sensitivity of PIC_{BW}（S=K6/K3）	BIC_{BW}（%）		BIC_{BW} 灵敏度 Sensitivity of BIC_{BW}（S=K6/K3）
		K=3	K=6		K=3	K=6		K=3	K=6	
8249	NJ	5.1	8.7	1.706	4.2	10.2	2.429	120.1	389.3	3.241
	SNPhylo	4.0	8.2	2.050	5.0	9.7	1.940	160.7	345.8	2.152
511	NJ	1.4	2.8	2.000	2.0	3.8	1.900	164.2	452.7	2.757
	ADMIXTURE+SNPs	3.3	3.9	1.182	4.2	5.1	1.214	171.9	461.1	2.682
929	NJ	3.8	7.5	1.974	3.9	7.6	1.949	143.2	455.5	3.181
	ADMIXTURE+TagSNPs	4.1	7.7	1.878	4.0	7.7	1.925	156.3	463.1	2.963
	NJ 平均	3.4	6.3	1.893	3.4	7.2	2.092	142.5	432.5	3.060

3 讨论

3.1 不同评估指标的特点和适用性

本研究采用3类不同指标：空间可视化指标PCA，群体变异度指标GD、PIC和统计学指标BIC评估不同分群方法的分群功效。一个良好的分群功效评估指标应该满足以下3点：（1）显示亚群边界清晰度和亚群间成员的混杂程度；（2）有效地定量表达亚群间变异度的相对大小；（3）灵敏地反映由不同分群方法或亚群个数变化带来的亚群间和亚群内变异的变化。从PCA图上可以清楚看出ADMIXTURE + TagSNPs分群法产生的亚群间的边界清晰度比SNPHylo分群法高，亚群成员间的混杂度比SNPHylo分群法低。本研究结果表明PCA图可以清楚地表达不同分群方法所产生的亚群在边界清晰度和亚群间成员混杂度方面的差异，是评估不同分群方法分群功效的有效指标。PCA作为表达一个分群方法分群功效的直观图示工具在种质资源遗传研究中已经有多年应用并且至今仍在广泛应用[31,37]，它的直观性和准确性在本研究再次得到验证。但PCA并不能有效地定量表达亚群间变异度的相对大小，因此本研究评估了GD、PIC和BIC作为定量指标的有效性。如果分群方法是有效的，一个有效的定量表达亚群间变异度的相对大小的指标所探测到的亚群间相对变异度的预期值应该大于1，本研究结果表明BIC_{BW}均显著大于1而GD_{BW}和PIC_{BW}实际值远小于1，所以BIC是评价分群方法相对分群功效高低的有效定量指标，而GD和PIC是无效指标。本研究结果还表明BIC度量值和PCA图上显示的群间群内变异的相对大小一致。贝叶斯信息指数（BIC）作为度量处理间、处理内变异相对大小的显著性测验的统计指标已被广泛应用[32,35]，它也被用作评估不同RNA基因表达谱数据分群方法的分群功效的指标，表现较好[32]。GD和PIC作为度量基因位点内等位基因多态性的指标在度量群体遗传多样性和筛选分子标记方面得到广泛应用[7-8,38-39]，但作为度量亚群间、亚群内相对变异方面尚未见报道。本研究结果表明GD和PIC显著地低估了亚群间相对变异，不能有效度量亚群间、亚群内变异，不是评估分群方法的分群功效的合适定量指标。

3.2 不同分群方法分群功效的综合评估

本研究结果表明PCA是评估一个分群方法分群功效的直观指标，BIC_{BW}和灵敏度是评估分群功效的定量指标，建议将三者结合，作为评价一个分群方法的分群功效的综合指标。PCA图显示ADMIXTURE+TagSNPs分群法产生的亚群边界清晰，群间混杂少，优于其他3个分群方法，从BIC_{BW}指标值可以看出，当分为3个亚群时（K=3），ADMIXTURE+SNPs分群法BIC_{BW}值最高，为1.719；当分为6个亚群时（K=6），ADMIXTURE+TagSNPs分群法BIC_{BW}值最高，为4.631。从亚群个数由K=3到K=6时不同分群方法对群间相对变异的变化的灵敏度看，ADMIXTURE+TagSNPs分群法灵敏度最高，SNPhylo法最低。利用PCA散点图，BIC_{BW}指标估值和灵敏度估值综合评估4个分群方法得出：ADMIXTURE+TagSNPs分群法产生的亚群边界清晰，亚群间个体混杂少，相对群间异度大，对分群数变化的灵敏度高，总体表现最好，SNPhylo法总体表现最差。本研究需要指出一个分群方法的表现受众多因素的影响，包括输入的SNP个数、软件内部使用的SNP筛选标准和方法，分群方法计算个体间遗传距离的方法和分群（或构图）用的计算机算法等。本研究在比较2种不同分群方法时，考虑到不同软件内部SNP筛选标准和筛选方法的不同，确保2种方法实际被采取的SNP标记相同，有效避免了因输入数据的不同对分群结果的影响。分群软件通常允许使用者选择不同的遗传距离，比如邻接法，允许使用者选择内置的多种不同遗传距离计算方法的一种，不同遗传距离计算方法产生不同的距离矩阵，进而输出不同的分群结果。所以在使用NJ分群法时，一定要了解软件实际使用的是何种遗传距离，确保选用的遗传距离与所输入的标记类型匹配，这样利用NJ法才可以得出功效最好的分群结果。Nei等[40]的研究表明修正欧几里德距离是与离散型数据（如RFLP、SSR、SNP等DNA分子标记数据）最匹配的遗传距离算法[40-41]。如果分群用的输入数据类型是连续性数据（例如转录组和多数代谢组数据），应该采用不同的遗传距离。

3.3 SNP、标签SNP、试验成本和数据质量

最近几年随着高通量DNA测序技术的进步带来的测序成本的快速下降，SNP分子标记越来越广泛地被用于种质资源研究和种质材料分群[9,12-13]。利用重测序和各种简化基因组测序，很容易获得几十万至几百万SNP分子标记位点数据，由于其中不少SNP位点之间存在高度的连锁不平衡或相关性和冗余重复，直接利用大量SNP数据对分群和关联

分析软件构成巨大挑战。分子标记的优点是能够用有限的试验室鉴定成本的增加换来田间大量土地、人力成本和时间成本的节省，如果在后续种质材料鉴定或分子标记辅助育种过程中需要对大量 SNP 位点进行 DNA 鉴定，导致试验室成本的大量增加，分子标记将难以被育种家接受。在实践中研究者通常需根据有限知识挑选使用一小部分 SNP 标记，因此标签 SNP 受到重视。在人类遗传学中，已经开发出标签 SNP 筛选软件，用计算机软件辅助挑选少量的可以代表整个样本主要的遗传信息的 SNP，即标签 SNP（tagSNP），并利用这些标签 SNP 进行类群划分和亲缘关系研究[26-28]，但在农作物中，尚未见到利用标签 SNP 进行种质材料分群的报道。尽管 4 个分群方法所用数据均来源于同一套 525141 个 SNP 位点数据，不同分群方法对原始 SNP 数据的接受和过滤利用方式不同，实际需要输入软件的 SNP 个数和软件实际采用的 SNP 个数可以有显著不同。考虑到使用 SNP 标记的试验室成本（主要包括产生 SNP 的 DNA 测序成本和分子标记使用过程中基因型鉴定的成本），评估一个分群方法不仅要看分群效果也要考虑使用成本特别是试验室成本。

ADMIXTURE+SNPs 分群法、NJ 法和 SNPhylo 分群法需要输入全部 525141 个 SNP 数据，尽管实际上仅使用部分 SNP 位点。本研究开发的 ADMIXTURE+tagSNPs 分群法是 ADMIXTURE+SNPs 分群法的改进，实际输入软件的是通过 Haploview V4.2 软件挑选出的 4849 个 TagSNP 位点，比 ADMIXTURE+SNPs 分群法输入数据少很多，所以试验室成本更低。

影响分群方法的分群效果的一个重要因素是分子标记数据的质量，分子标记数据应该有足够大的遗传变异多态性。多态性信息含量（PIC）是评估分子标记数据变异多态性的一个有效指标[21-22]，尽管本研究的结果表明 PIC 不是评估分群方法分群功效的合适指标。在比较不同分群方法时，本研究采用同一套 SNP 数据，并对数据进行相同的前处理，用 Tassel V5.2 软件以最小等位基因频率（MAF，minimum allele frequency）不小于 5% 作为筛选条件，将分布在 10 条染色体上的 876305 个 SNPs 过滤后得到 525141 个 SNP。本项目数据的最小等位基因频率均值为 0.217，PIC 均值为 0.304。赵久然等[9]报道在 344 份自交系中，3072 个 SNP 标记所检测到的多态信息含量（PIC）为 0.028~0.570，平均 PIC 值为 0.344。吴金凤等[12]利用 1041 个 SNP 位点对 51 份玉米自交系进行基因型分析，最小等位基因频率平均值为 0.359，多态性信息含量（PIC）的变化范围为 0.186~0.375，平均值为 0.345。Yang 等[42]利用 926 个 SNP 对 527 份玉米自交系进行遗传多样性分析，发现平均 MAF 是 0.3，平均 PIC 值是 0.31。Wu 等[43]利用 GBS 测序对 538 份 CIMMYT 玉米自交系的 362008 个 SNP 的分子特征进行分析，发现平均 MAF 值是 0.22，平均 PIC 值是 0.25。可见与发表的研究相比，本研究使用的分子标记数据的质量较好。

参考文献

[1] 王懿波，王振华，王永普，张新，陆利行 . 中国玉米主要种质杂交优势利用模式研究 . 中国农业科学，1997，30（4）：16-24
Wang Y B，Wang Z H，Wang Y P，Zhang X，Lu L X. Studies on the heterosis utilizing models of main maize germplasms in China. Scientia Agricultura Sinica，1997，30（4）：16-24

[2] 王懿波，王振华，王永普，张新，陆利行 . 中国玉米主要种质杂种优势群的划分及其改良利用 . 华北农学报，1998，13（1）：74-80
Wang Y B，Wang Z H，Wang Y P，Zhang X，Lu L X. Division，utilization and the improvement of main germplasm heterosis of maize in China. Acta Agriculturae Boreali-Sinica，1998，13（1）：74-80

[3] 黎裕，王天宇 . 我国玉米育种种质基础与骨干亲本的形成 . 玉米科学，2010，18（5）：1-8
Li Y，Wang T Y. Germplasm base of maize breeding in China and formation of foundation parents. Journal of Maize Sciences，2010，18（5）：1-8

[4] 郭晋杰，赵永锋，张冬梅，祝丽英，黄亚群，陈景堂 . 不同杂种优势群玉米子粒脱水速率分析 . 植物遗传资源学报，2018，19（1）：39-48
Guo J J，Zhao Y F，Zhang D M，Zhu L Y，Huang Y Q，Chen J T. Analysis of grain dehydration rate in different maize heterotic groups. Journal of Plant Genetic Resources，2018，19（1）：39-48

[5] 黎裕，王天宇 . 玉米种质创新——进展与展望 . 玉米科学，2017，25（3）：11-18
Li Y，Wang T Y. Germplasm enhancement in maize：advances and prospects. Journal of Maize Sciences，2017，25（3）：11-18

[6] 刘旭，李立会，黎裕，方沩 . 作物种质资源研究回顾与发展趋势 . 农学学报，2018，8（1）：1-6
Liu X，Li L H，Li Y，Fang W. Crop germplasm resources：advances and trends. Journal of Agriculture，2018，8（1）：1-6

[7] Lu Y，Yan J，Guimarães C T，Taba S，Hao Z，Gao S，Chen S，Li J，Zhang S，Vivek B S，Magorokosho C，Mugo S，Makumbi D，Parentoni S N，Shah T，Rong T，Crouch J H，Xu Y . Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theoretical and Applied Genetics，2009，120：93-115

[8] Guo Z F，Wang H W，Tao J J，Ren Y H，Xu C，Wu K S，Zou C，Zhang J N，Xu Y B. Development of multiple SNP marker panels affordable to breeders through genotyping by target

sequencing (GBTS) in maize. Molecular Breeding, 2019, 39: 37

[9] 赵久然,李春辉,宋伟,王元东,张如养,王继东,王凤格,田红丽,王蕊 . 基于 SNP 芯片揭示中国玉米育种种质的遗传多样性与群体遗传结构 . 中国农业科学, 2018, 51 (4): 626-634
Zhao J R, Li C H, Song W, Wang Y D, Zhang R Y, Wang J D, Wang F G, Tian H L, Wang R. Genetic diversity and population structure of important Chinese maize breeding germplasm revealed by SNP-Chips. Scientia Agricultura Sinica, 2018, 51 (4): 626-634

[10] Cao G Q, Gao Y M, Zhu J. QTL analysis for flag leaf length in a rice DH population under multi environments.Acta Agronomica Sinica, 2007, 33 (2): 223-229

[11] 袁力行,傅骏骅,Warburton M,李新海,张世煌,Khairallah M,刘新芝,彭泽斌,李连城 . 利用 RFLP、SSR、AFLP 和 RAPD 标记分析玉米自交系遗传多样性的比较研究 . 遗传学报, 2000, 27 (8): 725-733
Yuan L X, Fu J H, Warburton M, Li X H, Zhang S H, Khairallah M, Liu X Z, Peng Z B, Li L C. Comparison of genetic diversity among maize inbred lines based on RFLPs, SSRs, AFLPs and RAPDs. Acta Genetica Sinica, 2000, 27 (8): 725-733

[12] 吴金凤,宋伟,王蕊,田红丽,李雪,王凤格,赵久然,蔚荣海 . 利用 SNP 标记对 51 份玉米自交系进行类群划分 . 玉米科学, 2014, 22 (5): 29-34
Wu J F, Song W, Wang R, Tian H L, Li X, Wang F G, Zhao J R, Wei R H. Heterotic grouping of 51 maize inbred lines by SNP markers. Journal of Maize Sciences, 2014, 22 (5): 29-34

[13] 王文斌,徐淑兔,高杰,张兴华,郭东伟,李向阳,薛吉全 . 基于 SNP 标记的玉米自交系遗传多样性分析 . 玉米科学, 2015, 23 (2): 41-45
Wang W B, Xu S T, Gao J, Zhang X H, Guo D W, Li X Y, Xue J Q. Analysis of genetic diversity of maize inbred lines based on SNP markers. Journal of Maize Sciences, 2015, 23 (2): 41-45

[14] Elshire R J, Glaubitz J C, Sun Q, Poland J A, Kawamoto K, Buckler E S, Mitchell S E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 2011, 6 (5): e19379

[15] Glaubitz J C, Casstevens T M, Lu F, Harriman J, Elshire R J, Sun Q, Buckler E S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One, 2014, 9 (2): e90346

[16] Sun X W, Liu D Y, Zhang X F, Li W B, Liu H, Hong W G, Jiang C B, Guan N, Ma C X, Zeng H P, Xu C H, Song J, Huang L, Wang C M, Shi J J, Wang R, Zheng X H, Lu C Y, Wang X W, Zheng H K. SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. PLoS One, 2013, 8 (3): e58700

[17] Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics, 2000, 155: 945-959

[18] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 2009, 19: 1655-1664

[19] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology & Evolution, 1987, 4 (4): 406-425

[20] Lee T H, Guo H, Wang X Y, Kim C, Paterson A H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics, 2014, 15 (1): 1-6

[21] Felsenstein J. PHYLIP-Phylogeny Inference Package (Version 3.2). Cladistics, 1989, 5 (2): 163-166

[22] Schliep K P. Phangorn: phylogenetic analysis in R. Bioinformatics, 2011, 27 (4): 592-593

[23] Felsenstein J. Confidence limits on phylogenies. An approach using the bootstrap. Evolution, 1985, 39: 783-791

[24] Kenneth L, Alexander D H. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics, 2011, 12 (1): 1-6

[25] Tang H, Tang H, Peng J, Wang P, Risch N J. Estimation of individual admixture: analytical and study design considerations. Genetic Epidemiology, 2005, 28: 289-301

[26] Barrett J C, Fry B, Maller J, Daly M J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 2005, 21 (2): 263-265

[27] Zhang K, Qin Z S, Liu J S, Chen T, Waterman M S, Sun F Z. Haplotype block partitioning and Tag SNP selection using genotype data and their applications to association studies. Genome Research, 2004, 14 (5): 908-916

[28] Chia J M, Song C, Bradbury P J, Costich D, Leon N D, Doebley J, Elshire R J, Gaut B, Geller L, Glaubitz J C, Gore M, Guill K E, Holland J, Hufford M B, Lai J S, Li M, Liu X, Lu Y L, McCombie R, Nelson R, Poland J, Prasanna B M, Pyhäjärvi T, Rong T Z, Sekhon R S, Sun Q, Tenaillon M I, Tian F, Wang J, Xu X, Zhang Z W, Kaeppler S M, Jeffrey R I, McMullen M D, Buckler E S, Zhang G Y, Xu Y B, Ware D. Maize HapMap2 identifies extant variation from a genome in flux. Nature Genetics, 2012, 44: 803-807

[29] Reich D, Price A L, Patterson N. Principal component analysis of genetic data. Nature Genetics, 2008, 40 (5): 491-492

[30] Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 2012, 28 (24): 3326-3328

[31] 徐国平,范濂 .Timopheevi 胞质杂种小麦育性恢复性的基因型 × 环境互作分析 . 遗传学报, 1986, 13 (6): 437-446
Xu G P, Fan L. Analysis of the genotype × environment interaction of fertility restoration in Timopheevi CMS hybrid wheat. Acta Genetics Sinica, 1986, 13 (6): 437-446

[32] Shu G P, Zeng B Y, Wright D, Smith O. Impact of data transformation on the performance of different grouping methods and group number determination statistics for analyzing gene expression profile data. Proceedings of 13th KSU Conference on Applied Statistics in Agriculture, 2002, 13: 94-110

[33] Liu K, Muse S V. PowerMaker: An integrated analysis environment for genetic maker analysis. Bioinformatics, 2005, 21 (9): 2128-2129

[34] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika, 2008, 95: 759-771

[35] Gao X, Song P. Composite likelihood Bayesian information

criteria for model selection in high-dimensional data. Journal of the American Statistical Association，2010，105：1531-1540

［36］ 李念念 . 黄淮海地区玉米种质群体结构和遗传多样性的 GBS-SNP 分析 . 郑州：郑州大学，2018
Li N N. Analysis of population structure and genetic diversity of maize germplasm in Huang-Huai-Hai region by GBS-SNP. Zhengzhou：Zhengzhou University，2018

［37］ 张晨，云岚，李珍，王俊，郭宏宇，盛誉，石子英，徐学宝 . 新麦草种质的 SSR 遗传多样性及群体结构分析 . 植物遗传资源学报，2019，20（1）：48-59
Zhang C，Yun L，Li Z，Wang J，Guo H Y，Sheng Y，Shi Z Y，Xu X B. Genetic diversity and structure analysis in Psathyrostachys Nevski population using SSR markers. Journal of Plant Genetic Resources，2019，20（1）：48-59

［38］ 崔永霞，张名昌，白建荣，程宇坤，张效梅，任元 . 利用 SSR 分析山西省玉米地方品种的遗传多样性 . 植物遗传资源学报，2012，13（5）：810-818
Cui Y X，Zhang M C，Bai J R，Cheng Y K，Zhang X M，Ren Y. Analysis of genetic diversity of maize landraces in Shanxi by SSR markers. Journal of Plant Genetic Resources，2012，13（5）：810-818

［39］ 刘海忠，宋炜，王宝强，王江浩，张全国，张动敏，李兴华，魏剑锋，李荣改 . 120 份欧美玉米自交系的遗传多样性分析 . 植物遗传资源学报，2018，19（4）：676-684
Liu H Z，Song W，Wang B Q，Wang J H，Zhang Q G，Zhang D M，Li X H，Wei J F，Li R G. Genetic diversity analysis of 120 European and American maize inbred lines. Journal of Plant Genetic Resources，2018，19（4）：676-684

［40］ Nei M，Li W H. Mathematical model for studying genetic variation terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America，1979，76：5269-5273

［41］ SAS/STAT® 9.2 User′s Guide. SAS Institute Inc.，Cary，NC，USA，2008

［42］ Yang X H，Gao S B，Xu S T，Zhang Z X，Prasanna B M，Li L，Li J S，Yan J B . Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Molecular Breeding，2011，28（4）：511-526

［43］ Wu Y S，Vicente F S，Huang K J，Dhliwayo T，Costich D E，Semagn K，Sudha N，Olsen M，Prasanna B M，Zhang X C，Babu R. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theoretical and Applied Genetics，2016，129：753-765