

网络出版时间:2025-05-27

网络出版地址:https://link.cnki.net/urlid/61.1359.S.20250526.1605.004

基于 RNA-seq 的冰草 EST-SSR 标记开发及验证

杨东升¹, 李冉¹, 宝格日乐¹, 王海伟¹, 卢倩倩¹, 郝水源¹, 刘建华²

(1. 河套学院/硬质小麦技术创新中心/河套地区绿色农产品安全生产与预警控制实验室,
内蒙古巴彦淖尔 015000; 2. 巴彦淖尔市临河区农牧和科技局, 内蒙古巴彦淖尔 015000)

摘要:冰草是小麦抗逆遗传改良的重要近缘植物,开发冰草基因组分子标记对于加快冰草的遗传育种效率和利用具有重要意义。本研究基于转录组测序(RNA-seq)技术,对2个冰草材料的幼嫩根、茎、叶片及结实期的花穗、种子进行生物信息学分析,开发高多态性的EST-SSR标记并进行验证。结果表明,共获得23 491 793条高质量序列,其碱基长7.03 Gb,GC含量56.44%,Q30为93.04%。共获得33 993条Unigene,总长度为29 309 958 bp。将所获Unigene与9大功能数据库进行比对,共有25 614条Unigene获得功能注释,其中Nr数据库共注释24 557条,GO数据库注释到66 193条Unigene,按功能分为3个大类和42个亚类;共14 412条Unigene注释到KEGG数据库的136条代谢通路;有11 321条Unigene在KOG数据库获得注释,归属为25个功能类别;共预测到大于100 bp的CDS有7 288条,SSR位点2 443个,其中三碱基重复1 439个,占总数的58.90%;获得EST-SSR多态性引物42对,从中随机选择10对在冰草的30个F₂代分离单株进行有效性验证,平均多态性比率为48.33%。以上结果说明利用RNA-seq技术大量开发的EST-SSR引物,可高效地应用于四倍体冰草产量、品质性状相关的标记开发、优异新种质指纹图谱构建及精准分子标记辅助育种。

关键词:冰草;RNA-seq;生物信息学分析;EST-SSR标记开发;验证

中图分类号:S512.9;S336

文献标识码:A

文章编号:1009-1041(2025)07-0891-11

EST-SSR Marker Development and Verification of Wheatgrass Based on RNA-seq

YANG Dongsheng¹, LI Ran¹, BAO Gerile¹, WANG Haiwei¹,
LU Qianqian¹, HAO Shuiyuan¹, LIU Jianhua²

(1. Hetao College/Technological Innovation Center for Durum Wheat/Laboratory of the Green Agricultural Product Safety Production and Early Warning Control in Hetao Area, Bayannur, Inner Mongolia 015000, China; 2. Agricultural Industries, and Science and Technology Bureau of Linhe District, Bayannur, Inner Mongolia 015000, China)

Abstract: *Agropyron* is an important relative plant for genetic improvement of stress resistance in wheat. The development of molecular markers is of great significance to accelerate the efficiency of genetic breeding and innovative utilization of wheatgrass. In this study, based on transcriptome sequencing technology, the young roots, stems, leaves, spikes at grain-filling stage, and seeds of two wheatgrass accessions have been bioinformatics analyzed, and highly polymorphic EST-SSR markers were developed and verified. The results showed that a total of 23 491 793 high-quality sequences were obtained, with a base length of 7.03 Gb, the GC content of 56.44%, and the Q30 content of 93.04%. 33 993 Unigenes were obtained, with a total length of 29 309 958 bp. A total of 25 614 Unigene annotations

收稿日期:2024-10-03 修回日期:2025-04-20

基金项目:内蒙古自治区自然科学基金面上项目(2023MS03049);河套学院高层次人才引进科研启动经费资助项目(HYRC202310);内蒙古自治区“科技兴蒙”行动重点专项(NMKJXM202201);巴彦淖尔市科技计划项目(NMKJXM202408)

第一作者 E-mail:yangdongsheng007@163.com(杨东升)

通讯作者 E-mail:haoshuiyuan@126.com(郝水源)

were obtained by comparing Unigene with nine functional databases, including 24 557 annotations in the Nr database and 66 193 annotations in the GO database, which were divided into 3 categories and 42 subcategories according to their functions. A total of 14 412 Unigene pathways were annotated into 136 metabolic pathways in the KEGG database. A total of 11 321 Unigenes were annotated in the KOG database, categorized into 25 functional categories. A total of 7 288 CDSs with 2 443 SSR sites greater than 100 bp were predicted, of which 1 439 were tribase repeats, accounting for 58.90% of the total. A total of 42 pairs of primers with EST-SSR polymorphisms were obtained, and 10 pairs were randomly selected for validation using 30 F_2 individuals. The average polymorphism ratio was 48.33%. The results indicated that the EST-SSR primers, which were developed in large quantities by RNA-seq technology, can be efficiently applied to the development of markers related to yield and quality traits of tetraploid wheatgrass, the construction of excellent new germplasm fingerprints, and precise molecular marker-assisted breeding.

Keywords: Wheatgrass; RNA-seq; Bioinformatics analysis; EST-SSR marker development; Validation

冰草属 (*Agropyron*) 植物隶属于小麦族 (Triticeae) 多年生的禾本科牧草, 多生长在荒漠草原、高山、沙丘、草甸等土壤贫瘠的环境, 主要分布于中国的西北、华北、东北等干旱区^[1-3]。中国常见的有蒙古冰草 (*A. mongolicum* Keng, $2n = 2x = 14$)、沙生冰草 (*A. desertorum* Schult, $2n = 2x = 14$)、西伯利亚冰草 (*A. fragile* Canad, $2n = 2x = 14$)、根茎冰草 (*A. michnoi* Roshev, $2n = 2x = 14$)、扁穗冰草 (*A. cristatum* Gaertn, $2n = 2x = 14$) 等 5 个种, 有多花冰草 (*A. cristatum* var. *pluriflorum*)、光穗冰草 (*A. cristatum* var. *pecriforme*)、毛沙生冰草 (*A. desertorum* var. *pilosiusculum*)、毛沙芦草 (*A. desertorum* var. *vilosum*) 等 4 个变种, 以及毛稈冰草 (*A. sibiricum* f. *pubiflorum*) 1 个变型^[4-5]。其中蒙古冰草在内蒙古地区分布极为广泛, 具有耐寒耐旱、适应性强等特点; 扁穗冰草中的航道冰草 (*A. cristatum* cv. *Fairway*, $2n = 2x = 14$) 具有叶量大、茎秆柔软、适口性好等特点^[6]。冰草春季返青早, 秋季枯黄晚, 是早春和晚秋时节牛、羊、马、骆驼等家畜易采食的优质牧草, 也常被用于人工草地建植、退化草场补播、麦类作物杂交改良和遗传转化, 极具饲用价值、生态价值和遗传研究价值^[6-8]。

近年来, 关于冰草的研究已有较多报道, 常怡然等^[9]对荒漠草原不同自然条件下的蒙古冰草种群生物量和养分权衡特征进行了分析, 发现在一定程度上蒙古冰草种群可调节自身资源分配来应对特化的生境; 杨靖等^[10]通过对内蒙古地区的 30 份冰草属种质资源进行表型性状的遗传多样性分析, 将其聚为 3 大类, 矮株密穗细茎型的冰草占总样本量 86.67%; 杨东升等^[11]基于 GBS-seq 技术

构建了一张含 14 个连锁群、标记间距为 0.427 cM 的四倍体冰草超高密度分子遗传连锁图谱; 樊璐等^[12]基于转录组测序技术筛选获得 6 个高表达量的蒙古冰草抗旱调控关键基因。

目前, 关于冰草分子标记开发研究报道甚少, 多集中在冰草重要性状相关基因序列的分析与鉴定上。王贺萍等^[13]基于全长转录组测序技术, 筛选并克隆到 1 个参与木质素合成的候选基因 CAD 序列, 其长度为 1 083 bp。黄馨田等^[14]通过对根茎冰草转录组测序, 克隆到 1 个长度为 930 bp、与发育相关的 *AmrbHLH137* 基因 cDNA 序列。闫秀秀等^[15]通过对干旱胁迫下蒙古冰草转录组测序数据进行分析, 获得 6 个高表达量的 GDSL 酯酶基因序列。本研究采用高通量测序平台 PE150 对航道冰草和蒙古冰草的幼嫩根、茎、叶片及结实期的花穗、种子进行转录组测序, 通过组装建立冰草转录组数据库, 对获得的 Unigene 进行功能注释、代谢通路及 EST-SSR 分析, 以期在今后冰草重要性状相关分子标记开发、功能基因挖掘、种质资源创新利用及分子遗传育种提供科学依据。

1 材料与方 法

1.1 供试材料

材料为多年生蒙古冰草、航道冰草, 及其种间杂种 F_1 经秋水仙碱溶液进行染色体加倍、套袋自交获得的 30 个 F_2 单株, 种植于河套学院农作物试验田, 分别取航道冰草和蒙古冰草苗期的根、茎、叶片及结实期的花穗、种子各 5 份组织样品, 均为混合取样, 每份组织材料 3 个重复。液氮速冻后, 于 $-80\text{ }^{\circ}\text{C}$ 的超低温冰箱中保存备用。

1.2 RNA、DNA 提取与检测

采用天根生化科技有限公司的植物基因组 RNA 提取试剂盒,通过 Trizol 法分别提取蒙古冰草和航道冰草各样本的总 RNA。利用 Agilent 2100 检测 RNA 完整性,用 NanoDrop 2000 超微量分光光度计检测 RNA 浓度,检测合格后用于 cDNA 文库构建。

利用植物基因组试剂盒分别对各样品材料进行 DNA 提取。用紫外凝胶成像系统检测 DNA 纯度,利用超微量分光光度计 NanoDrop ND 2000 测定其浓度。将样品 DNA 稀释至 $50 \text{ ng} \cdot \mu\text{L}^{-1}$,用于 EST-SSR 标记引物的 PCR 扩增。

1.3 cDNA 文库构建与质控

将 2 种冰草的各组织材料 RNA 浓度均稀释至 $80 \text{ ng} \cdot \mu\text{L}^{-1}$,利用移液枪等体积吸取各组织材料的 RNA 至同一个 2 mL 的离心管中,充分混匀。将冰草的 RNA 混样,寄到北京百迈客生物科技有限公司进行 cDNA 文库构建。利用 Qubit 3.0 荧光定量仪对构建的文库进行初步定量,要求浓度大于 $1 \text{ ng} \cdot \mu\text{L}^{-1}$,然后用 Qsep 400 高通量分析系统对文库的插入片段进行检测,插入片段符合预期后,通过 qPCR 法对文库的有效浓度 ($>2 \text{ nmol} \cdot \text{L}^{-1}$) 进行准确定量,以确保文库质量。

1.4 转录组测序、数据组装

利用高通量测序平台 PE150 进行转录组测序(RNA-seq),通过 Trinity 软件对测序获得的 Raw Reads 进行短片段(K-mer)打断,然后延伸成长片段(Contig),最后对高质量的 Clean Data 进行组装,获得转录本序列(Transcript),并统计分析 Unigene。

1.5 Unigene 功能注释

利用 DIAMOND 软件将 Unigene 序列与 Nr、Swiss-Prot、COG、KOG、eggNOG、KEGG、GO、Pfam、TrEMBL 等九大数据库进行比对,其中用 InterPro 整合的数据库对新基因的 GO 结果进行分析,用 HMMER 软件与 Pfam 数据库比对,获得 Unigene 的注释信息。

1.6 CDS 预测

基于开放阅读框(ORF)长度、对数似然函数值(Log-likelihood score)和 Pfam 数据库蛋白质结构域序列的比对等信息,使用 Trans Decoder 软件对 Unigene 的编码区序列(coding sequence, CDS)进行预测。

1.7 EST-SSR 分子标记开发及多态性引物筛选与验证

用 STAR 软件对 Unigene 序列进行比对,通

过 MISA 软件与获得的 Unigene 进行 SSR 位点挖掘。利用 3 个杂交冰草 F_2 分离单株及其双亲的 DNA 为模板,对初步获得的 EST-SSR 引物进行多态性筛选,完整的引物序列委托生工生物工程(上海)股份有限公司合成。从筛选获得的 42 对多态性引物中,随机选取 10 对进行四倍体冰草 30 个 F_2 单株的 PCR 扩增,验证其有效性。参照于肖夏等^[16]提出的方法进行 EST-SSR 的 PCR 基因片段扩增及聚丙烯酰胺凝胶电泳检测。

2 结果与分析

2.1 高通量测序、数据组装

通过对冰草进行 RNA-seq,共获得高质量序列 23 491 793 条,其碱基长 7.03 Gb,GC 含量为 56.44%,错误率 0.01%,Q20 和 Q30 分别为 97.44% 和 93.04%,说明冰草的测序数据较丰富,完全满足后续的生物信息学分析要求。

由表 1 可知,共获得 Unigene 33 993 条,Unigene 长度为 200~300 bp 的最多,有 8 915 条(26.23%); $>2 000 \text{ bp}$ 的最少,为 2 999 条(8.82%) 条;总长度为 29 309 958 bp,且 N50 长度为 1 387 bp,远大于其平均长度(862.24 bp),表明组装完整度较高。

2.2 冰草 Unigene 总体注释

利用软件 BLAST 和 HMMER 对组装的 Unigene 与 9 大功能数据库进行比对、注释,分别设置参数 E 值为 $E \leq 10^{-5}$ 和 $E \leq 10^{-10}$,结果见表 2。冰草注释成功的 Unigene 有 25 614 条,其中注释 Unigene 数目最多的是 TrEMBL 和 NR,分别为 25 285 条(占 98.72%)和 24 557 条(占 95.87%);eggNOG 有 19 870 条,占 77.57%;GO 有 19 321 条,占 75.43%;Pfam 有 14 669 条,占 57.27%;KEGG 有 14 412 条,占 56.27%;Swissprot 有 14 225 条,占 55.54%;KOG 有 11 321 条,占 44.20%;GO 数据库注释到的 Unigene 最少,仅有 5 519 条。结果显示,注释到基因序列长度在 $300 \text{ bp} \leq$ 长度 $<1 000 \text{ bp}$ 的有 10 354 条,长度 $>1 000 \text{ bp}$ 的有 10 268 条, $<300 \text{ bp}$ 的有 4 992 条,说明 Unigene 总体注释结果良好。

2.3 Nr 功能注释

如图 1 所示,将 Unigene 与 Nr 数据库进行相似序列匹配,共有 24 557 条 Unigene 被成功注释,其中冰草与粗山羊草(*Aegilops tauschii* subsp.)同源的序列最多,占 Nr 总注释的 33.40%;其次为硬粒小麦(*Triticum turgidum* subsp.)、大

表 1 冰草组装结果统计表
Table 1 Statistics of assembly transcriptome sequences in wheatgrass

长度范围 Length range/bp	转录本数(占总数的比例) Transcripts (Percentage of transcripts in total)	单基因数(占总数的比例) Unigene (Percentage of Unigene in total)
200~300	10 232(18.44%)	8 915(26.23%)
300~500	9 921(17.88%)	6 798(20.00%)
500~1 000	14 521(26.17%)	7 541(22.18%)
1 000~2 000	15 441(27.83%)	7 740(22.77%)
>2 000	5 364(9.67%)	2 999(8.82%)
总数 Total number	55 479(100%)	33 993(100%)
总长度 Total length	53 206 311	29 309 958
N50 length	1 403	1 387
平均长度 Mean length	959.04	862.24

表 2 冰草 Unigenes 注释统计
Table 2 Statistics of Unigenes annotations of wheatgrass

数据库 Database	注释数量 Annotated number	占总注释数量的比例 Percentage of annotated number in total/%	各片段数量 Number of segments		
			<300 bp	300 bp≤长度 length<1 000 bp	≥1 000 bp
COG	5 519	21.55	328	1 372	3 819
GO	19 321	75.43	3 555	7 309	8 457
KEGG	14 412	56.27	2 140	4 973	7 299
KOG	11 321	44.20	1 337	3 720	6 264
Pfam	14 669	57.27	1 323	4 944	8 402
Swissprot	14 225	55.54	1 920	4 934	7 371
TrEMBL	25 285	98.72	4 892	10 174	10 219
eggNOG	19 870	77.57	3 546	7 545	8 779
Nr	24 557	95.87	4 614	9 796	10 147
All	25 614	100	4 992	10 354	10 268

麦(*Hordeum vulgare*)、青稞(*Hordeum vulgare* subsp.)、乌拉尔图小麦(*Triticum urartu*)、普通小麦(*Triticum aestivum*)、二穗短柄草(*Brachypodium distachyon*), 占比分别为 31.64%、12.62%、6.32%、6.23%、3.24%、1.41%; 与水稻 *Oryza sativa* Japonica、玉米 *Zea mays*、弯叶画眉草 *Eragrostis curvula* 这 3 种植物注释最少, 分别占 0.41%、0.27%、0.27%; 与黑麦草(*Lolium perenne* L.)、老芒麦(*Elymus sibiricus* Linn.)、偃麦草(*Elytrigia intermedia* Nevski) 等其他物种相匹配的占比为 4.18%。

2.4 GO 功能注释分类

在 GO 数据库中对冰草的 Unigene 进行功能分类, 注释到分子功能(molecular function)、细胞组分(cellular component)和生物进程(biological process)3 大类, 包含 42 个亚类(图 2)。其中, 注释

到分子功能的 16 个亚类中的 Unigene 共有 20 922 条, 以分子结合功能亚类和催化活性亚类占比最高, 注释数量分别为 10 267 和 8 308 条; 注释到细胞组分的 11 个亚类中的 Unigene 共有 17 374 条, 以细胞解剖实体和催化活性亚类占比最高; 注释到生物学过程的 23 个亚类中的 Unigene 共有 27 897 条, 以细胞进程和代谢过程亚类占比最高, 注释数量分别为 9 399 条和 8 404 条, 而与节律过程、生物之间的种内相互作用、碳利用、移动等相关的基因表达较少。

2.5 KEGG 代谢通路分析

通过对冰草 RNA-seq 获得的 Unigene 进行 KEGG 代谢通路富集分析, 共注释到 14 412 条 Unigene, 占总 Unigene 的 56.27%, 涉及代谢通路 136 条, 富集显著的代谢通路有 27 条, Unigene 数目为 121~824 条, 其中数目最多的是植物病原

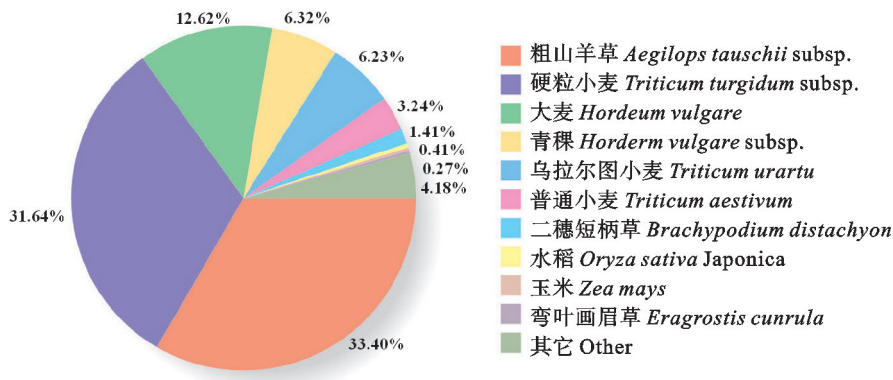


图 1 Nr 注释的 Unigenes 物种分布

Fig. 1 Species distribution based on Nr annotation of the Unigenes

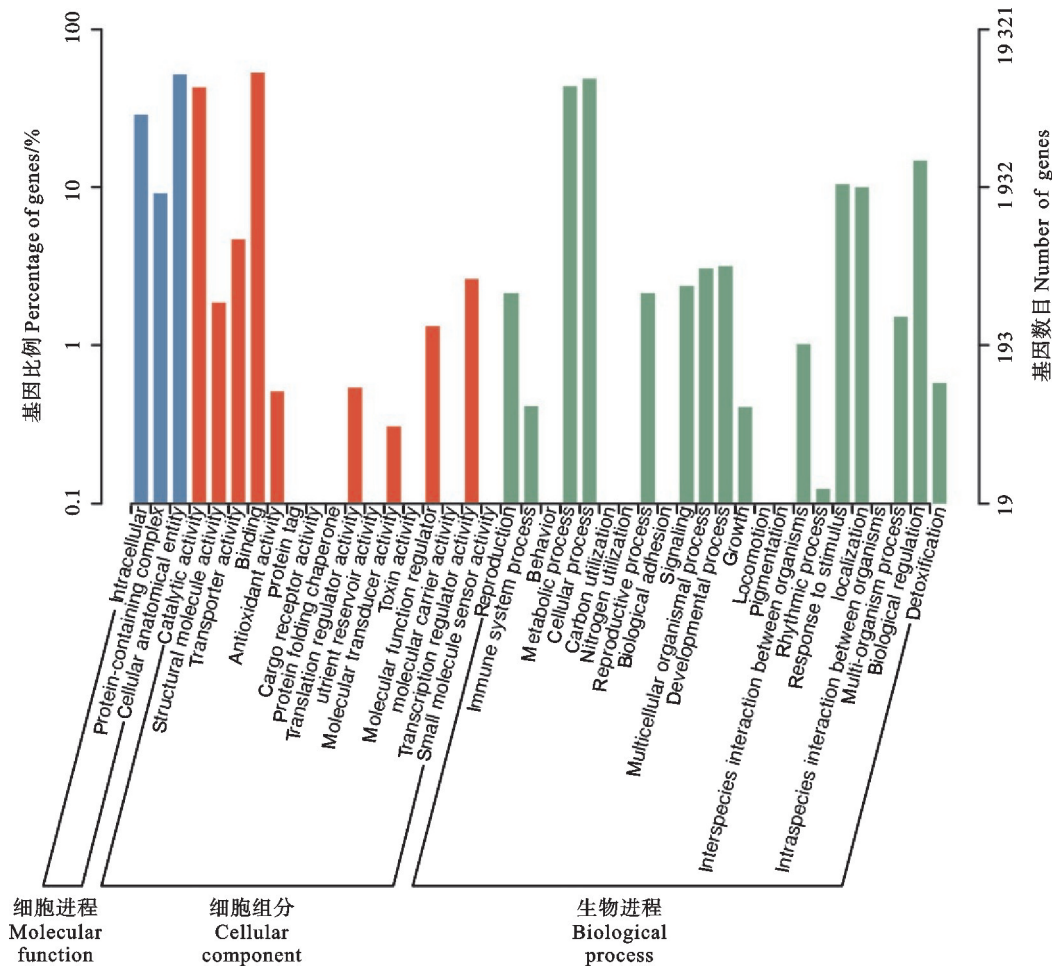


图 2 Unigenes 的 GO 功能分类

Fig. 2 GO function classification of the Unigenes

菌互作、碳代谢、植物激素信号转导和氨基酸的生物合成,分别包含 824、358、345 和 307 条 Unigene;内质网中的蛋白质加工和核糖体代谢通路,均富集 271 条 Unigene;糖酵解/糖异生和苯丙烷类物质生物合成代谢通路,均有 173 条 Unigene 参与;数目最少的是过氧化物酶体和光合生物中的碳固定代谢通路,别包含 128 和 121 条 Uni-

gene(表 3)。

2.6 KOG 注释分类

对冰草的 Unigene 进行 KOG 功能注释分类,共有 11 321 条 Unigene 获得注释,占总 Unigene 的 44.20%,分为 25 个功能类别(图 3)。其中,一般功能预测、蛋白质周转及信号转导机制 3 类的基因数量最多,分别为 1 995 条(17.62%)、

表 3 冰草转录组 Unigenes 的部分 KEGG 代谢通路
Table 3 Some KEGG metabolic pathways of the Unigenes in wheatgrass transcriptome

代谢通路编号 Pathway ID	代谢通路 Pathway	Unigene 数目 Unigene number
ko04626	植物病原菌互作 Plant-pathogen interaction	824
ko01200	碳代谢 Carbon metabolism	358
ko04075	植物激素信号转导 Plant hormone signal transduction	345
ko01230	氨基酸的生物合成 Biosynthesis of amino acids	307
ko03040	拼接体 Spliceosome	299
ko04016	MAPK 信号通路 MAPK signaling pathway	295
ko03013	RNA 转运 RNA transport	276
ko03010	核糖体 Ribosome	271
ko04120	泛素介导的蛋白水解 Ubiquitin mediated proteolysis	245
ko04144	内吞作用 Endocytosis	239
ko00500	淀粉和蔗糖代谢 Starch and sucrose metabolism	232
ko03018	RNA 降解 RNA degradation	183
ko00010	糖酵解/糖异生 Glycolysis/Gluconeogenesis	173
ko00940	苯丙烷类物质生物合成 Phenylpropanoid biosynthesis	173
ko03015	mRNA 监视通路 mRNA surveillance pathway	156
ko00564	甘油酯代谢 Glycerophospholipid metabolism	150
ko00190	氧化磷酸化 Oxidative phosphorylation	140
ko00195	光合作用 Photosynthesis	134
ko00230	嘌呤代谢 Purine metabolism	132
ko04146	过氧化物酶体 Peroxisome	128
ko04141	内质网中的蛋白质加工 Protein processing in endoplasmic reticulum	271
ko00520	氨基糖和核苷酸糖的代谢 Amino sugar and nucleotide sugar metabolism	168
ko03008	真核生物中的核糖体生物发生 Ribosome biogenesis in eukaryotes	149
ko00270	半胱氨酸和蛋氨酸代谢 Cysteine and methionine metabolism	142
ko00630	乙醛酸盐和二羧酸盐代谢 Glyoxylate and dicarboxylate metabolism	136
ko04070	磷脂酰肌醇信号系统 Phosphatidylinositol signaling system	132
ko00710	光合生物中的碳固定 Carbon fixation in photosynthetic organisms	121

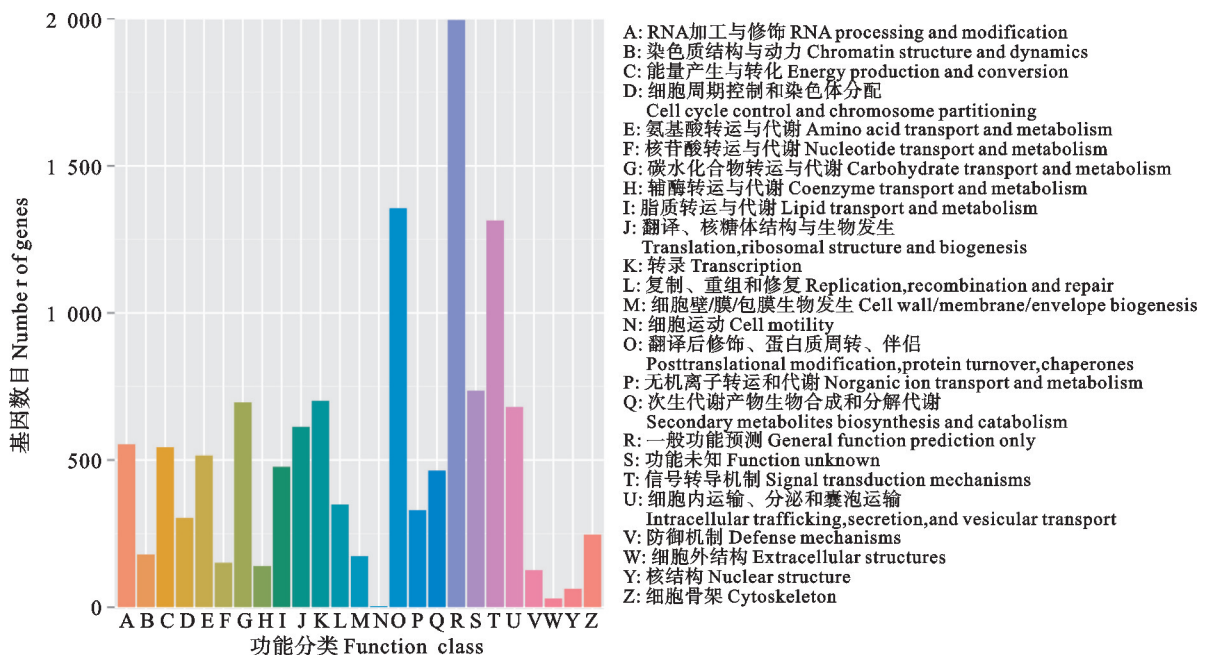


图 3 Unigenes 的 KOG 注释分类
Fig. 3 KOG classification of the Unigenes

1 356 条 (11.98%) 和 1 315 条 (11.62%) Unigene, 其次是转录 (701 条, 6.19%)、碳水化合物运输和代谢 (696 条, 6.15%)、细胞内运输、分泌和囊泡运输 (680 条, 6.01%)、翻译、核糖体结构和生物发生 (613 条, 5.41%)。染色质结构和动力学、核苷酸转运与代谢、辅酶转运与代谢、细胞壁/膜/包膜生物发生及防御机制匹配数目较少 (均小于 2%)。只有极少数 Unigene 被分配到细胞运动、胞外结构及细胞核结构, 且均小于 0.6%。因此, 冰草中信号转导、功能预测、翻译与修饰的基因最多, 氨基酸与碳水化合物的运输和代谢、能源

生产和转换等参与的基因次之, 核苷酸转运与代谢、细胞运动最少。

2.7 冰草 CDS 预测与 SSR 检测分析

共预测到大于 100 bp 的 CDS 有 7 288 条, 随序列长度的增加, 对应数量成降低趋势 (图 4)。其中长度为 100~200 bp 的数目最多, 为 1 144 条, 占总 CDS 的 15.70%, 其次是 200~300 bp 有 1 000 条, 占比 13.72%; 300~2 000 bp 之间, 各长度下的 CDS 数量均在 105 条以上, 且 2 900~3 000 bp 的数目最少 (19 条), 占比 0.26%。

对预测获得的 CDS 进行 SSR (简单序列重复)

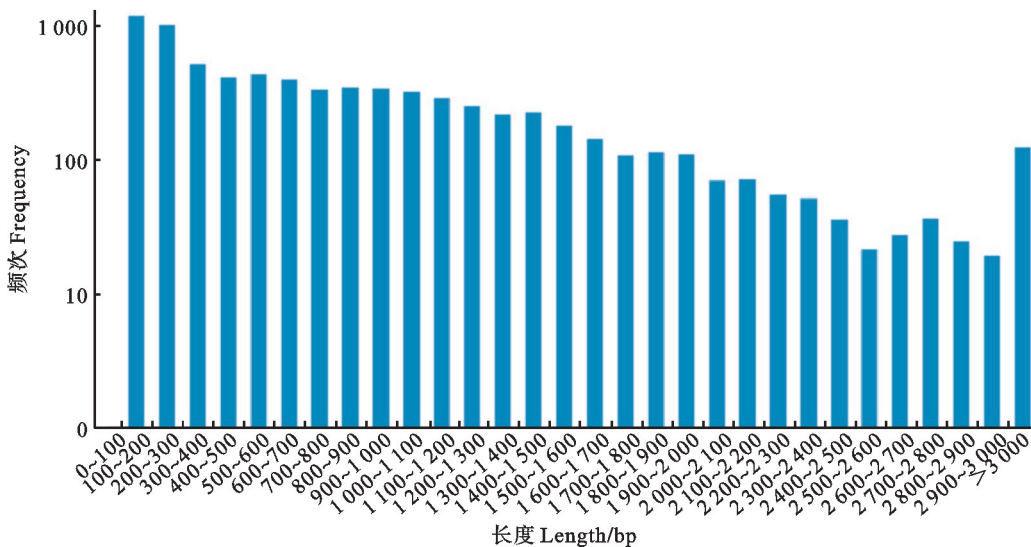


图 4 Unigenes 的 CDS 长度分布

Fig. 4 CDS length distribution of the Unigenes

检测, 共检测到 2 443 个, 分布于 8 种不同的核苷酸重复类型中 (表 4)。其中三碱基重复的数量最多, 为 1 439 个, 占总数的 58.90%, 其次是单碱基重复, 为 420 个 (17.19%), 其余为二碱基重复 (383 个, 15.68%)、混合 SSR (123 个, 5.03%), 且四碱基、五碱基、六碱基重复和以复合物存在的 SSR 最少, 分别为 59、12、5、2 个, 占 1.57%。单碱基重复共 4 种, SSR 频率最高的为 A/T, 最低的是 G/C; 二碱基重复共 12 种, SSR 频率最高的为 CT/AG、TC/GA、GA/TC 和 AG/CT, 最低的是 AT/AT; 三碱基重复共 57 种, CCG/CGG、CGC/GCG、CGG/CCG 和 GGC/GCC 的 SSR 频率最高, TTA/TAA、TGT/ACA、TAG/CTA、TAC/GTA 和 ATA/TAT 最低; 四碱基重复中, SSR 频率最高是 ACCT/AGGT、TAAA/TTTA、TC-TA/TAGA 和 TGCT/AGCA, ATGC/GCAT 和 AAAG/CTTT 最低。

表 4 冰草转录组 SSR 分析

Table 4 SSR analysis of wheatgrass transcriptome

SSR 种类 Type of SSR	数目 Number
以复合物存在的 SSR SSR in complex form	2
混合 SSR Mixed SSR	123
单碱基重复 Nucleotide	420
二碱基重复 Di-nucleotide	383
三碱基重复 Tri-nucleotide	1 439
四碱基重复 Tetra-nucleotide	59
五碱基重复 Penta-nucleotide	12
六碱基重复 Hexa-nucleotide	5
总计 Total	2 443

2.8 冰草 EST-SSR 标记开发及多态性引物筛选与验证

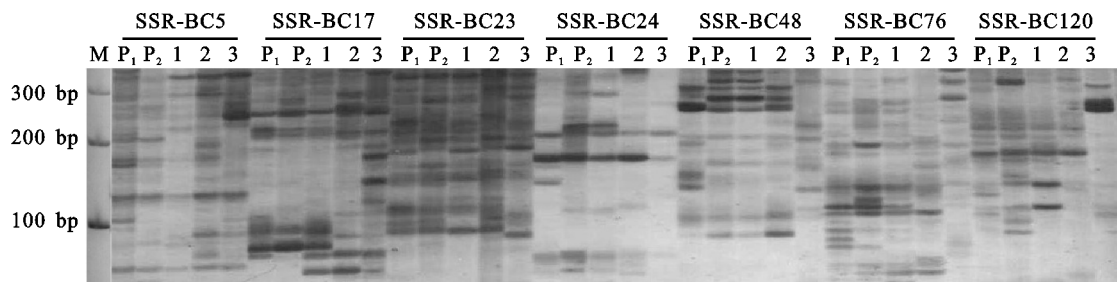
根据 SSR 检测结果, 二碱基和三碱基重复的多态性位点较丰富, 去掉 SSR 位点最前端或最后端序列, 将其导入 Primer 5.0 引物设计软件中,

开发出 158 对 EST-SSR 标记引物序列,从中筛选出 42 对多态性丰富、扩增条带清晰的适宜引物,并进行了命名编号(表 5、图 5),其中引物长度为 19~20 bp,预期产物片段介于 200~279 bp 之间。用随机选择的 10 对 EST-SSR 引物对冰草的

30 个 F₂ 代分离单株及其亲本进行 PCR 扩增,其多态性位点在 4~10 个之间,多态性比率为 35.71%~63.64%,平均多态性比率为 48.33%(图 6、表 6),表明这 10 对 EST-SSR 标记引物的多态性较好、可靠性较强。

表 5 冰草多态性 EST-SSR 引物信息
Table 5 Information of polymorphic EST-SSR primers

引物编号 Primer code	引物序列 Primer sequence		SSR 单元 SSR unit	预期产物 Expected product/bp
	正向 Forward	反向 Reverse		
SSR-BC1	CAGAGTCCTCGATTCCGAAC	CCTCACCAGACCAATCACCT	(GA)10	211
SSR-BC2	CCGTACATAGTTCGTCGCCT	TGTCCAGTGTCAATTGCAT	(CA)12	244
SSR-BC3	CTTCTCCAGCCACATGACG	CATCACCATTCTTCCCCATC	(GA)9	241
SSR-BC4	ATCAACATCCAAGGCCTCCT	AAACGGGTTTAGCTGGAGGT	(AC)13	273
SSR-BC5	CCATCACCCATTTCTTCTCCTC	TCTTCTCCTCCTCCTCCTC	(CT)12	226
SSR-BC6	ACCTGCAGATCAAATCCACC	AAGTGAAACCACAGCCAACC	(AG)11	243
SSR-BC7	GATCGTGTTCGAGTGGGTCT	CCCCTCCACTCTTGTGGTT	(TC)10	223
SSR-BC8	AGTTGCTTCCCTTCTCCAAA	CTTCCAATCCCTCTTCTCCTC	(AG)14	225
SSR-BC9	CGGTGACGTAGAGGAAGAGG	AACTTCTCAGAGCAGCAGCC	(CT)8	264
SSR-BC10	TCCGAATACAACACGCCAC	TGTCCCGCTTCACTTCTTCT	(AC)11	208
SSR-BC11	CGCTGCTTGATTAGGAGAGG	AGGTCTCTGCGACCAGAATG	(TC)12	264
SSR-BC12	TTGATGCAGCCTACACCGTA	TGATGAACAAGCCTTCATGG	(TG)7	232
SSR-BC13	GGCAAATGGCTGTGATCTT	CGCACCTTTCCCAACTCATA	(CA)9	261
SSR-BC14	TGACATCGCTGCTACCTGAC	TACAAACGAGGTGACACAGC	(GT)9	227
SSR-BC15	GCACCTTGAGGCAAACACTT	CAAGAAATCTGGCCAAGAGC	(TG)10	253
SSR-BC16	CCGTTCTCTATGCCAATAGA	GCACTGGAAGGGATGGTAAG	(AC)10	279
SSR-BC17	TTGTATCCCGCTCTCAAAG	TGACGACTGATGATCGGAAA	(GT)7	200
SSR-BC18	CCGTACCAACAACACTGTGTC	TTCAAATGGAGATGGCACAA	(AC)6	266
SSR-BC19	CTCGACCTTGAGTAGTTCGG	GGAGGACGAGGATGAGGAC	(CG)6	202
SSR-BC20	GAAGAGCACTCGAGGACCTG	GAAGATGTTCCGGTCCATGT	(GT)11	209
SSR-BC21	TGTCTCTCTCCTTCTCTCA	TTGCAGAGCCTGGTAGTTGA	(AT)6	279
SSR-BC22	AACAAGATTGGTCGATTGCC	CTGTATGCAGGGGAACATCA	(GA)9	244
SSR-BC23	TTTCCCCTACTCCTCTGCT	GCTGAAGTCTCGTAGAGGG	(TC)10	277
SSR-BC24	GCAGGCAAGGAGAAAAGACAC	TACCAATGAACACCACCACC	(GT)12	248
SSR-BC25	CTCTCCCTCTCTCTCGTT	TATTGACTCCGCGAGAGGTT	(CT)8	210
SSR-BC26	CCTTGAACCTTCTCCTCGTCG	CTCTCTCTCCCTCTGCATCC	(GA)8	268
SSR-BC27	AAGGTGAGGACCGTCAAGG	ATGGACGTCATGTTGGAGGT	(CCA)7	257
SSR-BC28	TTCGAAGCTGAGTCCACCAT	CACAGCCAACCTCCAAGATT	(CTC)7	274
SSR-BC29	CCTTTTGACCTGTGCTGTGA	ATAGCATGCCAAGCACTCT	(TGC)7	223
SSR-BC30	AGCGTTCTGCACGTAATCCT	TCCGGTCTACCATTATCCA	(GGC)6	207
SSR-BC31	TCGCATCTCAGTTCCAATCC	GCCTATTTGTTCCGGTTGTC	(CCG)5	279
SSR-BC32	GACGATGACGCAGATGAAGA	AAGATACACCTCGCTTGCGT	(GAT)5	262
SSR-BC33	CTACCTCTTCTCTGCGCC	ACTCGACGTGGAACGAGTG	(CGG)5	272
SSR-BC34	GACGATTCACACCGGAATCT	GACAAACGACAGCCACAACA	(GTG)5	263
SSR-BC35	GATCTCTACCTCCGCCTTCC	AAGACTGACTGTTGGGGTGG	(CCG)7	266
SSR-BC36	ATATCCAATTCATCCGCAA	CTACCCCTACCACCCTCC	(CGG)7	215
SSR-BC37	AAACGGGTCACTATCGTGG	AAAAGATGGAGGTGTCCGC	(CGC)7	263
SSR-BC38	CTGATAGATGGCGACCTCGT	GGGGTCTCCGGTCTTAACTC	(GGC)6	260
SSR-BC39	CTCTCCTCGCTCTCAGCTA	ACTATGGTTCGAGGACCCTT	(TTC)6	270
SSR-BC40	AAAAACAGAGGCTTCAGGCA	CTTCCAGAACTGAGGCACC	(CGA)7	246
SSR-BC41	CAGCTACCAGCAATGCAGAA	ATACTGGGTGGAGGTGATGG	(TGC)8	274
SSR-BC42	AAAATCAAACCCAGTGGCTG	GAGGAGGAGGAGAGCGAGAG	(GCC)6	213

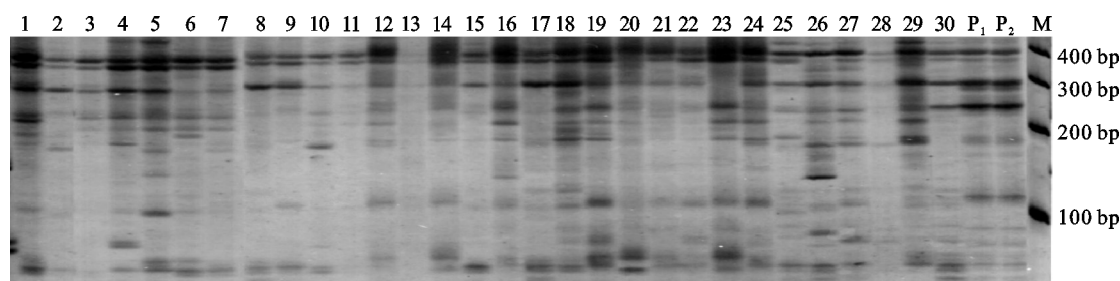


M: 100 bp marker; P₁: 蒙古冰草; P₂: 航道冰草; 1~3: 3 个 F₂ 单株。下同。

M: 100 bp marker; P₁: *A. mongolicum* Keng; P₂: *A. cristatum* cv. Fairway; 1-3: Three F₂ individuals. The same in Fig. 6.

图 5 蒙古冰草×航道冰草及其 3 个杂交 F₂ 单株对部分 EST-SSR 引物的筛选结果

Fig. 5 Screening of partial EST-SSR primers using three F₂ individuals derived from *A. mongolicum* Keng×*A. cristatum* cv. Fairway



1~30: 30 个 F₂ 单株。1-30: 30 F₂ individuals.

图 6 引物 SSR-BC8 对杂交 F₂ 代 30 个单株及其双亲的 EST-SSR 扩增结果

Fig. 6 Amplicons of EST-SSR primer SSR-BC8 from 30 F₂ individuals and their parents

表 6 SSR 多态性引物扩增结果

Table 6 Amplification results of partial EST-SSR polymorphic primers

引物名 Primer name	扩增位点总数 Number of amplified loci	多态性位点数 Number of polymorphic loci	多态性位点比率 Percentage of the polymorphic loci/%
SSR-BC1	13	7	53.85
SSR-BC4	15	6	40.00
SSR-BC5	12	5	41.67
SSR-BC8	11	7	63.64
SSR-BC11	16	8	50.00
SSR-BC13	9	4	44.44
SSR-BC14	13	5	38.46
SSR-BC24	10	6	60.00
SSR-BC25	14	5	35.71
SSR-BC36	18	10	55.56
平均 Average	13.1	6.3	48.33

3 讨论

随着分子生物学的迅猛发展,高通量测序技术逐渐进入人们的视野,越来越多的科研人员开始运用第二代、三代测序技术探索更多的生物学现象。RNA-seq 是基于 RNA 分子水平上的测序

技术,其具有高效快速、价格低廉、准确的特点,且不依赖于物种的全基因组信息就能够获得转录本序列的海量信息,已被广泛应用于诸多植物的分子标记开发、差异基因表达分析等领域^[17-21]。SSR 作为一种重要的分子标记,具有共显性遗传、位点稳定、多态性高、带型简单易读等优点,其来源主要有基因组 SSR 和表达序列标签(EST)SSR 这 2 类,其中基因组 SSR 广泛存在于植物的整个基因组中;而 EST-SSR 仅存在于基因的某一表达区域,高通量测序是获取该标记的最佳途径,有研究表明其在近缘种间的通用性较好,这可能与功能基因存在密切关系^[22],因此基于转录组数据开发的 EST-SSR 标记有助于与重要产量、品质等性状进行关联分析,进而提高育种目标性状的定向选择效率。

目前,有关牧草的基因组信息挖掘、EST-SSR 标记开发等研究报道较多,叶远俊等^[23]通过对萱草进行转录组测序,共鉴定出 7 335 个 EST-SSR 位点,其中三碱基重复的数量占总位点的 54.4%,这与本研究中三碱基重复数占比最大的结果相一致。孙禄娟等^[24]利用盐生草全长转录

组数据对其进行 SSR 位点检测,共获得 SSR 位点 327 种、29 640 个,其中三碱基重复类型 ATC/GAT 和 AAG/CTT 占比较高,而本试验高占比的三碱基重复类型为 CCG/CGG、CGC/GCG、CGG/CCG 和 GGC/GCC。逢洪波等^[25]对鬼针草转录组数据进行组装,共得到总长 106 457 436 bp 的 Unigenes 为 120 122 条,检测到 SSR 位点 17 140 个,其数量远高于本研究获得的 2 443 个冰草 EST-SSR 位点,这可能与物种间的基因组大小存在差异有密切联系。本研究对无参考基因组的冰草进行 RNA-seq 分析,共获得 33 993 条 Unigene,平均长度 862.24 bp,其中有 25 614 条 Unigene 成功注释到 GO、Nr、KEGG 等 9 大功能数据库中,占总 Unigenes 的 75.35%,这些被注释的 Unigenes 为冰草属植物的代谢通路分析、目的基因表达及植物信号传导等研究提供了参考依据。而未得到功能注释的 Unigenes 有 8 379 条,可能是当前公共基因数据库信息尚不完善或冰草存在某些特异性基因与其无法比对。此外,通过 SSR 检测,本试验最终获得 2 443 个 SSR 标记位点,分布于 8 种不同的核苷酸重复类型中,说明二倍体冰草的基因位点较丰富。其中单核苷酸中 A/T 重复最多,共 213 个,这与盐生草^[24]、苦荞^[26]、金银花^[27]、山桐子^[28]等植物的研究结果一致;二核苷酸重复最多的为 AG/CT、CT/AG,这与小麦^[29]、老芒麦^[30]、天麻^[31]等植物相同;三核苷酸重复以 CCG/CGG 最多,这与 Morgante 等^[29]在小麦、水稻、玉米等单子叶植物中 SSR 位点分析结果一致。试验以 3 个杂交冰草 F₂ 分离单株及双亲的 DNA 为模板,对新设计的 158 对 EST-SSR 标记进行引物筛选,最终获得多态性丰富、扩增条带清晰的适宜引物 42 对,其余 116 对引物没有稳定的扩增产物,这可能是 Unigenes 拼接不完整或出现误差所导致^[32],后续可通过冰草与小麦的 EST 序列进行比对,设计出多态性强、特异性高的冰草特异序列,用于小麦族植物分子遗传和进化研究中。

本试验利用 RNA-seq 技术对 2 个冰草的各组织样本进行高通量测序分析,获得了高质量的序列信息,进而基于生物信息学对冰草 Unigene 进行功能注释、代谢通路分析、EST-SSR 标记位点开发及多态性引物验证试验,其结果可为后续开展四倍体冰草基因组功能分析、种质资源创新及标记辅助育种等研究提供参考依据。

参考文献:

- [1] 云锦凤,米福贵.冰草属牧草的种类与分布[J].中国草地,1989,11(3):14.
YUN J F, MI F G. Species and distribution of *Agropyron* grass [J]. *Grassland of China*, 1989, 11(3): 14.
- [2] 康如龙,刘万龙,许冬梅,等.荒漠草原蒙古冰草种群分布格局及种内关联性[J].生态学报,2024,44(21):9910.
KANG R L, LIU W L, XU D M, et al. The distribution pattern and intraspecific associations of *Agropyron mongolicum* populations in desert steppe [J]. *Acta Ecologica Sinica*, 2024, 44(21): 9910.
- [3] 李鸿雁,李悦焯,李俊,等.内蒙古 143 份冰草属种质资源表型多样性分析与综合评价[J].植物遗传资源学报,2024,25(8):1254.
LI H Y, LI Y X, LI J, et al. Phenotypic diversity analysis and comprehensive evaluation of 143 *Agropyron* germplasm resources in Inner Mongolia, China [J]. *Journal of Plant Genetic Resources*, 2024, 25(8): 1254.
- [4] 耿以礼.中国主要植物图说——禾本科[M].北京:科学出版社,1959:65.
GENG Y L. Illustration of main plants in China——Gramineae [M]. Beijing: Science Press, 1959: 65.
- [5] 徐娜.根茎冰草新品系生物学性状的研究[D].呼和浩特:内蒙古农业大学,2011:9.
XU N. Study on the biological characteristics in a new strain of *Agropyron michnoi* Roshev [D]. Hohhot: Inner Mongolia Agricultural University, 2011: 9.
- [6] 杨东升,于卓,于肖夏,等.四倍体杂交冰草 SRAP 和 SSR 分子标记遗传连锁图谱构建[J].麦类作物学报,2018,38(8):941.
YANG D S, YU Z, YU X X, et al. Construction of the genetic linkage map of tetraploid hybrid crested wheatgrass based on SRAP and SSR markers [J]. *Journal of Triticeae Crops*, 2018, 38(8): 941.
- [7] 杨东升,于卓,于肖夏,等.冰草蛋白质含量等 4 个重要性状的 QTL 定位研究[J].麦类作物学报,2019,39(7):775.
YANG D S, YU Z, YU X X, et al. QTL mapping for four quality and yield related traits in wheatgrass [J]. *Journal of Triticeae Crops*, 2019, 39(7): 775.
- [8] 王棕瑶,徐世锐,吉夏洁,等.小麦-冰草不同 2P 附加系的特征分析[J].植物遗传资源学报,2023,24(4):944.
WANG Z Y, XU S R, JI X J, et al. Characterization of different 2P wheat-*Agropyron cristatum* addition lines [J]. *Journal of Plant Genetic Resources*, 2023, 24(4): 944.
- [9] 常怡然,史佳梅,许冬梅,等.荒漠草原不同自然种群蒙古冰草生物量和养分权衡特征[J].草业学报,2024,33(11):186.
CHANG Y R, SHI J M, XU D M, et al. Trade-off relationships between biomass and nutrient allocation in different natural populations of *Agropyron mongolicum* on the desert steppe [J]. *Acta Prataculturae Sinica*, 2024, 33(11): 186.
- [10] 杨靖,张晓明,陈越,等.内蒙古 30 份冰草属种质资源表型性状遗传多样性分析[J].中国草地学报,2023,45(9):1.
YANG J, ZHANG X M, CHEN Y, et al. Genetic diversity analysis of phenotypic traits of 30 *Agropyron* genus germplasm resources in Inner Mongolia [J]. *Chinese Journal of Grassland*, 2023, 45(9): 1.
- [11] 杨东升,于卓,于肖夏,等.基于 GBS 技术构建四倍体杂交冰草超高密度遗传连锁图谱[J].麦类作物学报,2021,41(10):1197.
YANG D S, YU Z, YU X X, et al. Construction of ultra-high density genetic linkage map for tetraploid hybrid crested

- wheatgrass based on GBS technology [J]. *Journal of Triticeae Crops*, 2021, 41(10):1197.
- [12]樊璐,唐俊,马琳,等. 基于转录组测序筛选蒙古冰草抗旱相关基因[J]. 草地学报, 2024, 32(11):3344.
FAN L, TANG J, MA L, *et al.* Screening genes related to drought resistance in *Agropyron mongolicum* Keng based on transcriptome [J]. *Acta Agrestia Sinica*, 2024, 32(11):3344.
- [13]王贺萍,孙震,刘雨辰,等. 蒙古冰草肉桂醇脱氢酶基因序列鉴定及功能分析[J]. 植物学报, 2024, 59(2):204.
WANG H P, SUN Z, LIU Y C, *et al.* Sequence identification and functional analysis of cinnamyl alcohol dehydrogenase gene from *Agropyron mongolicum* [J]. *Chinese Bulletin of Botany*, 2024, 59(2):204.
- [14]黄馨田,韩家帆,杜锦瑜,等. 根茎冰草 *AmrbHLH137* 基因克隆及表达分析[J/OL]. 分子植物育种, <https://link.cnki.net/urlid/46.1068.S.20240226.1118.002>.
HUANG X T, HAN J F, DU J Y, *et al.* Cloning and expression analysis of *AmrbHLH137* gene from *Agropyron michnoi* Roshev [J/OL]. *Molecular Plant Breeding*, <https://link.cnki.net/urlid/46.1068.S.20240226.1118.002>.
- [15]闫秀秀,范波波,张学峰,等. 蒙古冰草 GDSL 酯酶/脂肪酶基因的序列及表达分析[J]. 中国农业大学学报, 2023, 28(4):41.
YAN X X, FAN B B, ZHANG X F, *et al.* Sequence and expression analysis of GDSL esterase/lipase gene in *Agropyron mongolicum* [J]. *Journal of China Agricultural University*, 2023, 28(4):41.
- [16]于肖夏,杨东升,于卓,等. 冰草 SSR-PCR 体系优化及验证[J]. 草业科学, 2020, 37(9):1760.
YU X X, YANG D S, YU Z, *et al.* Optimization and validation of a simple sequence repeat (SSR)-PCR system in wheatgrass [J]. *Pratacultural Science*, 2020, 37(9):1760.
- [17]TOPU M, SESIZ U, BEKTAŞ H, *et al.* Next-generation-sequencing-based simple sequence repeat(SSR) marker development and linkage mapping in lentil (*Lens culinaris* L.) [J]. *Life*, 2023, 13(7):1579.
- [18]PAN G, XIE J, QIN Y, *et al.* Development of SSR markers for genetic diversity analysis and species identification in *Polygonatum odoratum* (Mill.) Druce based on transcriptome sequences [J]. *PLoS One*, 2024, 19(9):e0308316.
- [19]SAGAR T, KAPOOR N, MAHAJAN R. Development of genomic SSR markers for characterization of genetic diversity in wild-pomegranate germplasm [J]. *Genetic Resources and Crop Evolution*, 2024, 71(4):1401.
- [20]洪森荣,刘佳凝,袁昕,等. 苏丹草和高丹草转录组测序及其差异基因表达分析[J]. 草地学报, 2024, 32(3):714.
HONG S, LIU J, YUAN X, *et al.* Transcriptome sequencing and their differential gene expression analysis of *Sorghum sudanense* (Piper) Stapf. and *Sorghum bicolor* (Linn.) Moench. × *Sorghum sudanense* (Piper) Stapf [J]. *Acta Agrestia Sinica*, 2024, 32(3):714.
- [21]吴林,杨苏渝,杨阳,等. 基于转录组测序挖掘马铃薯块茎采后发芽关键基因[J]. 西南大学学报(自然科学版), 2023, 45(11):70.
WU L, YANG S Y, YANG Y, *et al.* Identification of critical genes involved in postharvest sprouting of potato tubers based on RNA-seq analysis [J]. *Journal of Southwest University(Natural Science Edition)*, 2023, 45(11):70.
- [22]李永平,陈敏氢,康建坂,等. 辣椒 EST-SSRs 标记的通用性研究[J]. 福建农业学报, 2017, 32(7):724.
LI Y P, CHEN M D, KANG J B, *et al.* Cross-species applicability of EST-SSRs from *Capsicum annuum* L. [J]. *Fujian Journal of Agricultural Sciences*, 2017, 32(7):724.
- [23]叶远俊,周熠玮,谭健俊,等. 萱草 EST-SSR 分子标记开发及开花性状的连锁分析[J/OL]. 分子植物育种, <http://link.cnki.net/urlid/46.1068.S.20231009.1127.004>.
YE Y J, ZHOU Y W, TAN J J, *et al.* EST-SSR markers development and association analysis of flowering traits in *Hemerocallis* accessions [J/OL]. *Molecular Plant Breeding*, <http://link.cnki.net/urlid/46.1068.S.20231009.1127.004>.
- [24]孙禄娟,何建军,汪军成,等. 基于全长转录组测序的盐生草 SSR 标记开发及其遗传多样性分析[J]. 草业学报, 2022, 31(8):199.
SUN L J, HE J J, WANG J C, *et al.* Development of SSR markers based on full-length transcriptome sequencing and genetic diversity analysis of *Halogeton glomeratus* [J]. *Acta Pratacultural Sinica*, 2022, 31(8):199.
- [25]逢洪波,解元坤,张璐,等. 基于转录组测序的鬼针草 SSR 标记开发及其应用[J]. 分子植物育种, 2018, 16(16):5359.
PANG H B, XIE Y K, ZHANG L, *et al.* Development and application of SSR markers in *Bidens bipinnata* L. based on transcriptome sequencing [J]. *Molecular Plant Breeding*, 2018, 16(16):5359.
- [26]HOU S Y, SUN Z X, LINGHU B, *et al.* Genetic diversity of buckwheat cultivars (*Fagopyrum tartaricum* Gaertn.) assessed with SSR markers developed from genome survey sequences [J]. *Plant Molecular Biology Reporter*, 2016, 34(1):233.
- [27]李慧,刘东超,徐瑞瑞,等. 基于 RAD-seq 技术的金银花 SSR 标记开发及鉴定[J]. 北京林业大学学报, 2021, 43(6):108.
LI H, LIU D C, XU R R, *et al.* Development and identification of SSR markers based on RAD-seq of *Lonicera japonica* [J]. *Journal of Beijing Forestry University*, 2021, 43(6):108.
- [28]邱文明,严莉,全铸,等. 基于 SLAF-seq 技术的山桐子 SSR 标记开发及应用[J]. 分子植物育种, 2024, 22(12):3929.
QIU W M, YAN L, TONG Z, *et al.* Development and application of SSR markers in *Idesia polycarpa* based on SLAF-seq technology [J]. *Molecular Plant Breeding*, 2024, 22(12):3929.
- [29]MORGANTE M, HANAFEY M, POWELL W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. *Nature Genetics*, 2002, 30(2):194.
- [30]ZHANG Z Y, XIE W G, ZHAO Y Q, *et al.* EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species [J]. *BMC Plant Biology*, 2019, 19(1):235.
- [31]WANG Y S, SHAHID M Q, GHOURI F, *et al.* Development of EST-based SSR and SNP markers in *Gastrodia elata* (herbal medicine) by sequencing, *de novo* assembly and annotation of the transcriptome [J]. *3 Biotech*, 2019, 9(8):292.
- [32]崔雨,鲍印广,王洪刚,等. 基于 RNA-seq 技术开发中间偃麦草基因组特异分子标记[J]. 麦类作物学报, 2016, 36(6):699.
CUI Y, BAO Y G, WANG H G, *et al.* Development of specific molecular markers for *Thinopyrum intermedium* using RNA-seq data [J]. *Journal of Triticeae Crops*, 2016, 36(6):699.