

# 黄瓜多样性固定核心样本集的构建与评价

张 微, 李斯更, 沈 镒, 李锡香

(中国农业科学院蔬菜花卉研究所/农业部蔬菜作物基因资源与种质创制北京科学观测实验站, 北京 100081)

**摘要:** 利用 149 对具有多态性的 InDel 引物对 473 份黄瓜初选核心种质自交系进行遗传多样性分析。采用 3 种方法 12 种取样比例对该初级遗传多样性固定群体进行抽样获得候选多样性固定核心样本集 (*GDFCC*), 使用等位基因数 ( $N_a$ )、有效等位基因数 ( $N_e$ )、Shannon's 信息指数 ( $I$ )、基因多样性指数 (gene diversity)、多态性信息含量 ( $PIC$ )、总等位点数 (total number of loci)、等位点保留百分率 (retention rate of loci) 评价候选多样性固定核心样本集的多样性和代表性, 结果表明, 采用逐级聚类 + 稀有基因优先取样法并按照 15% 取样比例构建出的多样性固定核心样本集的效果较好。比较发现, 该核心样本集的  $N_e$ 、 $I$ 、基因多样性指数和  $PIC$  值均接近或高于初级遗传多样性固定群体, 且对原始群体的等位点的保留百分率为 99.68%。入选多样性固定核心样本集的材料来自 15 个国家和国内 18 个省市。该研究为今后黄瓜优异基因资源的挖掘利用提供了代表性强、覆盖度广、遗传稳定的研究群体, 将有利于黄瓜种质资源的高效研究利用。

**关键词:** 黄瓜; InDel 标记; 核心种质; 遗传多样性固定群体; 遗传多样性

## Establishment and Analyses of a Genetic Diversity Fixed Core Collection of Cucumber (*Cucumis sativus* L.)

ZHANG Wei, LI Si-geng, SHEN Di, LI Xi-xiang

(Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences/Beijing Research Station of Vegetable Crop Gene Resource & Germplasm Enhancement, Ministry of Agriculture, Beijing 100081)

**Abstract:** In this study, 149 pairs of InDel primers were used to analyze the genetic diversity of 473 inbred lines of a primary core collection in cucumber. Three methods with twelve sampling proportions were applied to sample the candidate genetic diversity fixed core collections (*GDFCC*), number of alleles ( $N_a$ ), number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), gene diversity, polymorphism information content ( $PIC$ ), total number of loci and retention rate of loci were used to test diversity and representativeness of the candidate *GDFCC*s. The results indicated that the method of stepwise clustering plus rare genes preferred sampling with 15% sampling proportion was more efficient for the *GDFCC* construction. By comparison,  $N_e$ ,  $I$ , gene diversity and  $PIC$  of the *GDFCC* were higher than those of the primary core collection, and the retention rate of loci in the *GDFCC* reached 99.68%. The materials of the *GDFCC* comes widely from eighteen provinces in China and fifteen foreign countries with a rich diversity and representation. This study provides with a valuable population with genetic stability and extensive representativeness for promoting efficient research and utilization of cucumber germplasm resources.

**Key words:** cucumber; InDel marker; core collection; genetic diversity fixed core collection; genetic diversity

黄瓜 (*Cucumis sativus* L.) 属于典型的异花授粉作物, 是世界普遍栽培的重要蔬菜。由于黄瓜杂优品种的大面积推广, 栽培品种的遗传背景变得越来  
越狭窄, 拓展黄瓜作物的遗传多样性显得十分必要。

收稿日期: 2015-04-26 修回日期: 2015-06-02 网络出版日期: 2016-04-06

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20160406.1114.004.html>

基金项目: 国家自然科学基金项目 (31171961); 农业部园艺作物生物学与种质创制重点实验室项目; 国家农作物种质资源平台项目

第一作者研究方向为蔬菜种质资源。E-mail: vivizh2007@sina.com

通信作者: 李锡香, 研究方向为蔬菜种质资源。E-mail: lixx0612@163.com

我国国家蔬菜种质中期库收集黄瓜资源 2000 余份,其中绝大部分属于地方品种,遗传杂合度高,严重制约了资源的研究和利用。虽然研究人员对黄瓜种质资源开展了一系列表型<sup>[1-6]</sup>和分子多样性鉴定<sup>[6-10]</sup>以及核心种质研究<sup>[11-15]</sup>,但是前人的研究多基于有限的典型材料和天然混交的原始种质,研究结果在指导资源库大量资源的研究和利用中的作用非常有限。为了提高作物遗传多样性鉴定和核心种质研究结果的效度,英国国际园艺研究中心的 G. J. King 等<sup>[16]</sup>提出了多样性固定基础群体(*DFFS*)的概念,即代表某物种基因库内多样性的、遗传固定的自交系收集品种。他们在构建芸薹属作物核心种质的基础上,通过小孢子培养和连续多代自交的方法,构建了 3 个多样性固定基础群体并得到了有效应用,这为我们解决上述资源研究中的问题提供了有益的借鉴。

InDel 分子标记与 SSR 标记所反映出的基因组遗传信息不尽相同,具有多态性高、特异性强、稳定性好、检测容易、在基因组分布广等优点,在水

稻<sup>[17-18]</sup>、玉米<sup>[19]</sup>、大白菜<sup>[20-21]</sup>、甘蓝<sup>[21-22]</sup>等作物上应用较多,在黄瓜苦味基因挖掘<sup>[23]</sup>、品种纯度鉴定<sup>[24]</sup>和抗病基因定位<sup>[25-27]</sup>等方面也有应用。本研究基于黄瓜表型初选核心种质多代自交构建的初选遗传多样性固定群体,利用黄瓜重测序信息设计的分布于黄瓜 7 条染色体上的 149 对 InDel 标记对其遗传多样性进行分析,结合多种取样方法和取样比例构建多样性固定核心样本集,为促进黄瓜种质资源高效研究和利用奠定基础。

## 1 材料与方法

### 1.1 试验材料

试验材料为基于国家蔬菜种质资源中期库保存的 2000 多份国内外黄瓜种质资源构建的表型初级核心种质经过多代自交形成的 473 份自交系,即初级多样性固定群体,其中来自印度、埃及、匈牙利、美国、俄罗斯、日本等 35 个国家的材料 116 份,来自中国东北、华北、华中、华东、华南、西北、西南等 7 个地区的材料 357 份(表 1)。

表 1 黄瓜初级核心种质的来源

Table 1 The source of cucumber germplasm in the primary core collection

原产地	份数	原产地	份数	原产地	份数
Origin	Quantity	Origin	Quantity	Origin	Quantity
阿富汗	4	西班牙	3	中国江苏	8
捷克	2	叙利亚	4	中国江西	4
刚果	1	土耳其	7	中国山东	33
丹麦	1	乌克兰	2	中国浙江	3
埃及	2	美国	12	中国福建	6
塞尔维亚	1	赞比亚	2	中国广东	11
前苏联	3	津巴布韦	1	中国台湾	1
法国	1	阿曼	1	中国河南	16
匈牙利	5	埃塞俄比亚	1	中国湖北	15
印度	19	西德	1	中国湖南	6
伊朗	3	希腊	1	中国甘肃	3
伊拉克	1	以色列	2	中国宁夏	2
日本	11	中国黑龙江	6	中国青海	1
哈萨克斯坦	1	中国吉林	22	中国陕西	1
肯尼亚	1	中国辽宁	21	中国新疆	4
毛里求斯	1	中国北京	3	中国广西	1
尼泊尔	1	中国河北	28	中国贵州	4
荷兰	5	中国内蒙古	8	中国四川	25
巴基斯坦	2	中国山西	8	中国云南	78
波兰	3	中国天津	9	中国重庆	3
波多黎各	1	中国安徽	7	未知	20
俄罗斯	10				
总数 Total	473				

## 1.2 试验方法

**1.2.1 DNA 提取** 在黄瓜植株 4 叶 1 心时,每份材料选 5 个生长正常单株的幼嫩叶片作为混合样品,采用改良 CTAB 法提取基因组 DNA。使用微量分光光度仪检测 DNA 浓度,采用 0.8% 琼脂糖凝胶电泳检测 DNA 纯度和完整性。最终稀释浓度至 20 ng/L, -20 °C 冰箱保存备用。

**1.2.2 PCR 扩增与产物检测** InDel 引物根据黄瓜基因组重测序预测得到的 InDel 位点设计,这些引物分布于黄瓜的 7 条染色体上。引物由英杰生命科技有限公司 (Invitrogen) 合成,通过试验筛选出多态性引物 149 对(表 2),建立并优化反应体系。PCR 反应体系为 15  $\mu$ L,其中 DNA 模板(15 ng/ $\mu$ L) 2  $\mu$ L, 10  $\times$  Buffer 1.5  $\mu$ L, Mg<sup>2+</sup> (25 mM) 1.2  $\mu$ L, dNTP(2.5 mM)0.2  $\mu$ L,上游引物(10 uM)0.2  $\mu$ L,下游引物(10 uM)0.2  $\mu$ L, Taq 酶(2.5 U/ $\mu$ L)

0.2  $\mu$ L, ddH<sub>2</sub>O 9.5  $\mu$ L。反应程序为:94 °C 5 min; 94 °C 30 s, 56 °C 30 s, 72 °C 1 min, 25 个循环;72 °C 7 min; 4 °C 保存。反应产物用 8% 的非变性聚丙烯酰胺凝胶电泳检测,银染拍照。

**1.2.3 数据统计与分析** 对 InDel 扩增产物进行统计,有带记为“1”,无带记为“0”,缺失记为“9”,并将原始数据分别转换为用于 PowerMarker V3.25 软件和 Popgen 3.2 软件使用的数据类型。使用 PowerMarker V3.25 软件分析总等位位点数(total loci)、等位位点保留百分率(retention rate of loci)、基因多样性(gene diversity)、多态性信息含量(PIC),使用 Popgen 3.2 软件分析等位基因数(*N<sub>a</sub>*)、有效等位基因数(*N<sub>e</sub>*)、Shannon's 信息指数(*I*)。使用 PowerMarker V3.25 软件和 MEGA 6.06,计算群体遗传距离(Nei'1972)并采用 UPGMA 法对群体进行遗传相似性聚类并构建系统进化树。

表 2 InDel 引物信息表

Table 2 Basic information of the InDel primers

引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')	引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')
CuIn0001F	Chr1	ATCAAACGGAGAGCTAAACA	CuIn0046F	Chr4	CAAAAACCAGAACCTGAAAG
CuIn0001R	Chr1	TCAAGGTCTAATGATCGAGG	CuIn0046R	Chr4	CTGGGAATAAGCATTGTAG
CuIn0009F	Chr6	CCCTTTCATTCTTCTAACTCAG	CuIn0047F	Chr4	CTTTTCCCTTCACACATCAT
CuIn0009R	Chr6	AAATGACATGAGTCAAAGCC	CuIn0047R	Chr4	TACGATATCACAACTGGTGG
CuIn0011F	Chr6	TATAGAATCCAGGGTCCAA	CuIn0052F	Chr6	GTACGACAAATGGAGAATGA
CuIn0011R	Chr6	AAAAAGCTGTCGTAATGGAG	CuIn0052R	Chr6	TGTTTTAAGGTTTCTGCCC
CuIn0019F	Chr6	CATCACTGTTTATCAACCCC	CuIn0055F	Chr6	AAACACACACCTTCTGCTCT
CuIn0019R	Chr6	CGTAGATGCTGACAACAAAA	CuIn0055R	Chr6	GCTCGTTGAAAAGAAGCAT
CuIn0020F	Chr6	AATAGTTGAGAAAAAGGGG	CuIn0059F	Chr7	CAAAGTTGGTGGTGTATTAT
CuIn0020R	Chr6	GAAGGTTTGTGTGTTGGAAT	CuIn0059R	Chr7	GTTGAGAATTTAGATCCCC
CuIn0021F	Chr6	GAAAGCAAGGAGGAAAAATCT	CuIn0062F	Chr3	GAGAATGGTGAAGAAGATGG
CuIn0021R	Chr6	CACTACTTCTCTCTCTCCT	CuIn0062R	Chr3	GTCCCACTCACTCTCTCTCA
CuIn0022F	Chr1	CTACAACAAAAACCCACCTC	CuIn0064F	Chr6	CCAATCTTACAATGGTGGT
CuIn0022R	Chr1	TCTCCTTGATTGATTGTCC	CuIn0064R	Chr6	AGCCGTCATCTTTTGTTTAG
CuIn0024F	Chr2	GATAGAGACTACTGTCGGTGAA	CuIn0065F	Chr6	GTGTAACCCAACCTGTATCAA
CuIn0024R	Chr2	AGAGACCATTATTGTTGGGT	CuIn0065R	Chr6	CTTTATCAACCCCTCGTTTG
CuIn0026F	Chr2	AATACCCACTTGTFTTGTGC	CuIn0066F	Chr6	GGTTCCTGGTGATTGATTAG
CuIn0026R	Chr2	ATGATAGGAGGGTGGAGATT	CuIn0066R	Chr6	GCTAGGACCTGCAAAAAGTTA
CuIn0028F	Chr2	ACCTTTCAGCTATCACCAAT	CuIn0070F	Chr2	AAAGAAAGTGTGGGTACGA
CuIn0028R	Chr2	TCACCCTCTTTATCACCTACA	CuIn0070R	Chr2	AGAGGTGTTAGGCAAAGACA
CuIn0032F	Chr3	GCTTGTACCATGCAATTGTT	CuIn0073F	Chr5	GACCTCTGCTGGTAAGGATT
CuIn0032R	Chr3	CCCTGCTTATTGAGATGAC	CuIn0073R	Chr5	TGAGAAAAATAGGAGTGGGA
CuIn0035F	Chr3	TCAGTTCAGCCTCTACCTA	CuIn0074F	Chr7	AGATTAAGAACCCTCTTGGG
CuIn0035R	Chr3	ACCAATGCCTAGACTGAAGA	CuIn0074R	Chr7	GAACAACAAGAACTCCAAGC
CuIn0036F	Chr3	TGATCAACCTCCACCTTTAC	CuIn0075F	Chr7	ATCTGCTTGCTATCCAATTC
CuIn0036R	Chr3	TTAGCGATGTTGAAGTGATG	CuIn0075R	Chr7	CGTTCCTAACCATATTTCTCG
CuIn0038F	Chr3	TTGAAACACACTCCAAGACA	CuIn0078F	Chr2	GATCAACACGAGTTCAGAT
CuIn0038R	Chr3	GAAGAGGGCCCATTTAGTAT	CuIn0078R	Chr2	GATCAGATTAGGTTTTTGGC
CuIn0041F	Chr4	TTAGTGCTGGAGAGATGGAG	CuIn101F	Chr1	CTAAGCTCAAGCCAGAAGAA
CuIn0041R	Chr4	TGTATGTGCATGACTTTTCGT	CuIn101R	Chr1	ACAGTTTTGTGATGAAAGGC
CuIn0043F	Chr4	AGTGGATATATTGTTGGGG	CuIn102F	Chr1	CATATTTACAATCTTGGGGC
CuIn0043R	Chr4	AGCTACCTCCTCTCAACT	CuIn102R	Chr1	CCTGATACGAAAAGAACCAC

表 2(续)

引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')	引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')
CuIn103F	Chr1	ACAGCAACATGTCCTACTAA	CuIn214F	Chr2	GATTTCCATGTTCTCATCGT
CuIn103R	Chr1	TGGAGATATCAATGTCGATG	CuIn214R	Chr2	TTTGTGAGACTCATTCTTCG
CuIn104F	Chr1	CACGTTCTGATTTTACACA	CuIn215F	Chr2	GTGGCCATTATCATTITGAC
CuIn104R	Chr1	ACTCACCAATTGATCTGGTA	CuIn215R	Chr2	ACCATGCCAAGACTTAACTG
CuIn105F	Chr1	ATGAAGAAGTCGGTTTCGTA	CuIn216F	Chr2	GACCGACACGTAAAAGTTGT
CuIn105R	Chr1	GCGCAGGAATAAAAAACATC	CuIn216R	Chr2	AAATGTAGAAAAGCATGTGGG
CuIn107F	Chr1	TTGTCAACTGTACGATCTGG	CuIn217F	Chr2	ACAATGTGCTAAGATGCACA
CuIn107R	Chr1	ATCTGGTACTGCCACAGAAG	CuIn217R	Chr2	GGCTTCTTTAGCCACAAATA
CuIn108F	Chr1	CAATTTGTGGGCTGTAGTTT	CuIn218F	Chr2	AAGGTAAGATTATCCCTCCG
CuIn108R	Chr1	TTATCTTCGTCCCTCTGCTA	CuIn218R	Chr2	GAGGCCCTTGATAAACTAT
CuIn109F	Chr1	TCTTCACTCCATGGTTCTTC	CuIn301F	Chr3	TCATTTCCGAACAGATTAGGG
CuIn109R	Chr1	GGTGAGAGAGAAAACATTTGC	CuIn301R	Chr3	GCTGTGGAAAATTCATATG
CuIn110F	Chr1	CATCGTGTGGACTATGATGA	CuIn302F	Chr3	GTTACCTAGTTCGATTCCCA
CuIn110R	Chr1	TATCCTCCATGAGATTTTCG	CuIn302R	Chr3	TTCCCTAATTTCGGTGAAGA
CuIn111F	Chr1	CATAAGCCCTCAGAGATTTG	CuIn303F	Chr3	ATGTAATGCAGCGGTAAAA
CuIn111R	Chr1	AGAGTGGCAACGAGATTCTA	CuIn303R	Chr3	CCACTAAAGACTCAGAGTTGC
CuIn112F	Chr1	GAATGCAGAGAGAGGAGTTG	CuIn304F	Chr3	TAGGGTGAGAGAGGATGAAA
CuIn112R	Chr1	CTGTTTTGTGAGGGTGTTT	CuIn304R	Chr3	TGTGTGATTAACATGGCATC
CuIn113F	Chr1	CCCTTGTAAGGTTTGTATTCC	CuIn306F	Chr3	ATATCTGGTCTGCCAATGTT
CuIn113R	Chr1	TTCTCTTCTCTTTTCGGTTTG	CuIn306R	Chr3	CTTGACCATTAATAACTGGG
CuIn114F	Chr1	TACTGTGAAAAGGTTATGCC	CuIn307F	Chr3	GACTGATGAACCTCAACCTG
CuIn114R	Chr1	CACATCAGCCCTAATGATTT	CuIn307R	Chr3	AGCTTCTGACTTTGTGGTTG
CuIn115F	Chr1	GTGACGAGAAATGAAAAGAG	CuIn308F	Chr3	ATTCCGAGCTGGAGTTGTAT
CuIn115R	Chr1	TAGGCATTGGTTTGTGTGTA	CuIn308R	Chr3	ACAATGCTACGAAGTACCAA
CuIn116F	Chr1	AAGCCCTCAACAGCACTAT	CuIn309F	Chr3	TGTTTTGATGTTGCTAGAGG
CuIn116R	Chr1	ATTTTACTCCGAGCAAGTGA	CuIn309R	Chr3	TTCTCAAACCATCCTTGACT
CuIn117F	Chr1	CATTGTCAAACCTCATGTTG	CuIn310F	Chr3	CCTTTCTGGATGTTCTACGA
CuIn117R	Chr1	CCATTTCTTGAAGTTCCCTT	CuIn310R	Chr3	CATGAATGGATGCAACTG
CuIn118F	Chr1	AAGACAACAATCACCTTGG	CuIn311F	Chr3	GTTGGTTGCACCTTTTTTC
CuIn118R	Chr1	CGAGACACAAGACATGAAAA	CuIn311R	Chr3	ATTCTAGGATCCCTGCAGTT
CuIn119F	Chr1	TTGACAGCAACCTAGGAACT	CuIn314F	Chr3	AAGAAACCTTTAGGACAGGG
CuIn119R	Chr1	GGTTTACTTTGATGCCACTG	CuIn314R	Chr3	TGAGTGGTAAAAATGCACTGA
CuIn201F	Chr2	CGTTGTTGGATCATCATGTA	CuIn315F	Chr3	CAACAAAGTTACAGTACCCCA
CuIn201R	Chr2	ACAATCACAAACGAAAGACC	CuIn315R	Chr3	GCATGCATACTAATCCCAAC
CuIn202F	Chr2	CTTTCACAATCCCATCTCTC	CuIn316F	Chr3	TTTGTCTAAACACTCCCTC
CuIn202R	Chr2	GATGCCATTAAGACAAAACC	CuIn316R	Chr3	TGTCCAAGGTTGACTCTCTT
CuIn203F	Chr2	AGCTTCAATGCAAAAAGGTAG	CuIn317F	Chr3	CCAACATCACATTTCAAGCTA
CuIn203R	Chr2	GTCGATTCTTGTTTGGGATA	CuIn317R	Chr3	GAGGTAGGGAAGTTTCAGG
CuIn204F	Chr2	TCTTTTTGACAAAAGTGGAC	CuIn318F	Chr3	GTAAAAAGAGCAAACGTCCA
CuIn204R	Chr2	TGTGAGAAGCAAGAAGGTTT	CuIn318R	Chr3	ACTCTTCTCCAAAACCTTCCC
CuIn205F	Chr2	TGTGCCGACAGATATAACAA	CuIn319F	Chr3	TAGTAAAGGACCAACCCAA
CuIn205R	Chr2	CAACGGGTAAAAGAGAAAGA	CuIn319R	Chr3	CGTCAATGAAATGGAGAGTT
CuIn206F	Chr2	CATGCATCAATCACCATCT	CuIn320F	Chr3	ACACCTCGAGTTTGAATGTT
CuIn206R	Chr2	AAGCCCTAGCTACCTCTAA	CuIn320R	Chr3	GACCTCCAACGTATCCATTA
CuIn207F	Chr2	GTCTTGAGTCATTCGTTTGAG	CuIn321F	Chr3	GGAGCCAAGTAAAGAAAGGTT
CuIn207R	Chr2	GCGCCTATAACTTACCTTTG	CuIn321R	Chr3	GGACCAAGTTACCAAGTACG
CuIn208F	Chr2	CTTTAGATCCCTTTTCGGATT	CuIn322F	Chr3	TGCTCAGTTTCTTTATCTACC
CuIn208R	Chr2	AACAACCGTTCCAAACTATG	CuIn322R	Chr3	GATATCACCAGACCAAAAAG
CuIn209F	Chr2	CTAGGTAAACTCCCTTTTCC	CuIn323F	Chr3	CAGAAGAAGGAAAGAAAGCAA
CuIn209R	Chr2	GGATCAAATCCTAAACCCA	CuIn323R	Chr3	AGGAAAAGGGAAGAAAGAA
CuIn210F	Chr2	TTTGGCTTAAAGATACTTGGC	CuIn324F	Chr3	GCATAACTTTTTCGAATGACC
CuIn210R	Chr2	AATCTTCTGCTTGGGATG	CuIn324R	Chr3	GCATCAAGGTTGTGGATAAT
CuIn211F	Chr2	ATCGTTAGACCTTCATGTGG	CuIn325F	Chr3	GGTTTTGATCCCTTTGTAAAG
CuIn211R	Chr2	TCATCCCTGCAGTATATGTG	CuIn325R	Chr3	ACTTAAAGCCCAATGCCATTTC
CuIn212F	Chr2	TGGTAATGGGAGATCATTGT	CuIn326F	Chr3	GATAATCCCTACACATCCAG
CuIn212R	Chr2	TATAGCAACAAAAAGGGGTG	CuIn326R	Chr3	TGTTTTCTCCCTGTTCATC
CuIn213F	Chr2	GAGCTCAATTTGGAGCATAAC	CuIn327F	Chr3	CTAATTTGGTTGTGGGAT
CuIn213R	Chr2	CAATCAATCCTTTTCTCTGC	CuIn327R	Chr3	TCGTATAAACGGCTCTTGT

表2(续)

引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')	引物编号 Primer code	染色体 Chromosome	序列(5'-3') Sequence(5'-3')
CuIn401F	Chr4	AGTGGATTGGATTTTCTGTG	CuIn604R	Chr6	CTTTTCATTGATGTTCTCC
CuIn401R	Chr4	CAACAAATTTGAGGTGAGGT	CuIn605F	Chr6	ATTCCTTCTGAGTTGGTTGA
CuIn402F	Chr4	ACAGCCACAAATATCATTTTC	CuIn605R	Chr6	CAGAGTCCAAGATTTCCAAG
CuIn402R	Chr4	CAGATTCGTCAAGTTCTCTC	CuIn606F	Chr6	ATCCACCTCCTAGCAACTTT
CuIn403F	Chr4	CCCCTCAACAAGAATAATGA	CuIn606R	Chr6	CCCTTGGGTGAATGTCTAAT
CuIn403R	Chr4	TGACATGGAAAGACTAATGG	CuIn607F	Chr6	GTGTTTCCAAAACGCTAGTA
CuIn404F	Chr4	GAATGTAGAGCAGTTCACAC	CuIn607R	Chr6	AGCACAAAGGTTAGAGCAAG
CuIn404R	Chr4	TACCATGAGAAAATTCACCC	CuIn608F	Chr6	AGCATGATCTCAAGTGAAT
CuIn405F	Chr4	TTCCTATGGAAGTACATGG	CuIn608R	Chr6	ATCCAGATGTTTGGTTGAG
CuIn405R	Chr4	CAGTGTGAAGAATTCAACTACC	CuIn609F	Chr6	ATGTGTGCATGCTAGGTCTT
CuIn406F	Chr4	CGCAAGTCAGTTGAACATTA	CuIn609R	Chr6	CTCCGTTCAATTTGTTAGTGG
CuIn406R	Chr4	GGCTAGCTCAACTTCAGAAA	CuIn610F	Chr6	ACTTACTGCATTTGGGTGTC
CuIn407F	Chr4	CCAGCTATTTACCACAAAACAC	CuIn610R	Chr6	TTAGAATCAACTGCCTCCTT
CuIn407R	Chr4	TCTAGGGCGAGTTGAGTTAC	CuIn611F	Chr6	GGCAAGTTGGTGAATAAAGA
CuIn408F	Chr4	ACTGATCTGTTTTCCCTTTT	CuIn611R	Chr6	CTACGGTGGATCTGTGATTT
CuIn408R	Chr4	TGAACTCGATGGTACCTTTT	CuIn613F	Chr6	CATATATTGTCAATGCCACG
CuIn409F	Chr4	TACGGGAATGAAGGAGATAA	CuIn613R	Chr6	CTTCCATGTCAAGATCCACT
CuIn409R	Chr4	GATCGAATGTAAACCTTGT	CuIn614F	Chr6	TAGCAGATTCAAAGAGGAGG
CuIn411F	Chr4	TGATTTTAAACAGCTCTTGGC	CuIn614R	Chr6	ATCAGTGTCTAAAAGGCAAC
CuIn411R	Chr4	GGGGATATAATTGAACAGCA	CuIn615F	Chr6	ATTATGTCCGGTAGGTTTTCG
CuIn412F	Chr4	ATATCCAACACAAAACCAGC	CuIn615R	Chr6	ACTCTTAAACACAGCCACAT
CuIn412R	Chr4	TGAAATCTATAAACCCACG	CuIn616F	Chr6	AGATCATAAAAACGTGGATGC
CuIn414F	Chr4	TTTAGCATTTAGTGGGGAAG	CuIn616R	Chr6	CTGCAACTTTGGGTCTTAT
CuIn414R	Chr4	GAGAATTCATCTTTACCCCC	CuIn618F	Chr6	ATAGCTTGTCCACCTCTTG
CuIn415F	Chr4	AGGAAGTTTGTCTGAAGGAA	CuIn618R	Chr6	CAAACAAGACTCCCAATA
CuIn415R	Chr4	GTGAGAGGTTCCATTTTCAG	CuIn619F	Chr6	CAATACCTTTGACGCTGTTA
CuIn502F	Chr5	GATTCCTGAGACAGAAAAG	CuIn619R	Chr6	GCGAGTTTTATCCCATCTAC
CuIn502R	Chr5	CTCGGGTATGTTCTTTCGTA	CuIn620F	Chr6	AAAAGGGTCTCTGCCTCTT
CuIn503F	Chr5	TGAAATAAGTTGGTTCGAG	CuIn620R	Chr6	CAACTGTGGTGAATGTTGT
CuIn503R	Chr5	GCCTGGTCGATAATATCTTT	CuIn621F	Chr6	CCCAAGTGTATCTCATGCTT
CuIn504F	Chr5	TCAGGACTCAACCTTAAAGC	CuIn621R	Chr6	TTTTACTTCGATTTCTGGAC
CuIn504R	Chr5	TGACAAAATGTACGACATGG	CuIn622F	Chr6	GTACGGAAITGCTTGGTG
CuIn505F	Chr5	CAATCGATGTGTTTCATCC	CuIn622R	Chr6	ACCCTCCATTGTCTCTCTCT
CuIn505R	Chr5	CTTAGTTTGCAAATCGAGCA	CuIn623F	Chr6	AAACTAGGGTGAAGAGGAC
CuIn506F	Chr5	GCTTTTCTAAGCTTGTTTGG	CuIn623R	Chr6	AACTTCCAAACCGCTTAGAT
CuIn506R	Chr5	TTGGCAGGATGAAAAGTAAC	CuIn701F	Chr7	GATTTGCTTACAGTCCAAT
CuIn510F	Chr5	GATGAGCAATCCATTAACCT	CuIn701R	Chr7	GCCCATGTTTACCAATACAG
CuIn510R	Chr5	TGGGTAAGTCTGCTTACAT	CuIn702F	Chr7	TCAACTGTAAAAGTCCCGTT
CuIn511F	Chr5	CCCATATCAACTCATCAAGG	CuIn702R	Chr7	GTATGGTGTGCTTGTGATTG
CuIn511R	Chr5	CCAACGGTAGAGATTTTCATC	CuIn703F	Chr7	TTTACTCTTTTACCCTGGGA
CuIn512F	Chr5	GCAAACCTTCTGACTGGCTAC	CuIn703R	Chr7	CACGAGTGATTTTAAACCGA
CuIn512R	Chr5	GGATGGACGTATTGTCAATTT	CuIn704F	Chr7	CAATTTGCAGTTTGTGACAG
CuIn514F	Chr5	TGATGATGACGACGAGTAAA	CuIn704R	Chr7	CCTTATTTTTGTCCGTGTTT
CuIn514R	Chr5	CAACATGTTTCATTTCCAGTG	CuIn706F	Chr7	AGTTGTGGCTCTTAGATGGA
CuIn515F	Chr5	CTCATCAAGATTTCAAAGAGC	CuIn706R	Chr7	TAGGTTTCAAATGAACTGGC
CuIn515R	Chr5	CGGTCAATTTCCATAGTTAG	CuIn707F	Chr7	GAAAGTGCCTTATTGTGACC
CuIn516F	Chr5	GGTTTTGTTTTGGGTAATG	CuIn707R	Chr7	TTAGCACATTTACAGGATC
CuIn516R	Chr5	AATTATCCCATTGTGCTCTCC	CuIn703F	Chr7	TTTACTCTTTTACCCTGGGA
CuIn517F	Chr5	GAAAAAGGGAGGCTATATC	CuIn703R	Chr7	CACGAGTGATTTTAAACCGA
CuIn517R	Chr5	TAGCCTTAGGCCAATAATTC	CuIn704F	Chr7	CAATTTGCAGTTTGTGACAG
CuIn518F	Chr5	CACAAACCAACCTGAAAT	CuIn704R	Chr7	CCTTATTTTTGTCCGTGTTT
CuIn518R	Chr5	TCATTCTGCTCTCTCTCTC	CuIn706F	Chr7	AGTTGTGGCTCTTAGATGGA
CuIn601F	Chr6	CTCATATTTGCATGGAGTTG	CuIn706R	Chr7	TAGGTTTCAAATGAACTGGC
CuIn601R	Chr6	GCTATCTATGCAAAGCCCT	CuIn707F	Chr7	GAAAGTGCCTTATTGTGACC
CuIn602F	Chr6	AGGAATACACAGACATACCACA	CuIn707R	Chr7	TTAGCACATGTACGAGCATC
CuIn602R	Chr6	TGAGAGGCATAAGGGTATTG	CuIn703F	Chr7	TTTACTCTTTTACCCTGGGA
CuIn603F	Chr6	GGAAAAGTACTCCATAGCA	CuIn703R	Chr7	CACGAGTGATTTTAAACCGA
CuIn603R	Chr6	GGGAGCATAAGTTGAGAA	CuIn704F	Chr7	CAATTTGCAGTTTGTGACAG
CuIn604F	Chr6	AATCACGTCATTCCTTGTGTC	CuIn704R	Chr7	CCTTATTTTTGTCCGTGTTT

**1.2.4 多样性固定核心样本集构建** 分别采用 3 种取样方法,即分组随机取样法、分组聚类+稀有基因优先取样法和逐级聚类压缩+稀有基因优先取样法,均分别按 5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90% 的比例对初级多样性固定群体进行取样,即每种方法按不同比例抽样构建出 12 个候选多样性固定核心样本集,3 种方法共构建 36 个候选多样性固定核心样本集。

表 3 分组随机取样法构建候选多样性固定核心样本集

Table 3 The candidate genetic diversity fixed core collections constructed by grouping random sampling method

分组 Groups	份数 Quantity	5%	10%	15%	20%	25%	30%	40%	50%	60%	70%	80%	90%
国外 Foreign countries	116	6	12	17	23	29	35	46	58	70	81	93	104
东北 Northeast China	49	2	5	7	10	12	15	20	25	29	34	39	44
华北 Northern China	56	3	6	8	11	14	17	22	28	34	39	45	50
华中 Central China	37	2	4	6	7	9	11	15	19	22	26	30	33
华东 Eastern China	55	3	6	8	11	14	17	22	28	33	39	44	50
华南 Southern China	18	1	2	3	4	5	5	7	9	11	13	14	16
西北 Northwest China	11	1	1	2	2	3	3	4	6	7	8	9	10
西南 Southwest China	111	6	11	17	22	28	33	44	56	67	78	89	100
其他 Others	20	1	2	3	4	5	6	8	10	12	14	16	18
总和 Total	473	25	49	71	94	119	142	188	239	285	332	379	425

**1.2.4.2 分组聚类+稀有基因优先取样法** 使用 PowerMarker V3.25 软件利用分子数据分别计算 9 组材料的遗传距离,根据遗传距离分别进行聚类分析并构建系统进化树。根据聚类结果,各组按 12 个取样比例分别对具最远遗传距离的样品进行优先抽取。此外,对每个 InDel 引物扩增出的位点按其出现的频率进行分类,对具有稀有等位基因位点(基因频率 < 5%)的样品进行优先取样。各组按比例的取样数量参照分组随机取样法,取样后将各组按相同比例所取样品进行合并,构成 12 个候选多样性核心样本集。

**1.2.4.3 逐级聚类+稀有基因优先取样法** 使用 PowerMarker V3.25 软件计算所有样品间的遗传距离,根据遗传距离进行聚类分析并构建系统进化树。根据聚类分组结果,按 12 个取样比例进行逐级聚类取样,即从最低分类水平组内(样品间差异最小)的 2 个样品中随机取一个样品或优先选取其中具有稀有基因位点的样品,而另外一个样品被剔除,入选的样品和其他保留下来的样品进入下一轮聚类分析,若组内只有一个样品则直接进入下一轮聚类分析;

**1.2.4.1 分组随机取样法** 将 473 份初级多样性固定材料按品种来源地分为 9 组,分别为国外组、东北组、华北组、华中组、华东组、华南组、西北组、西南组、其他(来源地不明)。按照 12 个取样比例,分别在 9 个组中进行随机取样,具体取样数量见表 3。各组按比例取样后,将按相同比例所取样品进行合并,构成 12 个候选多样性固定核心样本集。

对保留下来的样品再次计算遗传距离并进行聚类分析,按照相同的方法取样,再进入下一轮聚类分析和取样,直到剩余样品量达到相应比例要求的样本量范围,就构成了该比例的候选多样性核心样本集,即 12 个候选核心样本集。

### 1.2.5 不同候选多样性核心样本集的评价与比较

分析 3 种方法构建的 36 个候选核心种质的平均等位基因数、总等位位点数、等位位点保留百分率、有效等位基因数、Shannon's 信息指数、基因多样性、多态性信息含量等参数,对不同候选多样性固定核心样本集进行评价和比较,最终选出最优的候选多样性固定核心样本集。

## 2 结果与分析

### 2.1 多态性 InDel 引物筛选与初选多样性固定群体的遗传多样性分析

经过引物筛选,获得在不同材料间有多态性的 149 对 InDel 引物,这些引物在 473 个自交系中均扩增出具有明显多态性且带型稳定的产物,通过条带统计共检测到 310 个等位位点,每对 InDel

引物在所有材料中扩增出 2~3 个等位位点,平均等位基因数为 2.1074 个。各引物揭示的所有材料的有效等位基因数分布在 1.0043~2.7587 之间,平均为 1.7930;Shannon's 信息指数范围为 0.0154~1.0537,平均值为 0.6182;基因多样性变幅为 0.0043~0.6375,平均为 0.4230;多态性信息含量为 0.0043~0.5629,平均值为 0.3310,说明这些引物能很好地检验黄瓜种质资源的遗传多样性。

## 2.2 多样性固定核心样本集的构建

**2.2.1 分组随机取样法构建核心样本集** 对黄瓜初级遗传多样性固定群体按 5%、10%、15%、20%、25%、30%、40%、50%、60%、70%、80%、90% 的比例分别进行分组随机取样,计算不同比例构成的候选多样性固定核心样本集和初级遗传多样性固定群体(100%)的平均等位基因数、总等位位点数、等位位点保留百分率(表 4),有效等位基因数、Shannon's 指数、基因多样性、PIC 多态性信息含量(图 1)。

表 4 分组随机取样法构建的候选多样性固定核心样本集的参数比较

Table 4 The comparison of the candidate GDFCCs by grouping and randomly sampling with different proportions

取样比例 (%) Proportion	平均等位基因数 $N_a$	总等位位点数 Total loci	等位位点保留百分率 (%) Retention rate of loci	取样比例 (%) Proportion	平均等位基因数 $N_a$	总等位位点数 Total loci	等位位点保留百分率 (%) Retention rate of loci
5	2.0336	301	97.10	50	2.1007	309	99.68
10	2.0738	305	98.39	60	2.1074	310	100.00
15	2.0604	305	98.39	70	2.1007	310	100.00
20	2.0671	306	98.71	80	2.1074	310	100.00
25	2.0738	306	98.71	90	2.1074	310	100.00
30	2.094	308	99.35	100	2.1074	310	100.00
40	2.0872	307	99.03				

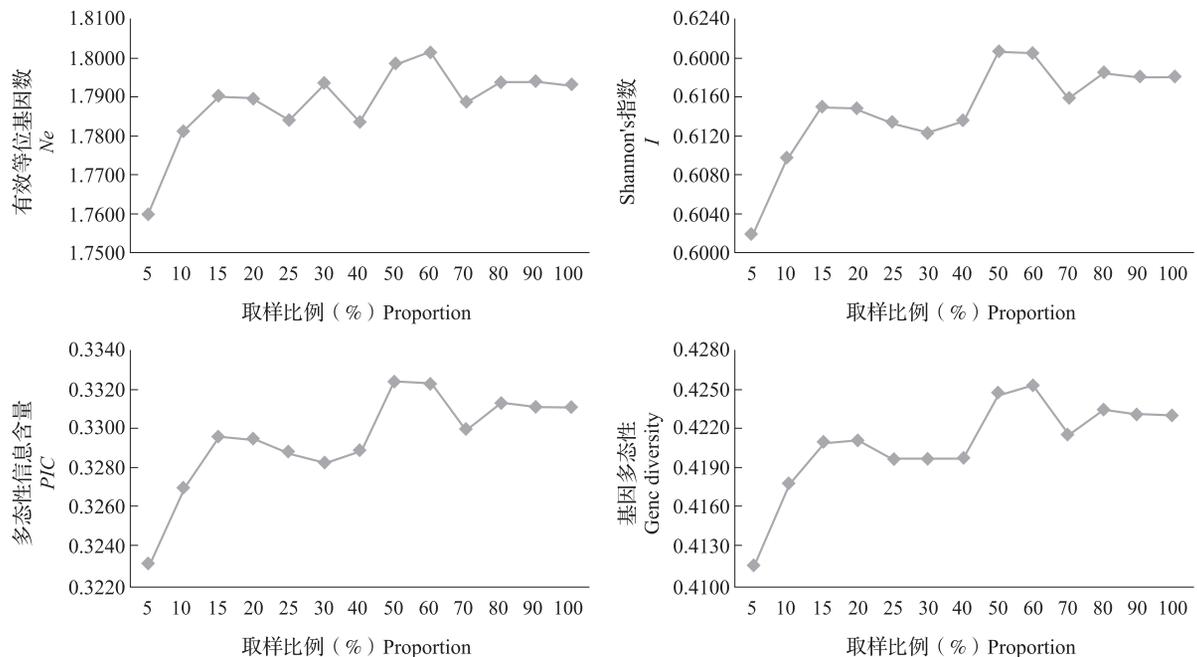


图 1 分组随机取样法构建候选核心样本集的参数变化趋势图

Fig. 1 The comparison of four parameters among the candidate GDFCCs by grouping and randomly sampling with different proportions

如表 4 和图 1 所示,随取样比例的增加,各指标均呈不规则上升趋势,且上述各参数在取样比例为

15% 时出现第 1 次拐点,说明从 5% 升高到 15% 的比例时群体多样性快速增加;当取样比例达到 50%

或 60% 时则达到最高,且高于或接近原始群体。同时,分析显示,从 50% 的抽样比例开始,所获得的候选核心样本集的等位位点保留百分率就已达到 99.68%,说明 50% 比例已经能很好地表现原始群体的多态性;比例过低,等位位点丢失严重;过高,样本集的冗余度提高。综合分析认为 50% 是分组随机取样法构建黄瓜多样性固定核心样本集的最佳取样比例。

**2.2.2 分组聚类 + 稀有基因优选取样法** 按照分组聚类和稀有基因优选的原则,按各比例取样最终获得 12 个候选核心样本集,计算不同比例构成的候选多样性固定核心样本集和初级遗传多样性固定群体(100%)的平均等位基因数、总等位位点数、等位位点保留百分率(表 5),有效等位基因数、Shannon's 指数、基因多样性、*PLC* 多样性信息含量(图 2)。

表 5 分组聚类 + 稀有基因优先取样法构建的候选核心样本集的比较

Table 5 The comparison of the candidate genetic diversity fixed core collections by grouping and clustering plus rare gene preferred sampling with different proportions

取样比例 (%) Proportion	平均等位基因数 <i>N<sub>a</sub></i>	总等位位点数 Total loci	等位位点保留百分率 (%) Retention rate of loci	取样比例 (%) Proportion	平均等位基因数 <i>N<sub>a</sub></i>	总等位位点数 Total loci	等位位点保留百分率 (%) Retention rate of loci
5	2.0403	304	98.06	50	2.1074	310	100.00
10	2.0671	306	98.71	60	2.1074	310	100.00
15	2.0738	307	99.03	70	2.1074	310	100.00
20	2.1007	310	100.00	80	2.1074	310	100.00
25	2.1007	310	100.00	90	2.1074	310	100.00
30	2.1007	310	100.00	100	2.1074	310	100.00
40	2.1074	310	100.00				

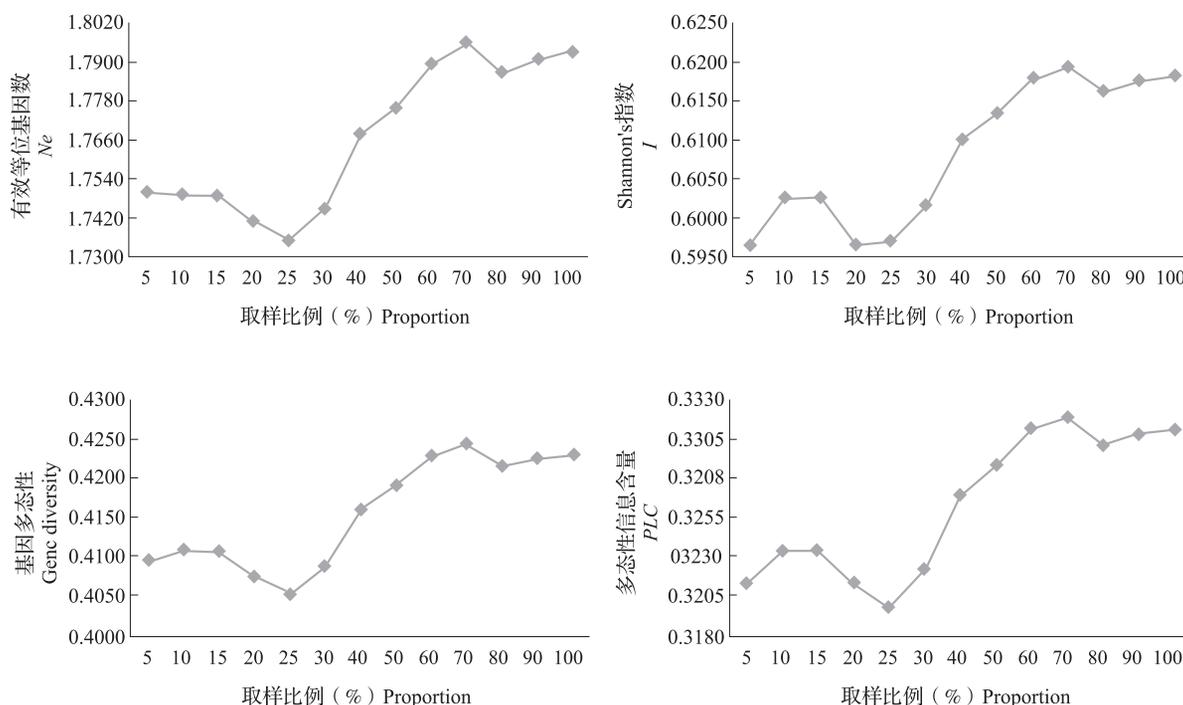


图 2 分组聚类 + 稀有基因优先取样法候选核心样本集的参数变化趋势图

Fig. 2 The four parameters comparisons among the candidate genetic diversity fixed core collections by grouping and clustering plus rare gene preferred sampling with different proportions

如表 5 和图 2 所示,各参数在取样比例达到 40% 开始快速上升,在 70% 时达到最高,之后趋缓。在 5% ~ 15% 的区段,有效等位基因数和基因多样性指数曲线是较低且平缓的,而 Shannon's 指数和多态性信息指数出现一个小峰值,说明从 5% 比例升高到 10% 比例时群体多样性在增加。从等位位点保留百分率情况看,20% 以上的抽样比例便能使其保留 100%,说明在抽样比例为 25% 以上时群体内出现严重冗余。综合多个参数的结果认为,抽样比例为 60% 时,不仅候选核心样本集的等位位点保留比例高、冗余程度低,而且多样性高于或接近原始

群体,是分组聚类 + 稀有基因优先取样法构建候选黄瓜核心样本集的较佳取样比例。

**2.2.3 逐级聚类 + 稀有基因优先取样法** 通过逐级聚类分析和取样,分别得到按 90% ~ 5% 的 12 个比例抽样的候选多样性固定核心样本集 426 份、378 份、330 份、282 份、237 份、189 份、141 份、116 份、95 份、71 份、47 份和 24 份。计算不同比例构成的候选多样性固定核心样本集和初级遗传多样性固定群体(100%)的平均等位基因数、总等位位点数、等位位点保留百分率(表 6),有效等位基因数、Shannon's 指数、基因多样性、PIC 多态性信息含量(图 3)。

表 6 逐级聚类 + 稀有基因优先取样法构建的候选核心样本集的比较

Table 6 The comparison of the candidate genetic diversity fixed core collections by clustering stepwise and rare gene preferred sampling with different proportions

取样比例 (%)	平均等位基因数	总等位位点数	等位位点保留百分率 (%)	取样比例 (%)	平均等位基因数	总等位位点数	等位位点保留百分率 (%)
Proportion	$N_a$	Total loci	Retention rate of loci	Proportion	$N_a$	Total loci	Retention rate of loci
5	2.0805	306	98.71	50	2.1074	310	100.00
10	2.1007	309	99.68	60	2.1074	310	100.00
15	2.1007	309	99.68	70	2.1074	310	100.00
20	2.1007	309	99.68	80	2.1074	310	100.00
25	2.1074	310	100.00	90	2.1074	310	100.00
30	2.1074	310	100.00	100	2.1074	310	100.00
40	2.1074	310	100.00				

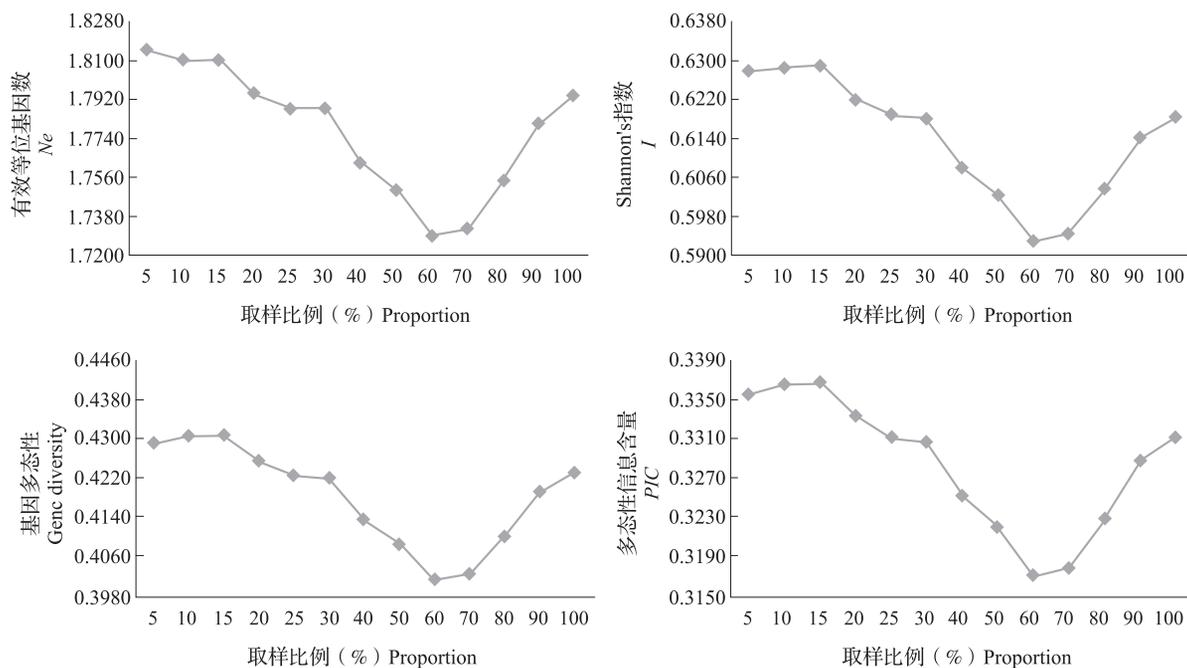


图 3 逐级聚类 + 稀有基因优先取样法候选核心样本集的参数变化趋势图

Fig. 3 The parameters comparisons of the candidate genetic diversity fixed core collections by clustering stepwise and rare gene preferred sampling with different proportions

如表 6 和图 3 所示,随取样比例的增加呈先逐渐下降后又逐渐上升的变化趋势。当取样比例在 5%~15% 时,各参数值处于高位。在取样比例为 30% 时,各参数的下降速度出现拐点,且与 90% 取样比例的各项参数接近。当取样比例为 60%~70% 时,各指标处于低谷。同时,抽样比例在 15% 时,候选核心样本集的等位位点保留百分率达到 99.68%,在 20% 比例及以上时,等位位点保留百分率达到 100%,说明从 25% 比例开始群体内出现冗余。综合比较各参数,认为 15% 是逐级聚类+稀有基因优先取样法构建候选黄瓜核心样本集的最佳取样比例,它既能很好地保留等位基因位点,同时也能较好地保持原有群体的遗传多样性。

表 7 对候选多样性固定核心样本集的多样性和代表性评价

Table 7 The diversity and representativeness of the candidate genetic diversity fixed core collections

候选核心 样本集 Code	平均等位 基因数 <i>N<sub>a</sub></i>	总等位 位点数 Total loci	等位位点保留 百分率(%) Retention rate of loci	平均有效等位 基因数 <i>N<sub>e</sub></i>	Shannon's 指数 <i>I</i>	基因多态性 Gene diversity	多态性信息 指数 <i>PIC</i>
Core 1	2.1007	309	99.68	1.7984	0.6208	0.4246	0.3324
Core 2	2.1074	310	100.00	1.7892	0.6178	0.4228	0.3311
Core 3	2.1007	309	99.68	1.8097	0.6287	0.4306	0.3365
原始群体	2.1074	310	100.00	1.7930	0.6182	0.4230	0.3311

从表 7 中可以看出,3 个核心样本集与原始群体的各项参数经 t 检验表现为不显著,即 3 个候选样本集均可以视为有效的核心样本集。比较而言,逐级聚类+稀有基因优先取样法构建的 Core 3 的平均有效等位基因数、Shannon's 指数、基因多态性和 *PIC* 值均高于或接近原始群体,同时保留了 99.68% 的原始群体的等位位点,表现了较好的去冗余的作用。最终确定以逐级聚类+稀有基因优先取样法的 15% 比例构建的候选核心样本集 Core 3 作为黄瓜多样性固定核心样本集。

构建的多样性固定核心样本集包括 71 份种质(表 8),这些材料分别来自中国东北、华北、华中、华东、华南、西北、西南等 7 个地区的 18 个省市,其中,来自黑龙江 1 份、吉林 4 份、辽宁 1 份、河北 2 份、内蒙古 1 份、河南 2 份、湖北 6 份、湖南 1 份、安徽 1 份、江苏 1 份、山东 2 份、广东 3 份、台湾 1 份、宁夏 1 份、新疆 1 份、广西 1 份、四川 3 份、云南 6 份。国外资源分别来自 15 个国家,其中来自捷克 1 份、法国 1 份、伊朗 1 份、荷兰 1 份、波兰 1 份、波多黎各 1 份、西班牙 1 份、赞比亚 1 份、津巴布韦 1 份、俄罗斯 3

### 2.3 候选多样性固定核心样本集的评价

确定了 3 种取样方法的最佳取样比例,分别构建 3 个最优候选多样性固定核心样本集,即按照分组随机取样法的 50% 比例构建的候选核心样本集被标记为 Core 1,按照分组聚类+稀有基因优先取样法的 60% 比例构建的候选核心样本集被标记为 Core 2,按照逐级聚类+稀有基因优先取样法的 15% 比例构建的候选核心样本集被标记为 Core 3。比较 3 个候选核心样本集的平均等位基因数、总等位位点数、等位位点保留百分率、平均有效等位基因数、Shannon's 指数、基因多样性、多态性信息指数(表 7)。

份、美国 4 份、日本 3 份、叙利亚 1 份、以色列 1 份、印度 7 份。另外还有 5 份材料来源未知。可见,多样性固定核心样本集的材料来源地覆盖面广,具有丰富的多样性和代表性。

## 3 讨论

不同作物由于授粉特性不同而导致同一种质内的杂合性不同,自花授粉种质杂合度较低而异花授粉种质的异质性较高。对于异花授粉的黄瓜来讲,纯合度低必将影响种质资源研究利用的效度,所以本研究利用黄瓜表型核心种质的自交系构建多样性固定核心样本集不失为提高异化授粉作物种质资源研究和利用效率的一种较好的选择。

核心种质是以最少的种质资源样品来最大程度地代表一个物种种质资源的遗传多样性<sup>[28-29]</sup>。核心种质的构建方法很多<sup>[30-31]</sup>,王丽侠等<sup>[32]</sup>认为类内随机取样法构建的核心种质表现最佳;姜慧芳等<sup>[33]</sup>采用分层聚类以及随机取样结合必选资源的方法,有效构建了花生核心种质;李国强<sup>[34]</sup>按照大白菜分类系统和结球大白菜生态型采用分组聚类法构建出

表 8 黄瓜多样性固定核心样本集 Core 3 中 71 份种质材料来源情况

Table 8 The source of 71 cucumber germplasms in the GDFCC of Core 3

种质编号 ID	原产地 Origin	分区 Area	种质编号 ID	原产地 Origin	分区 Area	种质编号 ID	原产地 Origin	分区 Area
QT329	捷克	国外	QT488	印度	国外	QT165	中国山东	华东
QT434	法国	国外	QT7	印度	国外	QT174	中国山东	华东
QT369	伊朗	国外	QT518	印度	国外	QT88	中国广东	华南
QT435	荷兰	国外	QT448	印度	国外	QT43	中国广东	华南
QT338	波兰	国外	QT185	中国黑龙江	东北	QT508	中国广东	华南
QT514	波多黎各	国外	QT120	中国吉林	东北	QT24	中国台湾	华南
QT1	西班牙	国外	QT114	中国吉林	东北	QT228	中国宁夏	西北
QT414	赞比亚	国外	QT39	中国吉林	东北	QT146	中国新疆	西北
QT412	津巴布韦	国外	QT191	中国吉林	东北	QT194	中国广西	西南
QT349	俄罗斯	国外	QT250	中国辽宁	东北	QT260	中国四川	西南
QT20	俄罗斯	国外	QT56	中国河北	华北	QT47	中国四川	西南
QT258	俄罗斯	国外	QT54	中国河北	华北	JD7	中国四川	西南
QT68	美国	国外	QT180	中国内蒙古	华北	BN34	中国云南	西南
QT9	美国	国外	QT155	中国河南	华中	QT244	中国云南	西南
QT10	美国	国外	QT121	中国河南	华中	QT221	中国云南	西南
QT13	美国	国外	QT163	中国湖北	华中	BN49	中国云南	西南
QT503	日本	国外	QT164	中国湖北	华中	BN1	中国云南	西南
QT403	日本	国外	QT206	中国湖北	华中	BN7	中国云南	西南
QT319	日本	国外	QT65	中国湖北	华中	QT301	高代自交系	未知
QT21	叙利亚	国外	QT84	中国湖北	华中	QT249	高代自交系	未知
QT25	以色列	国外	QT86	中国湖南	华中	QT281	未知	未知
QT332	印度	国外	QT265	中国湖北	华中	QT246	高代自交系	未知
QT8	印度	国外	QT202	中国安徽	华东	JD43	水果黄瓜	未知
QT312	印度	国外	QT60	中国江苏	华东			

284 份大白菜核心种质;郝晨阳等<sup>[35]</sup>采用适当调整的分层分组代表性取样法构建我国普通小麦核心种质,对著名品种、重要育种亲本和携带稀有等位变异基因等材料优先入选,即对遗传多样性高的地区增加取样量,反之减少;吴子龙等<sup>[36]</sup>和崔娜<sup>[37]</sup>采用有分子标记结合逐步压缩聚类法分别构建了山葡萄核心种质和萝卜核心种质。本研究采用 3 种取样方法构建候选核心样本集,最终认为逐级聚类 + 稀有基因优先取样法按 15% 比例取样构建的黄瓜多样性固定核心样本集最优,逐级聚类 + 稀有基因优先取样法在尽量保留具有稀有基因样品的同时,在整体上利用种质亲缘关系远近来去除冗余是行之有效的。

本研究构建的核心种质与吕婧<sup>[14]</sup>采用 M 策略从 3318 份原始黄瓜种质资源中筛选出 120 份黄瓜

核心种质相比较的最大特点在于:InDel 标记的应用主要反映的是黄瓜基因组插入和缺失变异信息;试验材料并非是随意获取的杂合度高低不一的原始种质,而是通过对种质库保存的资源的表型鉴定构建初级核心种质,再经过多代纯化形成遗传多样性丰富和遗传稳定的自交系;分子微核心种质对表型核心种质的代表性达到 99% 以上,更利于核心种质在科研和育种中的直接利用。

构建核心种质所用的数据也有很多类型,如表型性状数据(如质量性状<sup>[34,38-39]</sup>或数量性状<sup>[34,39-40]</sup>),分子数据(AFLP 引物<sup>[41-43]</sup>,SSR 引物<sup>[14,32,35,44-45]</sup>,EST-SSR 引物<sup>[37,46]</sup>,RAPD 引物<sup>[47-48]</sup>,InDel 引物<sup>[15]</sup>) 和表型数据结合分子数据<sup>[49]</sup>等,不同类型的数据会从不同的方面反映出种质资源的背景、特征和遗传多样性。本研究利

用 InDel 分子标记数据结合种质材料来源地对黄瓜核心种质自交系进行分析并构建多样性固定核心样本集,其所反映的遗传信息不同于吕婧<sup>[14]</sup>用 SSR 构建的核心种质。本研究中由单一标记获得的较少的数据类型可能会影响多样性固定核心样本集特征的全面反映,将来有必要综合利用不同类型的分子和表型鉴定数据对多样性固定核心样本集进行完善。

### 参考文献

- [1] 钟金仙,罗英. 黄瓜品种主要农艺性状相关与主成分分析[J]. 中国农学通报,2012,28(4):131-134
- [2] 李强. 黄瓜表皮毛相关基因的定位、同源克隆与功能研究[D]. 泰安:山东农业大学,2013
- [3] 刘书林,顾兴芳,苗晗,等. 黄瓜黑色果刺基因染色体定位及候选基因分析[J]. 中国农业科学,2014,47(1):122-132
- [4] 刘书林,顾兴芳,苗晗,等. 成熟黄瓜果皮红色性状的遗传分析及其基因定位[J]. 园艺学报,2014,41(2):259-267
- [5] 申晓青,陈书霞,潘玉朋,等. 黄瓜嫩果果皮颜色的遗传研究[J]. 农业生物技术学报,2014,22(1):37-46
- [6] 李锡香. 黄瓜种质遗传多样性的形态和分子评价及其亲缘关系研究[D]. 北京:中国农业科学院,2002
- [7] 王志坤. 黄瓜种质资源遗传多样性研究[D]. 哈尔滨:东北农业大学,2004
- [8] 司旻星,关媛,潘俊松,等. 黄瓜(*Cucumis sativus* L.)种质资源遗传多样性及亲缘关系分析[J]. 上海交通大学学报:农业科学版,2007,25(2):130-137
- [9] 沈颖. 西双版纳黄瓜群体遗传多样性分析及黄瓜果肉色 QTL 定位研究[D]. 北京:中国农业科学院,2009
- [10] 齐建建. 黄瓜遗传多样性和人工驯化的分子基础[D]. 北京:中国农业科学院,2014
- [11] 程嘉琪. 黄瓜核心种质的遗传多样性分析及白粉病抗性和低温耐受性评价[D]. 北京:中国农业科学院,2011
- [12] 程嘉琪,沈颖,李锡香,等. 黄瓜核心种质对白粉病的田间抗性评价[J]. 中国蔬菜,2011(20):15-19
- [13] 程嘉琪,沈颖,李锡香,等. 黄瓜核心种质低温耐受性的田间评价[J]. 植物遗传资源学报,2012,13(4):660-665
- [14] 吕婧. 黄瓜种质资源群体结构分析与核心种质集筛选[D]. 北京:中国农业科学院,2011
- [15] 李斯更. 黄瓜性型相关基因关联分析及性型稳定性的 DNA 甲基化研究[D]. 北京:中国农业科学院,2012
- [16] King G J, Teakle G R, Allender C J, et al. From trait to genome: characterizing *Brassica* diversity[C]//Plant & Animal Genomes XII Conference, 2004:1700-1702
- [17] 李斯更. 基于黄瓜基因组重测序的 InDel 标记开发及其应用[J]. 植物遗传资源学报,2012,14(2):278-283
- [18] 冯芳君,罗利军,李茨,等. 水稻 InDel 和 SSR 标记多态性的比较分析[J]. 分子植物育种,2005,3(5):725-730
- [19] 王友林,张礼霞,勾晓霞,等. 利用 InDel 标记鉴定水稻育种材料的籼粳属性[J]. 核农学报,2013,27(7):913-921
- [20] 葛敏,蒋璐,张晓林,等. 利用 Insertion/Deletion (InDel) 分子标记检测玉米互交种混杂的原理及应用[J]. 分子植物育种,2013,11(1):37-47
- [21] 仪泽会. 白菜 SSR 和 InDel 标记的开发及叶缘裂刻突变体的基因定位[D]. 陕西:西北农林科技大学,2012
- [22] 朱东旭,王彦华,赵建军,等. 结球甘蓝相对于大白菜连锁群特异 InDel 标记的建立及应用[J]. 园艺学报,2014,41(8):1699-1706
- [23] 王万兴. 结球甘蓝高密度遗传连锁图谱构建与主要农艺性状的 QTL 定位[D]. 北京:中国农业科学院,2013
- [24] 张圣平,苗晗,程周超. 黄瓜果实苦味(*Bt*)基因的插入缺失(InDel)标记[J]. 农业生物技术学报,2011,19(4):649-653
- [25] 兰青阔,张桂华,王永. 基于 InDel 标记快速检测黄瓜津优 38 种子纯度[J]. 种子,2011,30(6):19-23
- [26] 张海英. 黄瓜重要抗病基因的分子标记及遗传图谱的构建[D]. 北京:中国农业科学院,2006
- [27] 简德明. 黄瓜白粉病抗性基因紧密连锁的 AFLP 分子标记研究[D]. 北京:首都师范大学,2007
- [28] Frankel O H, Brown A H D. Plant genetic resources today: a critical appraisal[M]//Holden J H W, Williams J T. Crop genetic resources: conservation & evaluation. London: George Allen & Unwin, 1984:249-257
- [29] Brown A H D. Core collections: a practical approach to genetic resources management[J]. Genome, 1989,31:818-824
- [30] 刘宁宁. 植物资源核心种质构建与评价新方法的研究[D]. 杭州:浙江大学,2007
- [31] 张龙进. 山茱萸种质资源遗传多样性分析及核心种质构建方法研究[D]. 西安:陕西师范大学,2007
- [32] 王丽侠,李英慧,李伟,等. 长江春大豆核心种质构建及分析[J]. 生物多样性,2004,12(6):578-585
- [33] 姜慧芳,任小平,廖伯寿,等. 中国花生核心种质的建立[J]. 武汉植物学研究,2007,25(3):289-293
- [34] 李国强. 大白菜核心种质的构建与评价[D]. 北京:中国农业科学院,2008
- [35] 郝晨阳,董玉琛,王兰芬,等. 我国普通小麦核心种质的构建及遗传多样性分析[J]. 科学通报,2008,53(8):908-915
- [36] 吴子龙,王军,沈育杰,等. 不同分组取样方法初步构建山葡萄核心种质的研究[J]. 北方园艺,2011(7):117-122
- [37] 崔娜. 基于萝卜 EST 序列的 SSR 标记开发与应用研究[D]. 北京:中国农业科学院,2012
- [38] 李保印. 中原牡丹品种遗传多样性与核心种质构建研究[D]. 北京:北京林业大学,2007
- [39] 陈建华. 苎麻资源核心种质构建方法及遗传多样性研究[D]. 北京:中国农业科学院,2011
- [40] 刘遵春,张春雨,张艳敏,等. 利用数量性状构建新疆野苹果核心种质的方法[J]. 中国农业科学,2010,43(2):358-370
- [41] 杨美,付杰,向巧彦,等. 利用 AFLP 分子标记技术构建花莲核心种质资源[J]. 中国农业科学,2011,44(15):3193-3205
- [42] 张维瑞,袁王俊,尚富德. 基于 AFLP 分子标记的桂花品种核心种质的构建[J]. 西北植物学报,2012,32(7):1349-1354
- [43] 王红霞,赵书岗,高仪,等. 基于 AFLP 分子标记的核桃核心种质的构建[J]. 中国农业科学,2013,46(23):4985-4995
- [44] 张春雨,陈学森,张艳敏,等. 采用分子标记构建新疆野苹果核心种质的方法[J]. 中国农业科学,2009,42(2):597-604
- [45] 温景辉. 基于 SSR 分子标记的山葡萄种质遗传多样性研究与核心种质构建[D]. 长春:吉林农业大学,2011
- [46] 李丽,何伟明,马连平,等. 用 EST-SSR 分子标记技术构建大白菜核心种质及其指纹图谱[J]. 基因组学与应用生物学,2009,28(1):76-88
- [47] 田彬彬. 寒地苹果种质资源 RAPD 分子标记及其核心种质初步构建[D]. 长春:吉林农业大学,2007
- [48] 王庆芬. 抗寒梨种质资源 RAPD 的分子标记及其核心种质初步构建[D]. 长春:吉林农业大学,2007
- [49] 刘遵春,刘大亮,崔美,等. 整合农艺性状和分子标记数据构建新疆野苹果核心种质[J]. 园艺学报,2012,39(6):1045-1054