

基于简化基因组测序的稗草遗传多样性研究

聂豪杰^{1,2,3}, 李林傲^{2,3,4}, 张杨^{1,2,3}, 荆庆芳^{1,2,3}, 朱林^{1,2,3*}

(1. 宁夏大学生态环境学院, 宁夏 银川 750021; 2. 宁夏大学西北土地退化与生态恢复国家重点实验室培育基地, 宁夏 银川 750021; 3. 宁夏大学西北退化生态系统恢复与重建教育部重点实验室, 宁夏 银川 750021; 4. 宁夏大学农学院, 宁夏 银川 750021)

摘要:【目的】利用SNP分子标记,分析稗草(*Echinochloa* spp.)种质资源(野生材料和栽培品种)遗传多样性水平,揭示不同类型稗草种质资源之间的亲缘关系以及群体遗传结构差异,为稗草的遗传多样性保护、野生资源的开发利用奠定基础。【方法】以20份野生稗草和24份栽培稗草为材料,采用RAD-seq技术对各样本进行测序并开发SNP标记,通过构建系统进化树并进行主成分、群体遗传结构和遗传多样性参数分析,探讨遗传结构差异并评估不同群体间遗传多样性水平。【结果】测序产生的序列数据量为161.51 Gb,过滤后共获得31 284个高质量SNP位点。遗传结构分析结果将44份稗草资源划分为4个类群,类群I和II主要为野生资源,类群III和IV主要为栽培资源。4个类群间的遗传多样性指数(P_i)为0.000 061~0.000 154,遗传分化指数(F_{ST})为0.289 2~0.507 8,观测杂合度(H_o)为0.100 7~0.226 4,期望杂合度(H_e)为0.140 4~0.251 2,其中野生类群I和II的遗传多样性指数(P_i)、观测杂合度(H_o)、期望杂合度(H_e)均高于栽培类群III和IV;遗传分化结果表明4个类群两两之间均发生了显著的遗传分化,野生类群II与栽培类群IV遗传分化最为显著。【结论】SNP标记能将野生稗草和栽培稗草区分开来,证实了在稗草分类方面的可行性;野生稗草群体的遗传多样性水平高于栽培稗草群体,其中野生类群II(稗、水田稗、西来稗)遗传多样性水平最高,但总体上稗草遗传多样性水平偏低;并推测遗传漂变、人类活动以及授粉方式可能是导致稗草植物遗传多样性水平下降和产生遗传分化的重要原因。

关键词: 稗属; RAD; SNP; 遗传结构; 遗传多样性

中图分类号: S543 文献标志码: A 文章编号: 1009-5500(2024)03-0052-10

DOI: 10.13817/j.cnki.cyycp.2024.03.007



稗属(*Echinochloa*)为一年生或多年生禾本科(Graminea) C_4 植物,广泛分布于世界热带和温度地区^[1]。稗草具有很高的食用和饲用价值,加强对稗草资源的搜集和系统分类,并准确评估在形态以及分子

水平上的遗传多样性,有利于提高稗草资源的基因挖掘和种质创新效率,促进遗传育种工作的研究。早期对稗属植物遗传多样性的研究主要集中在形态特征,但由于稗草生存范围广泛,不同生境下的稗草形态性状如植株高度、穗色以及芒长等的高度变异导致对稗草的分类鉴定存在很大争议^[2]。分子标记作为一种直接分析核苷酸序列差异的方法,具有稳定性强、特异性好、标记数量多等特点,较为广泛地应用在稗属植物遗传多样性研究方面。Altop等^[3]通过RAPD标记并结合形态学标记对土耳其不同地理来源的稗草植物的遗传多样性进行了研究,说明了不同地理来源的稗草在基因型上存在很大差别,稗属植物种内变异丰

收稿日期: 2023-03-22; 修回日期: 2023-04-06

基金资助: 宁夏自然科学基金项目“基于简化基因组测序技术的湖南稷子与其近缘种遗传关系分析及分子标记开发”(2022AAC03086); 宁夏回族自治区农业育种专项(2019NYYZ0401)

作者简介: 聂豪杰(1998-), 湖北天门, 硕士研究生。

E-mail: 1691463793@qq.com

*通信作者。 E-mail: zhulinscience@126.com

富。Kaya等^[4]、Danquah等^[5]利用 AFLP 和 SSR 标记鉴定 24 份稗草,在稗属植物种内和种间都检测到了高水平的遗传变异。刘德好等^[6]利用 ISSR-PCR 反应体系筛选出了 9 对 SSR 多态性引物对 53 份稗属样本进行多样性分析,结果表明稗属植物种间多态性极其丰富。张婉茹^[7]利用 ISSR 分子标记技术对 8 种寒地稗草进行了研究,结果表明这些寒地稗草种质间的遗传变异较大。近年来,以第二代测序为基础发展起来的简化基因组测序技术能够快速有效地实现大规模的分子标记开发^[8]。在蓝花楹 (*Jacaranda mimosifolia*)^[9]、披碱草属 (*Elymus*)^[10]、鹅掌楸属 (*Liriodendron*)^[11]等物种的 SNP 标记开发、群体遗传结构和遗传多样性、系统演化分析等研究中得到广泛应用。

野生种质资源由于具有变异类型繁多的特点,相比于栽培种质资源遗传多样性也更高,野生资源为物

种的栽培驯化和育种提供了基础^[12]。本研究利用 RAD-seq 技术对国内稗草群体进行分子标记开发,通过获得的全基因组范围内的 SNP 标记,对国内野生稗草和栽培稗草进行遗传多样性和遗传结构分析,揭示野生稗草和栽培稗草间的遗传关系,从而为进一步开展稗草植物分子鉴定、种质资源遗传多样性保护、野生资源开发利用以及引种驯化奠定基础。

1 材料和方法

1.1 材料

所用的 44 份稗草资源包括 20 份野生稗草和 24 份栽培稗草,其中栽培稗草包含了宁夏本地的 2 个旱稗品系和 4 个湖南稷子品系,栽培稗草均由近几年课题组收集和引进;野生稗草除长芒稗和西来稗来源于宁夏试验地伴生杂草外,其余均为国家种子库提供(表 1)。

表 1 稗草材料来源

Table 1 Sources of *E. spp.* materials

编号	群体名称	类型	产地来源
A-1、A-2、A-3	旱稗 <i>E. hispidula</i> (Retz.) Nees(品系)	栽培	宁夏平罗
B-1、B-2、B-3	湖南稷子 <i>E. frumentacea</i> (Roxb.) Link(品系)	栽培	宁夏平罗
C-1、C-2、C-3	旱稗 <i>E. hispidula</i> (Retz.) Nees(品系)	栽培	宁夏平罗
D-1、D-2、D-3	湖南稷子 <i>E. frumentacea</i> (Roxb.) Link(品系)	栽培	宁夏平罗
E-1、E-2、E-3	宁夏无芒稗 <i>E. crus-galli</i> var. <i>mitis</i> (Pursh) Petermann	栽培	宁夏平罗
F-1、F-2、F-3	湖南稷子 <i>E. frumentacea</i> (Roxb.) Link(品系)	栽培	宁夏平罗
G-1、G-2、G-3	湖南稷子 <i>E. frumentacea</i> (Roxb.) Link(品系)	栽培	宁夏平罗
H-1、H-2、H-3	朝牧一号稗子 <i>E. crusgalli</i> (L.) Beauv 'Chaomu 1'	栽培	辽宁朝阳
I-1、I-2、I-3	长芒稗 <i>E. caudata</i> Roshev.	野生	宁夏平罗
J-1、J-2、J-3	西来稗 <i>E. crus-galli</i> var. <i>zelayensis</i> (Kunth) Hitchcock	野生	宁夏平罗
K-1、K-2、K-3	稗 <i>E. crus-galli</i> (L.) P. Beauv.	野生	吉林白山
L-1、L-2、L-3	湖南稷子 <i>E. frumentacea</i> (Roxb.) Link	野生	吉林长岭
M-1、M-2、M-3	短芒稗 <i>E. crus-galli</i> var. <i>breviseta</i> (Doll) Podpera	野生	四川西昌
N-1、N-2、N-3	无芒稗 <i>E. crus-galli</i> var. <i>mitis</i> (Pursh) Petermann	野生	甘肃白银
O-1、O-2	水田稗 <i>E. oryzoides</i> (Ard.) Flritsch	野生	云南昆明

注:A-O为群体编号,以下所示图表相同。

1.2 基因组 DNA 的提取

每份稗草材料选取 4~6 粒种子置于光周期(昼/夜)14 h/8 h,光照强度 4 000 lx/0 lx,温度 28 °C/16 °C,湿度 43% 的光照恒温培养箱中培养,待幼苗生长至 3~4 片叶时,剪取健康幼叶适量装入 5 mL 离心管中,放入液氮中速冻至少 1 h 后,保存在 -80 °C 冰箱备用。再取鲜叶 0.5 g,采用改良后的 CTAB 法^[13]提取总

DNA,并利用琼脂糖凝胶电泳技术对基因组 DNA 片段大小进行确定、浓度以及纯度进行检测,对质量合格的 DNA 进行下一步上机测序。

1.3 RAD 文库构建

利用限制性内切酶 *EcoR*-I 对基因组进行酶切,应用超声波将每个样本进行打碎,选择序列长度为 300~700 bp 的片段,插入文库。利用 Illumina Nova-

seq 6000 测序仪平台进行双末端 150 bp 测序,所有文库的构建及上机测序均由广州基迪奥生物科技有限公司完成^[14]。

1.4 测序 reads 过滤

将原始数据过滤,剔除带接头(adapter)的 reads pair,当单端 read 中含有的 N 含量超过该条 read 长度比例的 10% 时,需过滤 paired reads,单端 read 中含有的低质量(质量值 $Q \leq 5$)碱基数超过该条 read 长度比例的 50% 时,需去除此对 paired reads。最终利用得到的 Clean reads 作后续分析。

1.5 比对和变异检测

利用 BWA (<https://software.broadinstitute.org/gatk/>) 软件 men 模块将过滤后得到的 clean reads 比对稗(*E. crus-galli*V2.0)参考基因组序列,参数如下: MEN-t=4, -k=32, -M 根据对比结果,利用 GATK^[15] (<https://software.broadinstitute.org/gatk>) 软件开发 SNP 位点,使用 Vcftools (<http://vcftools.sourceforge.net/>) 软件对 SNP 进行过滤,参数为 miss 小于或等于 0.15、maf 大于或等于 0.05。

1.6 遗传数据分析

使用 Treebest^[16] 软件计算遗传距离矩阵,通过邻接法(neighbor-joining method)构建系统进化树。利用 ADMIXTURE^[17] (<http://software.genetics.ucla.edu/admixture/download.html>) 软件分析群体遗传结构,对研究群体预先设定亚群数目(K 值)为 2~9 进行聚类,根据交叉验证错误率的最低值确定最佳分群数。利用 GCTA^[18] (https://cns.genomics.com/software/gcta/bin/gcta_1.93.2beta.zip) 软件分析群体主成分(principle component analysis, PCA),得到样本的聚类情况。利用 Vcftools 软件计算野生稗草和栽培稗草群体间的观测杂合度(observed heterozygosity, H_o)、期望杂合度(expected heterozygosity, H_e)和遗传多样性指数(nucleotide diversity index, P_i)。使用 Stacks 软件中的 populations 模块计算群体间遗传分化系数(Genetic Differentiation Index, F_{ST}),根据公式 $N_m = (1 - F_{ST}) / 4F_{ST}$ 计算基因流大小。

2 结果与分析

2.1 测序质量

对 44 份稗属种质进行 RAD 简化基因组测序,共

获得数据量 165.27 Gb,平均每个个体 3.76 Gb。对测序数据严格过滤后,得到高质量的 clean data 为 161.51 Gb,平均每个个体 3.67 Gb,平均 GC 含量为 43.72%,平均碱基质量 Q30 达到 94.10%,本研究中所有样本的 Q30 值均在 93% 以上,说明碱基错误率低,测序数据合格(表 2);每份样本均获得高质量序列 25 973 373 条,平均有 25 699 374 条序列可以比对到参考基因组上,平均比率为 98.59%,平均测序深度为 14.23 \times 。平均测序深度和平均比率均满足分析要求。低质量数据过滤后共获得 31 284 个高质量 SNP 位点用于后续分析。以上结果表明本实验测序结果可靠。

2.2 SNP 突变类型

SNP 分型结果表明(图 1),44 份样本颠换变异程度占 40.68%,且 G/T 转换类型最多为 231 882 个,占所有碱基突变类型的 10.51%,A/C 颠换类型(231 576 个)占 10.50%;而转换变异程度高达 59.32%,其中 C/T 转换类型最多为 398 299 个,占所有碱基突变类型的 18.06%,A/G 转换类型(397 764 个)占 18.04%;转换与颠换(Ts/Tv)之比为 1.458。

2.3 系统发育分析

基于开发的 31 284 个 SNP 位点对 44 份野生与栽培稗草构建系统发育分析树(图 2),44 份稗草样本被划分为 4 个类群,分别为 2 个野生类群(类群 I 和类群 II)和 2 个栽培类群(类群 III 和类群 IV)。类群 I 包括 3 份栽培样本 H-1、H-2、H-3(辽宁朝阳)和 9 份野生样本 N-1、N-2、N-3(甘肃白银),I-1、I-2、I-3(宁夏平罗),L-1、L-2、L-3(吉林长岭);类群 II 包括 K-1、K-2、K-3(吉林白山),J-1、J-2、J-3(宁夏平罗)和 O-1、O-2(云南昆明)8 份野生个体;类群 III 包括来自宁夏平罗的 9 份栽培样本 A-1、A-2、A-3、F-1、F-2、F-3、G-1、G-2、G-3 和 3 份四川西昌的野生样本 M-1、M-2、M-3;类群 IV 的所有样本均来自于宁夏平罗的栽培种 B-1、B-2、B-3、C-1、C-2、C-3、D-1、D-2、D-3、E-1、E-2、E-3。聚类分析结果从整体来看,各类群在系统树上基本能够彼此分开。其中栽培类群 III 和 IV 距离相近,遗传关系近;野生类群 I 和 II 遗传关系相对较近。野生类群 II 和栽培类群 III 距离最远,遗传关系远。野生样本 M-1、M-2、M-3 样本完全归于栽培类群 III,

表2 稗草群体测序数据统计

Table 2 Summary of sequencing data of *E. spp.* populations

编号	有效数据/Gb	Q20/%	Q30/%	GC/%	有效片段	可对比片段	比对效率/%
A-1	3.41	98.11	93.83	44.38	24 548 212	23 916 387	98.90
A-2	3.69	98.08	93.75	44.02	26 522 732	25 880 286	99.08
A-3	3.50	98.12	93.87	43.76	25 233 462	24 422 015	98.33
B-1	3.58	98.20	94.03	43.41	25 703 432	25 130 473	99.24
B-2	3.38	97.92	93.33	43.76	24 277 030	23 631 114	98.85
B-3	3.55	98.11	93.83	43.73	25 534 974	24 979 917	99.20
C-1	3.28	98.10	93.79	43.29	23 526 014	23 025 275	99.19
C-2	3.61	98.20	94.02	43.82	25 865 868	25 315 131	99.11
C-3	3.27	98.29	94.23	43.49	23 448 538	22 957 504	99.20
D-1	3.23	98.08	93.70	43.51	23169448	22 567 546	98.87
D-2	3.45	98.15	93.92	43.72	24 7005 74	23 907 918	98.13
D-3	3.70	98.33	94.38	43.55	26 595 614	25 853 506	98.44
E-1	3.31	98.02	93.58	43.78	23 709 386	23 005 249	98.43
E-2	3.30	98.12	93.81	43.49	23 674 330	23 124 576	99.02
E-3	3.93	98.16	93.95	44.15	28 187 672	27 182 242	97.76
F-1	3.76	98.23	94.12	43.88	26 993 352	26 432 164	99.25
F-2	2.87	98.23	94.11	43.32	20 543 256	20 109 224	99.00
F-3	2.84	98.15	93.90	43.53	20 378 398	19 954 017	99.28
G-1	2.96	98.21	94.06	43.86	21 204 710	20 491 135	97.86
G-2	2.94	98.20	93.99	44.09	21 024 826	20 618 632	99.30
G-3	4.11	98.29	94.29	44.15	29 533 790	28 892 544	99.17
H-1	4.21	98.18	94.01	43.65	30 209 908	29 521 966	99.15
H-2	3.91	98.24	94.16	43.32	28 128 986	27 386 569	98.88
H-3	4.50	98.33	94.39	43.70	32 336 822	31 582 469	99.03
I-1	3.61	98.25	94.19	43.70	25 906 782	25 215 932	98.73
I-2	3.64	98.23	94.10	43.86	26 039 222	25 401 977	98.79
I-3	3.75	98.20	94.04	43.65	26 897 134	26 234 547	98.89
J-1	3.66	98.23	94.07	43.60	26 177 652	25 573 336	98.93
J-2	4.76	98.15	93.91	43.99	34 164 868	33 241 189	98.65
J-3	3.87	98.31	94.30	43.33	27 737 392	27 147 839	99.05
K-1	3.46	98.11	93.82	43.24	24 777 840	24 244 084	99.23
K-2	3.74	98.25	94.17	43.77	26 822 674	26 236 276	99.31
K-3	3.80	98.10	93.79	44.12	27 297 226	26 623 999	99.00
L-1	3.93	98.33	94.36	43.78	28 167 820	27 648 480	99.45
L-2	4.03	98.24	94.14	43.82	28 866 158	28 311 881	99.37
L-3	4.49	97.84	93.25	43.50	32 477 524	31 463 582	98.74
M-1	3.51	98.22	94.10	43.31	25 174 292	24 465 586	98.57
M-2	3.80	98.22	94.11	43.86	27 285 740	26 664 479	99.31
M-3	3.83	98.19	94.05	44.02	27 505 296	26 932 962	99.31
N-1	3.99	98.29	94.28	43.71	28 647 006	28 057 515	99.31
N-2	3.91	98.24	94.14	44.00	28 052 800	27 246 078	98.50
N-3	4.06	98.12	93.85	43.73	29 148 304	28 463 571	99.17
O-1	3.71	98.20	94.08	43.81	26 653 496	26 022 921	99.19
O-2	3.66	98.24	94.16	43.48	26 296 224	25 688 383	99.29

在系统发育树上与栽培类群Ⅲ的栽培稗草聚在一起；栽培样本H-1、H-2、H-3则完全归于野生类群Ⅰ

中，与野生类群Ⅰ中的其他野生样本聚在一起，且遗传关系介于栽培类型和野生类型两者之间。

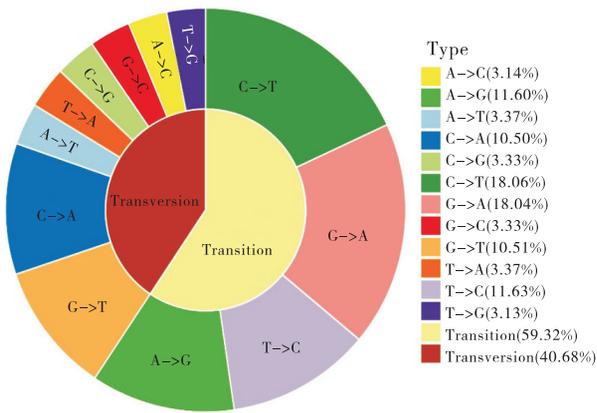


图1 SNP转换颠换图

Fig. 1 SNP conversion diagram

2.4 遗传结构及PCA分析

基于SNP分型数据,计算K=2~9时的交叉验证错误率(cross-validation error, CV error),结果表明(图3),当K=4时,交叉验证错误率最低,为稗草群体的最优分群数,即被划分为4个类群(图4),与系统发育

结果一致,可以较好地根据野生类型和栽培类型划分出4个类群,野生类群I为紫色,野生类群II为蓝色,栽培类群III为红色,栽培类群IV为绿色。类群I中的样本H-1、H-2、H-3、I-1、I-2、I-3、L-1、L-2、L-3只1种颜色组成,表明种群纯度高,而样本N-1、N-2、N-3由4种颜色组成,遗传结构最为复杂,表现为大比例的基因池混杂,说明这3份样本存在杂交的可能,与其他栽培群体和野生群体存在广泛的基因交流现象,可能是野生稗草和栽培稗草之间杂交渐渗的结果。类群II中的所有样本均由1种颜色组成,种群纯度高。类群III和类群IV中的样本除A-1、A-2、A-3、F-1、F-2、F-3、M-1、M-2、M-3、E-1、E-2、E-3只由1种颜色组成外,其余样本群体(BCDG)由2种相同的颜色组成,推测可能具有相同或相近的祖先来源,且归为类群III中的样本G-1、G-2、G-3与类群IV遗传关系更为密切。

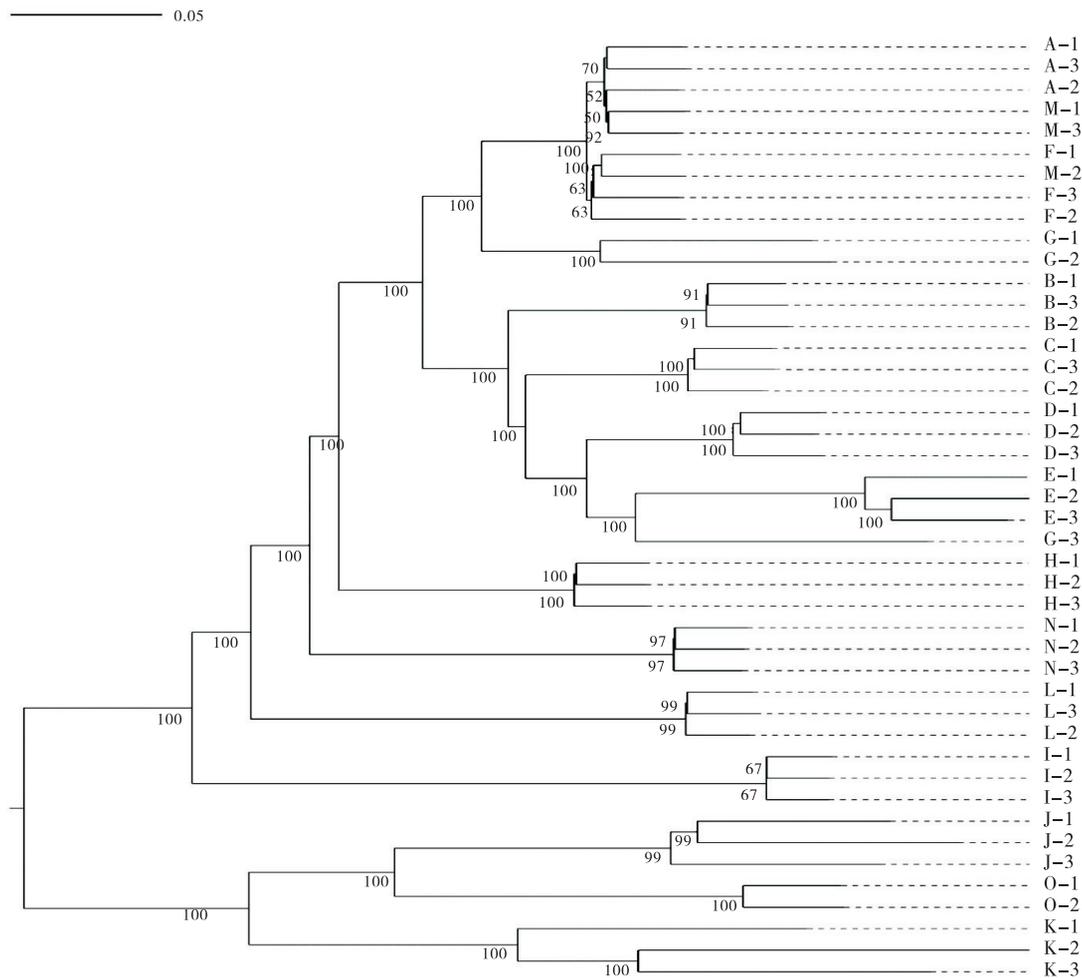


图2 44份稗草资源系统进化树

Fig. 2 The phylogenetic tree of the 44 *E. spp.* samples

以得到的 SNP 标记为基础进行主成分分析, PC1 和 PC2 累计方差贡献率为 41.52% (图 5)。24 份栽培个体和 20 份野生个体在空间上表现出较为明显的分离趋势, 野生类群 II 较为紧密地聚集在坐标轴的右侧, 与野生类群 I 和 2 个栽培类群 III 和 IV 在空间上距离较远, 遗传关系相对较远, 而野生类群 I 与两个栽培类群 III 和 IV 较为紧密的聚在一起, 遗传关系较近。2 个栽培类群紧密聚在主成分坐标轴左侧, 遗传基础总体狭窄, 遗传关系最为相近。

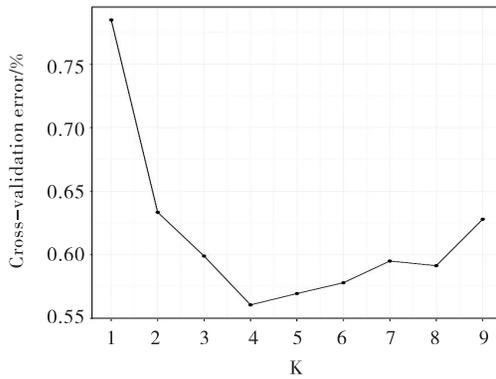


图 3 44 份稗草材料群体遗传结构分析, 计算 K 为 2~9 时的 CV error

Fig. 3 The population genetic structure of 44 *E. materials*, and the cross-validation (CV) error was

注: Calculated when K was 2~9.

2.5 遗传多样性及遗传分化分析

为研究野生与栽培稗草之间的群体遗传多样性, 比较野生稗草与栽培稗草的观测杂合度 (H_o)、期望杂合度 (H_e) 和遗传多样性指数 (P_i) (表 3)。 H_o 为 0.1007 (类群 IV)~0.2264 (类群 II); H_e 为 0.1404 (类群 III)~0.2512 (类群 I); P_i 为 0.000101 (类群 IV)

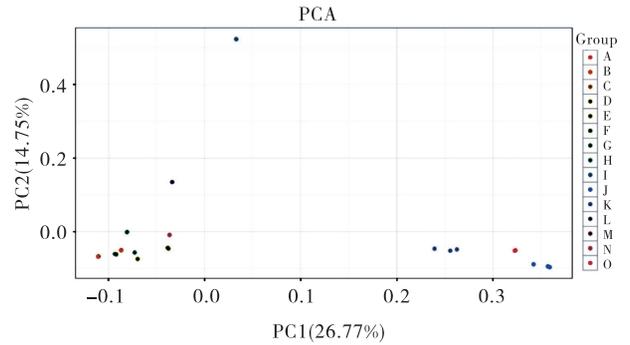


图 4 44 份稗草主成分分析

Fig. 4 Principal component analysis (PCA) of the 44 *E. spp. materials*

—0.000154 (类群 II)。由表可以看出, 野生类群 I 和 II 的 H_o 、 H_e 和 P_i 均高于栽培类群 III 和 IV, 野生稗草群体的遗传多样性水平明显高于栽培稗草群体, 其中野生类群 II 的遗传多样性水平最高, 说明该群体的遗传变异丰富, 但四个类群的 P_i 值很小, 表明稗草总体上遗传多样性水平偏低。基于群体 SNP 标记和群体遗传结构分析划分的结果 (类群 I、类群 II、类群 III 和类群 IV), 计算其 4 个类群间的遗传分化 F_{ST} 值和基因流 N_m 值 (表 4), 4 个类群间的遗传分化指数为 0.2892~0.5708, 类群 II 和类群 III 之间的 F_{ST} 值最大 (0.5708), 类群 I 和类群 III、类群 II 和类群 IV 之间的 F_{ST} 值较大 (0.3413 和 0.4816), 类群 I 和类群 IV 之间的 F_{ST} 值最小 (0.2892), 说明在野生稗草群体与栽培稗草群体间有显著的遗传分化, 其中野生类群 II 和栽培类群 III 之间的分化最为显著; 野生稗草与栽培稗草类群基因流结果矩阵显示, 两两种群之间的基因流 (N_m) 为 0.1880~0.6144, 4 个类群间的 N_m 值均小于 1, 说明群体间的基因交流很少, 存在显著的区域分化。

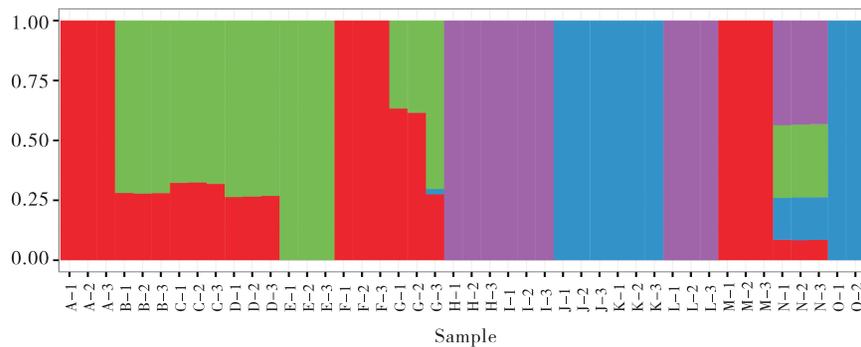


图 5 K=4 时 44 份稗草的群体结构

Fig. 5 The population structure of 44 *E. spp.* at K=4

表3 稗草群体遗传参数统计

Table 3 Statistics of genetic parameters in *E. spp.* populations

群体	观测杂合度(H_o)	期望杂合度(H_e)	遗传多样性指数(P_i)
类群 I	0.155 1	0.251 2	0.000 151
类群 II	0.226 4	0.250 7	0.000 154
类群 III	0.100 7	0.140 4	0.000 061
类群 IV	0.131 4	0.166 3	0.000 101

表4 群体间遗传分化指数(F_{ST})(对角线下)和基因流(对角线上)Table 4 Index of genetic differentiation (F_{ST}) (below diagonal) and N_m value above diagonal) between populations

群体	类群 I	类群 II	类群 III	类群 IV
类群 I		0.364 0	0.482 6	0.614 4
类群 II	0.407 1		0.188 0	0.269 1
类群 III	0.341 3	0.570 8		0.421 7
类群 IV	0.289 2	0.481 6	0.372 2	

3 讨论

遗传多样性是生物多样性的重要组成部分之一,是物种适应环境变化的基础,遗传多样性水平越高或遗传变异越强,物种的进化潜力和对复杂多变栖息地的适应能力就越强^[19-20]。基于开发的 SNP 位点,计算用于衡量群体遗传多样性水平高低的参数:观测杂合度(H_o)、期望杂合度(H_e)和遗传多样性指数(P_i)。研究结果发现野生稗草的遗传多样性指数 H_o 、 H_e 和 P_i 都比栽培稗草高得多,这与贺明月^[12]利用简化基因组测序技术研究不同类型糜子(*Panicum miliaceum*)间的群体遗传多样性所得出的结果相一致,野糜子群体遗传多样性水平高于栽培糜子群体。此外,赵亚琴等^[21]利用简化基因组测序技术对野生和栽培种啤酒花(*Humulus lupulus*)进行了 SNP 位点开发及遗传结构分析,结果表明野生个体和栽培个体之间存在较大的遗传分化,野生啤酒花个体具有较高的遗传多样性水平。以上研究中野生资源遗传多样性高于栽培资源的原因可能是因为:在作物驯化过程中,其遗传多样性与它们的祖先遗传类型相比会降低,说明作物驯化时发生了瓶颈效应^[22]。系统发育分析、主成分分析以及群体结构分析均表明野生稗草和栽培稗草间存在遗传差异,这可能与驯化瓶颈和人工选择有关。本研究中,群体遗传分化指数为 0.289 2~0.570 8,高于 0.25^[23],这表明群体间有很大的遗传分化,而 4 个类群之间的遗传分化指数均大于 0.25,说明稗草类群之间

存在很大的遗传分化,且野生类群和栽培类群之间的分化最为显著,亲缘关系最远。前人指出 $N_m < 1$ 时,遗传漂变可以引起种群间产生明显的遗传分化,导致遗传变异减少^[23],在本研究中稗草群体间的 N_m 值均小于 1,说明稗草之间存在很大的遗传分化是由遗传漂变引起的。而近交以及遗传漂变也是造成稗草遗传多样性水平偏低的原因之一,基于遗传多样性系数结果显示,两种类型稗草的期望杂合度(H_e)均高于观测杂合度(H_o),存在一定的近交现象,以致于纯合子过量,遗传结构也显示大部分参试材料只由一种颜色组成,种群纯度高,进而导致遗传多样性下降或者丧失;同时,稗属植物的生境受到人类活动的破坏与干扰,由于稗草多与水稻伴生于农田中,除草剂的使用等人为干扰导致稗草的生存和繁育受到威胁,稗草种群进一步缩小以及稗草是一种严格的自花授粉植物,与其他稗草进行基因交流的机会很小,导致遗传多样性丢失也是引起引起遗传多样性水平偏低的原因。因此,遗传漂变、人类活动以及授粉方式可能是导致稗草植物遗传多样性水平下降的重要原因。一般而言,珍稀植物以及极小种群的遗传多样性水平往往处于较低水平,如珍稀濒危植物太行菊属(*Opisthopappus*)^[24]、极小种群野生植物海南风吹楠(*Horsfieldia hainanensis*)^[25]。而稗草遗传多样性水平比一些珍稀甚至是极小种群的遗传多样性水平低,警示我们应该针对稗草资源提出合理的保护措施。而探究野生与栽培稗草之间的遗传关系,明晰不同类型稗草之间的遗传差

异,有利于栽培稗草农艺性状的改良和为分子驯化机制的研究提供有效的遗传信息,加快对野生稗草资源的开发利用,推动中国畜牧业发展。

多数作物的野生群体通常被认为是由野生型(wild type)和野化型(feral type)组成,而野化型进一步分为内源野化型(endo feral type)和外源野化型(exo feral type),其中栽培群体的单一谱系形成了内源野化型,而外源野化型是通过栽培型与其野生近缘种杂交或栽培群体不同谱系之间相互混合而形成的^[26-28]。本研究对野生和栽培稗草的遗传多样性以及遗传结构进行了分析,遗传多样性结果显示栽培稗草低于野生稗草,说明了这些野生稗草中应该包含了大量的野生型,其中可能就包含了栽培稗草祖先的遗传类型,假设这些野生稗草都是由栽培稗草逃逸而形成的野化型所构成的,那么野生稗草的遗传多样性应该小于栽培稗草。从遗传结构分析结果可以看出,两个野生稗草类群均能单独聚为一个组群,与栽培稗草类群区分开来,只有3个样本M-1、M-2、M-3归于栽培类群Ⅲ中,3个野生样本N-1、N-2、N-3与栽培类群Ⅲ和Ⅳ基因池有大比例混合,在系统发育树和PCA图上,野生种M-1、M-2、M-3与栽培稗草聚在一起,说明其与栽培稗草具有更近的亲缘关系,而野生种N-1、N-2、N-3与栽培种H聚在一起,介于野生稗草和栽培稗草类群样本之间,暗示其杂交起源的可能。栽培种H(朝牧一号)为辽宁省朝阳市畜牧所经多年选育而成^[29],却与野生类群Ⅰ聚在一起,可能是一方面由于其本身作为一种杂草,在驯化成为作物进行种植时通常驯化不彻底;另一方面在没驯化成为作物时多生长在主要作物不宜种植的区域,保留了部分杂草基因组特征利于其环境适应^[30]。因此推测:单独归于野生类群Ⅰ和Ⅱ的稗草可能为野生型,而野生样本M-1、M-2、M-3则可能属于栽培稗草中逃逸出来的内源野化型,野生样本N-1、N-2、N-3则可能属于栽培稗草与野生稗草杂交渐渗而形成的外源野化型。由此可见,基于全基因组水平的SNP多态性能够为野生型和内源、外源野化型稗草鉴定提供有效方法,而正确辨别种质是挖掘优质野生种质资源的前提,对揭示种质和种源驯化的传播研究意义重大。然而在本研究中所选取的野生稗草样本数量有限,在未

来的研究中应针对更多的野生样本与栽培样本进行研究,并针对人工选择的作用进行研究验证。

4 结论

运用RAD-seq测序技术对9个种源44份稗草样本进行测序,开发了31 284个高质量SNP位点,进而分析稗草群体的系统进化树、主成分、遗传结构以及多样性。15个群体被划分为4个类群,分别为2个野生类群和2个栽培类群,SNP标记可将野生稗草和栽培稗草区分开来,证明了其在稗草分类方面的可行性,此外遗传多样性结果表明野生稗草的遗传多样性明显高于栽培稗草,其中野生类群Ⅱ(稗、水田稗、西来稗)遗传多样性水平最高,群体之间的遗传分化程度很大。由遗传结构以及遗传多样性水平结果我们推测供试材料野生稗草中应该包含了大量的野生型,其中可能就包含了栽培稗草的祖先遗传类型,单独归于野生类群Ⅰ和Ⅱ的稗草可能为野生型,而野生稗草群体M则可能属于从栽培稗草中逃逸出来的内源野化型,野生稗草群体N则可能属于栽培稗草与野生稗草杂交渐渗而形成的外源野化型。

参考文献:

- [1] 陈守良. 中国植物志第九卷[M]. 北京:科学出版社, 1990:250-261.
- [2] 张婉茹,吴妹菊,王晓萍. 稗属植物种质资源及其应用价值研究进展[J]. 草业科学,2018,35(1):165-175.
- [3] ALTOP E K, MENNAN H. Genetic and morphologic diversity of *Echinochloa crusgalli*, populations from different origins[J]. *Phytoparasitica*, 2011, 39(1):93-102.
- [4] KAYA HB, DEMIRCI M, TANYOLAC B. Genetic structure and diversity analysis revealed by AFLP on different *Echinochloa* spp. from northwest Turkey[J]. *Plant Systematics and Evolution*, 2014, 300(6):1337-1347.
- [5] DANQUAH EY, JOHNSON D E, RICHES C, et al. Genetic diversity in *Echinochloa* spp. collected from different geographic origins and within rice fields in Cote d'Ivoire [J]. *Weed Research*, 2002, 42(5):394-405.
- [6] 陆永良,刘德好,郭水良,等. 基于ISSR标记的中国稻田稗属植物分类[J]. 浙江农业学报, 2014, 26(5):1309-1314.
- [7] 张婉茹. 寒地稗草种质资源遗传多样性研究[D]. 哈尔滨:哈尔滨师范大学, 2018.

- [8] 王洋坤,胡艳,张天真. RAD-seq技术在基因组研究中的现状及展望[J]. 遗传,2014,36(1):41-49.
- [9] 刘学锋,李小梅,张国武,等. 基于RAD高通量测序的蓝花楹群体遗传分析[J]. 亚热带植物学报,2022,30(5), 613-622.
- [10] 尹婷婷. 披碱草属种质资源表型性状与基于RAD-seq测序的遗传多样性分析[D]. 乌鲁木齐:新疆农业大学, 2021.
- [11] 潘文婷,孙建军,原勤勤,等. RAD-seq技术研究鹅掌楸属种源遗传多样性和遗传结构[J]. 林业科学,2022, 58(4):74-81.
- [12] 贺明月. 基于简化基因组测序的糜子驯化与传播研究 [D]. 长春:吉林大学,2020.
- [13] LIU L W, ZHAO L P, GONG Y Q, *et al.* DNA fingerprinting and genetic diversity analysis of late-bolting radish cultivars with RAPD, ISSR and SRAP markers[J]. *Scientia Horticulturae*,2008,116(3):240-247.
- [14] PETERSON B K, WEBER J N, KAY E H, *et al.* Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species [J]. *PLoS ONE*, 2012, 7(5) : e37135.
- [15] DEPRISTO M. A. ,BANKS E. ,POPLIN R. , *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data [J]. *Nature Genetics*,2011,43(5):491-498.
- [16] VILELLA A J, SEVERIN J, URETA-VIDAL A, *et al.* Ensembl Compara Gene Trees: Complete, duplication-aware phylogenetic trees in vertebrates [J]. *Genome Research*,2009,19(2):327-335.
- [17] ALEXANDER D H, NOVEMBER J, LANGE K. Fast model-based estimation of ancestry in unrelated individuals[J]. *Genome Research*,2009,19(9):1655-1664.
- [18] YANG J, LEESH, GODDARD ME, *et al.* GCTA: A Tool for Genome-wide Complex Trait Analysis [J]. *The American Journal of Human Genetics*,2011,88(1): 76-82.
- [19] LI X L, WANG J, FAN Z Q, *et al.* Genetic diversity in the endangered *Camellia nitidissima* assessed using transcriptome-based SSR markers [J]. *Trees*, 2020, 34(2):543-552.
- [20] ZHANG X, SU H L, YANG J, *et al.* Population genetic structure, migration, and polyploidy origin of a medicinal species *Gynostemma pentaphyllum* (Cucurbitaceae) [J]. *Ecology and Evolution*,2019,9(19):11145-11170.
- [21] 赵亚琴,樊丛照,张际昭,等. 基于简化基因组技术的啤酒花栽培种和野生种 SNP 位点开发及遗传结构分析 [J]. *中草药*,2021,52(20):6365-6372.
- [22] DOEBLEY J F, GAUT B S, SMITH B D. The Molecular Genetics of Crop Domestication [J]. *Cell*, 2006, 127(7):1309-1321.
- [23] WRIGHT S. Variability within and among natural populations [M]. Chicago:University of Chicago Press,2010.
- [24] 侯慧敏. 太行菊属(*Opisthopappus* Shih)群体遗传学分析[D]. 太原:山西师范大学,2020.
- [25] 蔡超男,侯勤曦,慈秀芹,等. 极小种群野生植物海南风吹楠的遗传多样性研究 [J]. *热带亚热带植物学报*, 2021,29(5):547-555.
- [26] GERING E, INCORVAIA D, HENRIKSEN R, *et al.* Getting Back to Nature: Feralization in Animals and Plants [J]. *Trends in Ecology & Evolution*, 2019, 34(12):1137-1151.
- [27] WANG H, VIEIRA F G, CRAWFORD J E, *et al.* Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice [J]. *Genome Research*,2017,27(6):1029-1038.
- [28] BARAZANI O, KEREN-KEISERMAN A, WESTBERG E, *et al.* Genetic variation of naturally growing olive trees in Israel: from abandoned groves to feral and wild? [J]. *BMC Plant Biology*, 2016, 16(1) : 1-11.
- [29] 张俊丽,杨斌,蒙静. 朝牧一号稗子引种试验研究 [J]. *宁夏农林科技*,2013,54(3):5-6.
- [30] WU D, SHEN E, JIANG B, *et al.* Genomic insights into the evolution of *Echinochloa* species as weed and orphan crop [J]. *Nature communications*,2022,13(1):689.
- [31] 杨娟弟,祁娟,贾燕伟,等. 10个披碱草属野生植物材料生产性能综合评价 [J]. *草原与草坪*, 2022, 42(5) : 54-61.

Research on genetic diversity of *Echinochloa* based on RAD-seq

NIE Hao-jie^{1,2,3}, LI Lin-ao^{2,3,4}, JING Qing-fang^{1,2,3}, ZHANG Yang^{1,2,3}, ZHU Lin^{1,2,3*}

(1. School of Ecology and Environmental Ningxia University, Ningxia Yinchuan 750021; 2. Breeding Base for State Key Lab. of Land Degradation and Ecological Restoration in Northwestern China Ningxia University, Ningxia Yinchuan 750021; 3. Key Lab. for Restoration and Reconstruction of Degraded Ecosystems in Northwestern China of Ministry of Education Ningxia University, Ningxia Yinchuan 750021; 4. School of Agriculture Ningxia University, Ningxia Yinchuan 750021)

Abstract: [Objective] The study was to analyze the genetic diversity level of *Echinochloa* spp. by SNP markers, to discover the genetic relationship and structure between wild and cultivated material and to provide a basis for the protection of genetic diversity, and utilization of wild resources of *Echinochloa* spp. [Method] 20 samples of wild *Echinochloa* and 24 samples of cultivated *Echinochloa* spp. were collected. The SNP markers of all samples were identified by RAD-seq. The phylogenetic analysis, principal component analysis, population genetic structure and genetic parameter were used to reveal the genetic structure and genetic diversity between wild and cultivated population. [Result] A total of 161.51 Gb sequence data was obtained, and 31,284 high-quality SNP loci were retained after filtering. The results of genetic structure indicated that 44 samples of *Echinochloa* could be divided into 4 groups, among which group I and II were wild resources, and group III and IV were cultivated resources. The nucleotide diversity index (P_i) for 4 groups of *Echinochloa* was 0.000 061~0.000 154, the index of genetic differentiation (FST) was 0.289 2~0.507 8. The observed heterozygosity (H_o) was 0.115 1~0.226 4 and the expected heterozygosity (H_E) was 0.100 7~0.251 2. The nucleotide diversity index (P_i), observed heterozygosity (H_o) and expected heterozygosity (H_E) of wild group I and II were higher than those of cultivated group III and IV. The genetic differentiation results indicated that genetic differentiation occurred between four groups, and wild group II and cultivated group IV were the most significant difference. [Conclusion] SNP markers could distinguish wild *Echinochloa* and cultivated *Echinochloa*, which proved the feasibility of classifying *Echinochloa*. The highest genetic diversity level was found in wild *Echinochloa* group II (*E. crus-galli* (L.), *E. oryzoides* and *E. crus-galli* var. *zelayensis*), however, in general, the level of genetic diversity of *Echinochloa* was lower; And it was speculated that genetic drift, human activities and pollination practices may be crucial reasons for the decline in genetic diversity and genetic differentiation of *Echinochloa*.

Key words: *Echinochloa*; RAD; SNP; genetic structure; genetic diversity