

· 研究论文 ·

基于流式细胞术和基因组Survey的黄缨菊基因组大小及特征分析

靳佳瑞^{1,2}, 刘玉萍^{1,2,3}, 苏旭^{1,2,3*}, 刘涛^{4*}, 余明君^{1,2}, 杨倩^{1,2}, 曲荣举^{1,2}
张朋辉^{1,2}, 才让扎西^{1,2}, 南措加¹, 周乐怡¹

¹青海师范大学生命科学学院, 西宁 810008; ²青海师范大学青海省青藏高原生物多样性形成机制与综合利用重点实验室, 西宁 810008; ³青海师范大学高原科学与可持续发展研究院, 西宁 810016; ⁴青海理工学院生态与环境科学学院, 西宁 810016

摘要 黄缨菊(*Xanthopappus subacaulis*)是菊科(Asteraceae)黄缨菊属(*Xanthopappus*)多年生高原特有单属种药用植物, 具有重要的经济、生态和药用价值。为确定适合黄缨菊全基因组测序的技术策略, 利用流式细胞术和基因组Survey分析评估黄缨菊基因组大小、杂合率、重复序列比例、GC含量和长末端重复反转录转座子(long terminal repeat retrotransposons, LTR-RTs)等信息。结果表明: (1) 以长裂太行菊(*Opisthopappus longilobus*)和番茄(*Solanum lycopersicum*)为参考物种, 采用流式细胞术预估黄缨菊为二倍体, 基因组大小分别为1.94 G和1.75 G, DNA-C值为0.99 pg; (2) 高通量测序得到约100.3 G的Clean reads, 其Q20均大于97.1%, Q30均高于90.8%, AT和GC碱基含量无明显分离, GC含量为38.5%, 测序质量良好; (3) K核苷酸序列分析(K-mer analysis, K-mer)显示黄缨菊基因组大小为2 198.50 Mb, 杂合度为0.69%, 重复序列占比为80.15%, 属于微杂合、高重复序列的复杂基因组; (4) LTR-RTs鉴定表明Copia家族数量最多, 占全基因组的30.72%, Gypsy家族和Unknown分别占全基因组的33.66%和16.54%, 插入时间始于约3 Mya, 在1 Mya内产生大量扩增。综上, 研究结果表明LTR的大量插入是导致黄缨菊基因组复杂化的重要原因之一; 测序数据和研究结果可为黄缨菊高质量基因组遗传图谱构建和关键功能基因挖掘提供重要参考。

关键词 黄缨菊, 流式细胞术, C值, 基因组Survey, 长末端重复反转录转座子(LTR-RTs)

靳佳瑞, 刘玉萍, 苏旭, 刘涛, 余明君, 杨倩, 曲荣举, 张朋辉, 才让扎西, 南措加, 周乐怡 (2025). 基于流式细胞术和基因组Survey的黄缨菊基因组大小及特征分析. 植物学报 60, 888–900.

黄缨菊(*Xanthopappus subacaulis*)是菊科(Asteraceae)、管状花亚科(Carduoideae)、菜蓟族(Cynareae)、飞廉亚族(Carduinae)、黄缨菊属(*Xanthopappus*)多年生无茎矮小草本植物, 也是青藏高原的特有单属种, 主要分布于我国青海、云南西北部、四川北部和甘肃东南部等地, 通常生长于海拔2 400–4 000 m的高寒草甸、草原及干旱山坡, 具有较强的耐寒、耐旱和耐盐碱等特性(史惠兰, 2011)。黄缨菊全草入药, 味苦、微寒, 主治出血、食物中毒等疾病, 具有重要的药用价值; 同时, 黄缨菊富含黄酮类和噻吩类化合物, 具有较好的止血功效和良好的抗菌功能(张丽,

2014)。迄今为止, 国内外诸多学者对黄缨菊的研究主要集中于其外部形态特征(马子兰等, 2020)、化学成分(田永清等, 2008)、药理活性(Zhang et al., 2014)、耐逆生理、系统发育(Wang et al., 2007)、群体遗传结构(张阳等, 2022)、简单序列重复(simple sequence repeats, SSR)引物开发与遗传多样性(郑长远, 2023)等领域。然而, 目前尚未见关于黄缨菊基因组大小及其特征的研究报道。先前研究认为, 物种基因组大小及其特征分析是进行全基因组*de novo*基因组测序、组装和注释的前提和基础(黄阿晶等, 2019)。例如, 对山茱萸(*Cornus officinalis*)的基因组

收稿日期: 2024-10-23; 接受日期: 2025-01-20

基金项目: 青海理工大学(筹)“昆仑英才”人才引进科研项目(No.2023-QLGKLYCZX-012)、青海省重大科技专项(No.2023-SF-A5)、国家自然科学基金(No.32360305)、2023年中央林业草原生态保护恢复资金野生动植物保护项目(No.QHSY-2023-016)和青海省省财政林业改革发展资金林草新技术推广项目(No.QSCZ-2023-001)

* 通讯作者。E-mail: xusu8527972@126.com; 532226527@qq.com

进行测序、组装和注释,并结合染色体核型数据构建了高质量的基因遗传图谱,为山茱萸群体进化历史和物种多样性研究提供了基础数据(Dong et al., 2022); Kong等(2023)通过对菊科11个代表种全基因组测序数据的比较分析,构建了系统发育树和古六倍体两步法模型,为菊科物种多倍化及祖先基因组重建和进化提供了新的组学佐证。因此,开展黄缨菊基因组大小及其特征分析,不仅可为其全基因组 *de novo* 测序和组装提供理论依据,还可为高质量遗传图谱构建提供基础数据,结合细胞学和物种繁育体系等研究,可为黄缨菊适应性进化机制探究和类黄酮化合物合成关键基因挖掘奠定理论基础。

目前,流式细胞术(flow cytometry, FCM)和基因组 Survey 是研究基因组大小和特征的有效手段(邓颖珂等, 2023)。基因组大小一般用C值表示,是指1个物种单倍体配子基因组中所有DNA的含量,且具有特异性,即每个物种细胞的染色体数目和DNA含量相对固定(Mishiba et al., 2000);大量研究还表明,C值与植物进化以及对生态环境的适应均有不同程度的联系(Bancheva and Greilhuber, 2006; 郭水良等, 2015)。流式细胞术基于细胞核内DNA与荧光染料结合后,其含量与发射荧光强度呈正比的原理,利用已知染色体倍性与基因组大小的内标作为参照,通过计算待测物种的染色体倍性和DNA相对含量来预估其基因组的相对大小(Fomicheva and Domblides, 2023)。虽然流式细胞术已成为植物科学研究中的重要工具,但由于自身细胞结构和实验材料状态等因素,操作过程仍存在局限(张晋丹和冯旻, 2023; 夏春皎等, 2024)。

随着DNA测序技术的迅速发展及测序成本的降低(Zhang et al., 2021),高通量测序为基因组学、转录组学和进化生物学等领域的研究提供了一种有效的技术手段(Goodwin et al., 2016),已广泛应用于基因组大小测定(马鹏举等, 2018)、注释组装(Morabito et al., 2020)和微卫星特征分析(杨尉等, 2023)等。先前的研究表明, Survey 分析可有效评估物种基因组大小、杂合度和重复序列比例等基因组特征,为后续全基因组 *de novo* 测序、组装和结构注释等提供重要参考依据,以减少测序的盲目性(Zhou et al., 2023)。国内外诸多学者利用这种方法成功评估了蓝花苜蓿(*Medi-*

cago sativa) (Li et al., 2020)以及沙鞭(*Psammochloa villosa*) (刘涛等, 2022)等饲草植物,地黄(*Rehmannia glutinosa*) (赵乐等, 2021)和川穹(*Ligusticum chuanyang*) (毛常清等, 2023)等药用植物的基因组大小。菊科植物作为双子叶植物的第1大科,具有重要的经济、观赏和药用价值(梁凤萍等, 2018)。迄今为止,许多学者对多种菊科植物(如刺苞菜蓟(*Cynara cardunculus*) (Scaglione et al., 2016)、万寿菊(*Tagetes erecta*) (Xin et al., 2023)和牛蒡(*Arctium lappa*) (Yang et al., 2022))进行了全基因组测序和染色体水平的高质量遗传图谱构建,这些物种绝大多数属于复杂基因组。转座子(transposable element, TE)的插入是植物进化过程中驱动基因组扩增和复杂化的主要机制(Mascher et al., 2017)。Li等(2023)比较了小麦族(Triticeae) 10个不同物种的基因组,发现转座子含量与基因组大小呈正相关($R^2=0.979$)。由于整合方式不同,转座子包括反转录转座子和DNA转座子,反转录转座子通过“复制-粘贴”的转座机制实现生物学功能(Finnegan, 1989)。其中,完整的长末端重复反转录转座子(long terminal repeat retrotransposons, LTR-RTs)在宿主基因组中可实现复制和扩增。Staton和Burke (2015)调查了14种菊科植物和1种外类群的TE系统发育动态演变模式,其外类群到菊科衍生谱系的向日葵族(Trib. Heliantheae), LTR-RTs主要成员Gypsy和Copia的占比呈线性增加,表明转座子的多样性和丰度在驱动菊科植物基因组扩增和系统发育演变过程中具有重要作用。据此,本研究利用流式细胞术与高通量测序和基因组 Survey 相结合的方法综合评估和分析黄缨菊的基因组大小与特征,旨在为黄缨菊全基因组测序、组装、注释及遗传图谱构建提供基础数据,并为挖掘类黄酮化合物合成关键基因提供参考依据。

1 材料与方法

1.1 种质资源采集

自2019年以来,项目组全面调查并采集了青藏高原及其毗邻地区黄缨菊(*Xanthopappus subacaulis* C. winkl.)不同居群的凭证标本、叶片材料和成熟种子,同时详细记录其经纬度、海拔高度、生境特征和伴生

物种等信息。本研究所用黄缨菊实验材料采自青海省门源县仙米乡(37°15'32"N, 102°0'57"E, alt. 2 702.9 m)。野外选取健康且长势良好的黄缨菊3株个体, 分别剪取其幼嫩叶片, 经消毒和清洗后置于冻存管备用。一部分样品用湿润滤纸包裹, 用于流式细胞术检测, 另一部分样品迅速放入液氮中速冻, 用于基因组Survey分析。流式细胞术所用参照物种为近缘种长裂太行菊(*Opisthopappus longilobus* C. Shih, 2.4 G)和已知基因组大小的番茄(*Solanum lycopersicum* L., 799.09 Mb), 分别由山西省太原市植物园和金迪未来生物科技有限公司提供。

1.2 流式细胞术估测基因组大小

称取待测黄缨菊新鲜叶片0.2 g, 将其置于培养皿中, 用锋利刀片一次性快速切碎, 于样品周围加入500 μ L细胞核裂解液提取60秒, 保证充分提取完整的细胞核; 采用50 μ m滤网过滤混液至样品管中, 随后加入2 000 μ L含RNA酶的荧光染液, 避光染色15分钟后, 用CyFlow Cube 6流式细胞仪上机检测。每个样品设3次重复。同样, 按照上述步骤检测长裂太行菊和番茄的荧光强度。以相对荧光强度为横坐标, 细胞核数量为纵坐标, 分别绘制黄缨菊、长裂太行菊和番茄的荧光信号图; 根据长裂太行菊和番茄的荧光峰值, 利用基因组大小预估公式来估测黄缨菊基因组的相对大小并计算C值。

待测样品基因组大小(Mb)=(待测样品荧光强度/对照样品荧光强度) \times 内参植物基因组大小(Mb); DNA-C值(pg)=待测样品基因组大小(Mb)/978 \times 倍性。

1.3 基因组DNA提取

利用植物基因组DNA提取试剂盒(Cat No.DP305, TIANGEN)提取黄缨菊新鲜叶片总DNA; 采用超微量核酸浓度测定仪(NanoDrop One)测定DNA浓度($\text{ng}\cdot\mu\text{L}^{-1}$)和纯度($\text{OD}_{260}/\text{OD}_{280}$), 并经1%琼脂糖凝胶电泳检测其完整性。

1.4 基因组测序与质控

运用Covaris超声波破碎仪将质检合格的DNA随机打断为300–500 bp的片段, 经末端修复、加“A”尾和测序接头、纯化以及PCR扩增等步骤完成整个文库制备; 采用BGI T7测序平台对构建好的文库进行PE150

双末端测序。为去除因PCR扩增等引起的Duplicated reads、带接头的Paired reads、N比例大于0.5%及低质量Paired reads, 本研究利用SOAPnuke (Chen et al., 2018)软件对测序获得的原始序列(raw reads)进行过滤得到Clean reads。主参数设置为-lowQual=20, -nRate=0.005, -qualRate=0.5, 其它参数为默认设置。此外, 我们采用FastQC软件从测序数据质量、测序错误率分布、碱基频率分布及GC含量分布4个方面评估黄缨菊的测序质量; 并从中随机抽取10 000对reads数据, 通过Blast软件比对到NCBI核苷酸数据库(NT数据库), 判断是否有外源污染。

1.5 K-mer分析

采用K=17的K-mer方法分析黄缨菊的Clean reads。若测序深度足够充分, 每条序列中每个碱基取出的全部K-mer可以覆盖整个基因组, 那么依据Lander waterman算法, 基因组大小(G)满足以下公式(Lander and Waterman, 1988):

$$C_{\text{base}} = \frac{C_{\text{K-mer}} \times L}{L - K + 1} \quad G = \frac{n_{\text{K-mer}}}{C_{\text{K-mer}}} = \frac{n_{\text{base}}}{C_{\text{base}}} \quad (1)$$

式中, n_{base} 和 $n_{\text{K-mer}}$ 分别为序列的碱基总数和K-mer数, C_{base} 和 $C_{\text{K-mer}}$ 分别为覆盖碱基的期望深度和K-mer期望覆盖深度。

若K-mer的覆盖深度频度分布服从泊松分布:

$$P_{\text{Kspecies}}(x) = \frac{c^x}{x!} e^{-c} \quad (2)$$

式中, c 为峰值对应的测序K-mer深度估计值。

当黄缨菊的K-mer数据同时满足上述假设条件时, 可以通过读取每个碱基的K-mer方式统计K-mer频度分布, 以K-mer深度(Depth)为横坐标, K-mer频度(Frequency)为纵坐标, 绘制黄缨菊的K-mer深度频度分布图, 以此预估黄缨菊K-mer的深度值, 并计算其基因组大小; 同时, 根据黄缨菊主峰深度1.8倍后的K-mer个数占比估算其基因组重复序列比例, 并依据杂合峰与纯合峰面积占比估算其杂合率。

1.6 LTR-RTs特征分析

通过SOAPdenovo (Vurture et al., 2017)对黄缨菊高质量Clean reads进行组装得到Contig, 经gap填充后, 继续拼接形成Scaffold, 获得含有重复序列的初级基因组序列; 使用LTR_Finder (Xu and Wang,

2007)和LTR_harvest (Ellinghaus et al., 2008)组合鉴定基因组序列中的LTR分布, 参数设置-w 2 -C -D 15000 -d 1000 -L 7000 -l 100 -p 20 -M 0.85, 并使用LTR_retreiver (Ou and Jiang, 2018)统计整合两者的预测结果; 同时借助向日葵(*Helianthus annuus*)的碱基平均替换速率(参数-u=1e-8), 依据公式 $T=K/2\mu$ 计算LTR插入时间(Badouin et al., 2017)。

2 结果与分析

2.1 流式细胞术预估基因组大小

采用流式细胞术对黄缨菊基因组大小进行预估, 结果显示黄缨菊有单一的主峰, 生长周期中大量细胞处于G0/G1期, 荧光峰值对应的荧光强度均值为8 652.28 (图1A)。同时, 参照长裂太行菊和番茄的荧光强度均值10 688.28和3 941.20 (图1B, C), 根据基因组大小预估公式获得黄缨菊基因组大小分别为1.94 G和1.75 G, DNA-C值为0.99 pg。

2.2 测序数据统计

利用SOAPnuke软件过滤文库构建与高通量测序得到的106 G Raw data (CRA021104), 质控后获得100 G 高质量的Clean data (表1)。使用FastQC软件评估Clean reads的测序质量, 结果显示Q20均大于97.1%, Q30均高于90.8%; AT和GC碱基含量无明显分离, GC含量比例为38.5%; GC含量密度图呈单峰形式(表1; 图2), 说明黄缨菊二代基因组测序数据质量良好。

2.3 样品污染评估

从NT数据库中随机抽取10 000条Clean reads进行同源比对, 发现黄缨菊比对占NT数据库中的总Reads数为2.36%、1.96%、0.97%、0.58%和0.48%的前5个物种分别为川木香(*Dolomiaea souliei*)、刺苞菜蓟、绒安菊(*Gynoxys mandonii*)、续断菊(*Sonchus asper*)和菜木香(*Dolomiaea edulis*) (表2), 表明黄缨菊的基

因组DNA文库测序数据不含有明显的外源污染, 可以用于后续相关分析。

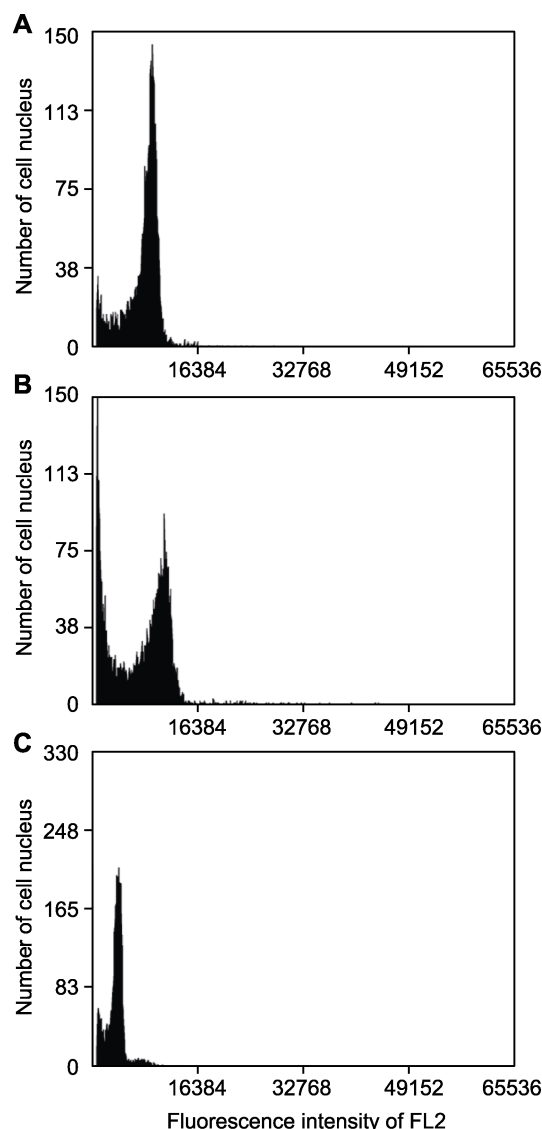


图1 流式细胞术(FCM)荧光信号图
(A) 黄缨菊; (B) 长裂太行菊; (C) 番茄

Figure 1 Flow cytometry (FCM) fluorescence pattern of different samples
(A) *Xanthopappus subacaulis*; (B) *Opisthopappus longilobus*; (C) *Solanum lycopersicum*

表1 黄缨菊测序数据统计

Table 1 Sequencing data statistics of *Xanthopappus subacaulis*

Date type	Library name	Read number	Base count (bp)	Read length (bp)	Q20 (%)	Q30 (%)	GC content (%)
Raw data	<i>Xanthopappus subacaulis</i>	710885064	106632759600	150; 150	98.7; 97.2	95.4; 91.0	38.6; 38.5
Clean data	<i>X. subacaulis</i>	669190080	100378512000	150; 150	98.6; 97.1	95.3; 90.8	38.5; 38.5

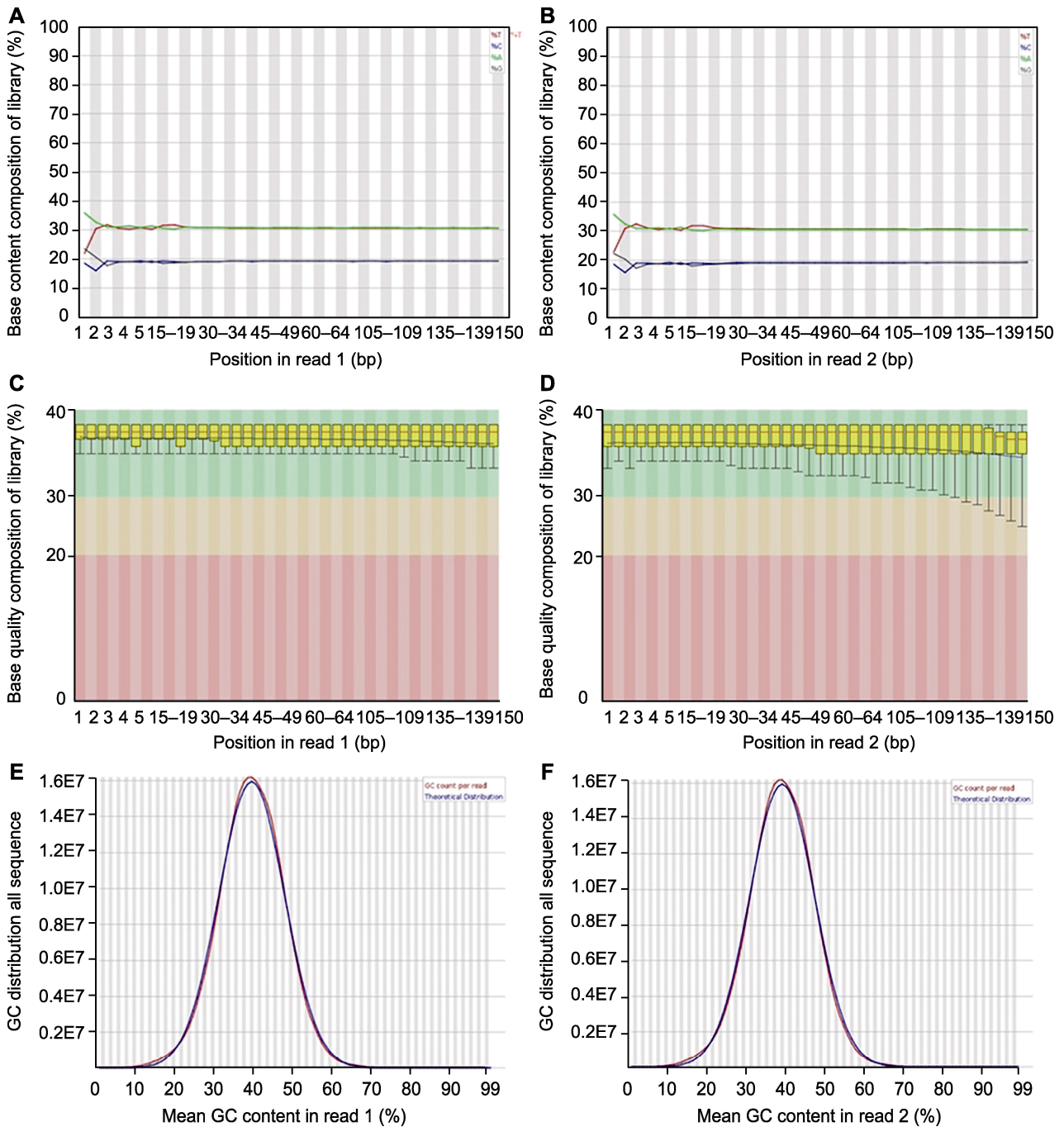


图2 黄缨菊测序质量检测与评估
(A), (B) Reads碱基分布图; **(C), (D)** Reads碱基质量分布图; **(E), (F)** Reads GC含量密度图

Figure 2 Sequencing quality evaluation for *Xanthopappus subacaulis*
(A), (B) Base distribution plot; **(C), (D)** Base quality distribution plot; **(E), (F)** GC content density plot

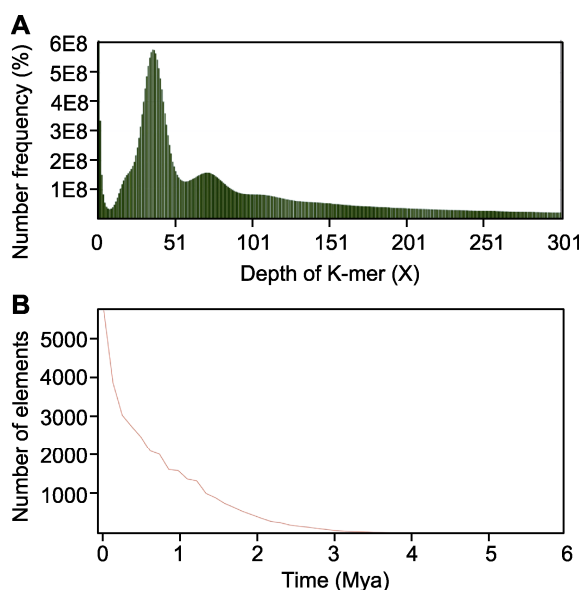
2.4 基因组大小测算及杂合率和重复序列评估

通过上述公式(1)和(2)计算得到黄缨菊基因组K-mer深度频率分布图。黄缨菊K-mer期望深度值为36, 单一主峰区域面积占比较大(图3), 应隶属于典型的二

倍体物种; 同时, 基于Lander waterman算法, 本研究推测黄缨菊的基因组大小(G)为2 198.5 Mb。此外, 研究结果还显示主峰左侧约K-mer深度1/2处有1个杂合峰, 依据杂合峰与纯合峰面积比可知黄缨菊基因组杂

表2 NCBI核苷酸数据库比对**Table 2** NCBI nucleotide database BLAST

Species	Comparison rate (%)
<i>Dolomiaea souliei</i>	2.36
<i>Cynara cardunculus</i>	1.96
<i>Gynoxys mandonii</i>	0.97
<i>Sonchus asper</i>	0.58
<i>D. edulis</i>	0.48

**图3** 黄缨菊基因组特征分析

(A) K-mer深度频率分布图; (B) 长末端重复反转录转座子(LTR-RTs)插入时间分析

Figure 3 Genomic characterization analysis of *Xanthopappus subacaulis*

(A) Depth frequency distribution of K-mer; (B) Insertion time estimation of long terminal repeat retrotransposons (LTR-RTs)

合率为0.69% (图3A); 主峰右侧K-mer深度约2倍处有1个重复峰, 通过计算纯合峰对应深度值1.8倍后的K-mer个数占比得到其重复序列比为80.15% (图3A)。因此, 黄缨菊的基因组是一种微杂合、高重复序列的复杂基因组。

2.5 LTR-RTs特征分析

使用软件LTR_Finder和LTR_retreiver对黄缨菊基因组序列中的LTR-RTs进行鉴定、统计和分析。黄缨菊基因组中共鉴定到238 282个完整的LTR反转录转座子, 其中Copia家族数量最多, 为95 073个, 占全基因组的30.72%, 表现出较高的活性; 而Gypsy家族和

表3 黄缨菊基因组长末端重复反转录转座子(LTR-RTs)分类与特征**Table 3** Classification and characterization of long terminal repeat retrotransposons (LTR-RTs) in the genome of *Xanthopappus subacaulis*

Transposable element	Class	Number	Base count (bp)	Percent (%)
LTR-RTs	Copia	95073	674533731	30.72
LTR-RTs	Gypsy	73190	739177424	33.66
LTR-RTs	Unknown	70092	363196738	16.54

Unknown分别为73 190个和70 092个, 占全基因组的33.66%和16.54% (表3)。利用碱基中性替换率 $\mu=1e-8$ 计算黄缨菊基因组中LTR的插入时间, 结果表明, LTR的插入开始于大约3 Mya并逐渐增加, 且在1 Mya内呈现爆发式增长, 说明黄缨菊中LTR反转录转座子扩增多集中于近期(图3B)。

3 讨论与结论

流式细胞术是一种可以鉴定染色体倍性和基因组相对大小的经典方法(陈婷婷, 2023)。研究表明, 选取不同的参考物种和进样方式对流式细胞术的荧光信号具有不同的影响(黄诗宇等, 2023)。本研究以长裂太行菊和番茄为参照物种, 采用“外标法”分别测定黄缨菊基因组的相对大小, 结果表明黄缨菊为二倍体, 预估其基因组相对大小分别为1.94 G和1.75 G, DNA-C值为0.99 pg, 不同参照物种对黄缨菊基因组相对大小的评估有显著影响。先前研究认为, 菊科植物具有丰富的染色体倍性和较大的基因组大小差异(孙浩男等, 2021), 其中春黄菊族基因组最大(138.88 G) (Garcia et al., 2014), 紫菀族基因组最小(335.00 Mb) (Peng et al., 2014); 同样依C值而言, 通过查阅植物DNA-C值数据库(Leitch et al., 2019), 其收录的649种菊科植物中C值最大为26.15 pg (*Senecio pendulus*), 最小为0.45 pg (*Erigeron canadensis*)。或许, 菊科植物如此大的基因组和C值差异正是其在全球范围内各类生境尺度上广泛分布的原因(解梦等, 2018)。本研究采用流式细胞术预估的黄缨菊基因组相对大小和C值均隶属于菊科植物的基因组特征范畴。

基因组Survey是评估物种基因组大小及其特征的有效手段(赵蕊蕊, 2023)。本研究基于BGI T7平台对黄缨菊基因组完成了双末端测序并进行质量评估

与比对,发现基因组测序质量良好且无外源污染,GC含量为38.5%。同时,K-mer分析显示黄缨菊为二倍体,基因组大小为2 198.50 Mb,杂合度为0.69%,重复序列占比为80.15%。通常情况下,真核生物基因组的GC含量变化范围为30%–50%。有研究表明,一方面GC含量越高,基因密度越大,抵抗高温和碱性等逆境条件的能力越强;另一方面GC含量的多种变换还可能在基因表达调控和基因突变时发挥重要作用(周慧琦,2014)。从目前已公布菊科植物基因组的GC含量看,青蒿(*Artemisia annua*)的GC含量为31.5%(Shen et al., 2018),刺苞菜蓟为32%(Scaglione et al., 2016),菊苣(*Cichorium intybus*)为36%(费星宇等,2022),长裂太行菊为36.63%(王祎玲等,2022),黄缨菊的GC含量高达38.5%,我们认为这可能与黄缨菊具有较强的抵御干旱、高温和强辐射等逆境条件有关。此外,研究还表明,菊科植物种内存在严重的自交不亲和现象,使得菊科植物基因组的杂合度较高(杨永恒等,2023)。同样,黄缨菊的细胞学和形态学证据也表明其为二倍体,染色体基数为9,小花两性,异花授粉,隶属于菊科植物特征范畴。高比例的重复序列是菊科植物较大基因组的共有特征(Garcia et al., 2014),如菊花脑(*Chrysanthemum indicum*)的重复序列占比为69.58%(Zhang et al., 2023)、薇甘菊(*Mikania micrantha*)为75.98%(Liu et al., 2020)、太行菊(*O. taihangensis*)为84.35%(王祎玲等,2022)。本研究表明,黄缨菊具有较高的杂合度(0.69%)和较大的重复序列占比(80.15%),进一步验证了先前研究结果(Garcia et al., 2014; 杨永恒等,2023)的正确性和合理性;此外,较多的重复序列也与基于黄缨菊转录组数据证实其具有丰富的SSR位点的结论相吻合(郑长远,2023)。然而,黄缨菊基因组的高杂合度和重复序列比例可能对其基因组测序、组装和注释产生影响(高胜寒等,2018)。

通过对黄缨菊测序数据进行组装以及对基因组中LTR-RTs的鉴定,发现其中Copia家族数量最多,表现出较高的活性;而Gypsy家族和Unknown次之。LTR可自主通过以RNA为中间体的逆转录在植物基因组中移动并通过DNA甲基化等作用改变基因活性,从而促进基因组分化和进化,对基因组的大小和结构造成较大影响(Amy Lyu et al., 2023)。本研究鉴定到的LTR-RTs在黄缨菊基因组中丰度相当高,该现象

与向日葵(Natali et al., 2013)和星香菊(McEvoy et al., 2024)等菊科植物基因组以及其它植物类群(Qiu and Ungerer, 2018)的鉴定结果相一致;但不同的是,黄缨菊Gypsy和Copia占全基因组的比例普遍小于菊科其它物种,该现象对基因表达过程是否产生影响还需通过更精确的组装、注释及比较基因组分析后进一步阐释。黄缨菊基因组中LTR-RTs的插入始于大约3 Mya,且在近期呈现大量扩增。同样,同属于管状花亚科的薇甘菊基因组中,约80%的LTR-RTs在1 Mya内产生大量爆发,有研究表明该现象与其基因组具有复杂的古多倍体历史有关(Liu et al., 2020)。

综上,本研究基于流式细胞术和基因组Survey分析了黄缨菊的基因组大小及特征,丰富了菊科菜蓟族的基因组信息库,也为黄缨菊高质量基因组遗传图谱构建和关键功能基因挖掘提供了基础数据。后续可采用二代与三代相结合并辅以Hi-C技术和转录组数据进行黄缨菊全基因组的*de novo*测序、组装和注释,通过比较基因组学研究更加深入全面地揭示黄缨菊基因组的动态演化历史。

作者贡献声明

靳佳瑞: 样品采集、数据处理分析和初稿撰写; 刘玉萍、苏旭和刘涛: 实验设计、资金支持、技术指导和论文修改; 余明君和杨倩: 协助完成实验; 曲荣举和张朋辉: 参与文章讨论; 才让扎西、南措加和周乐怡: 协助野外采样。

参考文献

- Amy Lyu MJ, Tang QM, Wang YJ, Essemine J, Chen FM, Ni XX, Chen GY, Zhu XG (2023). Evolution of gene regulatory network of C₄ photosynthesis in the genus *Flaveria* reveals the evolutionary status of C₃-C₄ intermediate species. *Plant Commun* 4, 100426.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Bérard A, Bergès H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadiou C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambesan S,

- Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Varès D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Muñoz S, Vincourt P, Rieseberg LH, Langlade NB (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546**, 148–152.
- Bancheva S, Greilhuber J (2006). Genome size in Bulgarian *Centaurea* s.l. (Asteraceae). *Plant Syst Evol* **257**, 95–117.
- Chen TT (2023). Ploidy Identification and Genome Size Determination of the Genus *Aspidistra* Based on Flow Cytometry. Master's thesis. Nanning: Guangxi Minzu University. pp. 1–60. (in Chinese)
- 陈婷婷 (2023). 基于流式细胞术对蜘蛛抱蛋属植物的倍性鉴定及其基因组大小测定的研究. 硕士学位论文. 南宁: 广西民族大学. pp. 1–60.
- Chen YX, Chen YS, Shi CM, Huang ZB, Zhang Y, Li SK, Li Y, Ye J, Yu C, Li Z, Zhang XQ, Wang J, Yang HM, Fang L, Chen Q (2018). SOAPnucle: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, gix120.
- Deng HK, Luo L, Wang RQ, Gao SY, Zhang WJ (2023). Genome size determination of *Scirpus mariqueter* and its related species. *Guihaia* **43**, 1838–1848. (in Chinese)
- 邓颀珂, 罗凌, 王若秋, 高少羽, 张文驹 (2023). 海三棱藨草及其近缘种基因组大小的测定. *广西植物* **43**, 1838–1848.
- Dong CC, Wang S, Zhang H, Liu JQ, Li MJ (2022). Karyotype evolution of the Asterids insights from the first genome sequences of the family Cornaceae. *DNA Res* **30**, dsac051.
- Ellinghaus D, Kurtz S, Willhoeft U (2008). *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18.
- Fei XY, Zhao H, Du Y, Duan MQ, Wang CC, Zhang B, Liu X, Li DY, Xu LX (2022). Genome size estimation and genome survey sequencing of chicory (*Cichorium intybus* L.). *Acta Agrestia Sin* **30**, 3207–3214. (in Chinese)
- 费星宇, 赵泓, 杜洋, 段梦琪, 王晨晨, 张彬, 刘雪, 李大勇, 许立新 (2022). 菊苣的基因组大小估算及基因组调查测序. *草地学报* **30**, 3207–3214.
- Finnegan DJ (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet* **5**, 103–107.
- Fomicheva M, Domblides E (2023). Mastering DNA content estimation by flow cytometry as an efficient tool for plant breeding and biodiversity research. *Methods Protoc* **6**, 18.
- Gao SH, Yu HY, Wu SY, Wang S, Geng JN, Luo YF, Hu SN (2018). Advances of sequencing and assembling technologies for complex genomes. *Hereditas (Beijing)* **40**, 944–963. (in Chinese)
- 高胜寒, 禹海英, 吴双阳, 王森, 耿佳宁, 骆迎峰, 胡松年 (2018). 复杂基因组测序技术研究进展. *遗传* **40**, 944–963.
- Garcia S, Leitch IJ, Anadon-Rosell A, Canela MÁ, Gálvez F, Garnatje T, Gras A, Hidalgo O, Johnston E, Mas de Xaxars G, Pellicer J, Siljak-Yakovlev S, Vallès J, Vitales D, Bennett MD (2014). Recent updates and developments to plant genome size databases. *Nucleic Acids Res* **42**, D1159–D1166.
- Goodwin S, McPherson JD, McCombie WR (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351.
- Guo SL, Yu J, Li DD, Zhou P, Fang Q, Yin LP (2015). DNA C-values of 138 herbaceous species and their biological significance. *Acta Ecol Sin* **35**, 6516–6529. (in Chinese)
- 郭水良, 于晶, 李丹丹, 周平, 方其, 印丽萍 (2015). 长三角及邻近地区138种草本植物DNA C-值测定及其生物学意义. *生态学报* **35**, 6516–6529.
- Huang AJ, Zhou JY, Li TZ, Xing YD, Gao F, Zhou YJ (2019). Flow cytometry and K-mer analysis estimates of genome size of *Sophora alopecuroides*. *Chin Tradit Herb Drugs* **50**, 6098–6102. (in Chinese)
- 黄阿晶, 周佳熠, 李天泽, 邢怡德, 高飞, 周宜君 (2019). 基于流式细胞术和K-mer分析的苦豆子基因组大小估测. *中草药* **50**, 6098–6102.
- Huang SY, Zhang SW, Yang TW, Gao MR, Zhang XJ, Li T, Yu WH, Shi Q, Meng P (2023). Genome size estimates of medicinal and edible plant *Erythralum scandens* based on flow cytometry and genome survey. *Mol Plant Breed* **21**, 5281–5288. (in Chinese)
- 黄诗宇, 张尚文, 杨天为, 高曼熔, 张向军, 李婷, 庾韦花, 石前, 蒙平 (2023). 基于流式细胞术和基因组Survey测定药食同源植物赤苍藤的基因组大小. *分子植物育种* **21**, 5281–5288.
- Kong XM, Zhang Y, Wang ZY, Bao ST, Feng YS, Wang JQ, Yu ZJ, Long F, Xiao ZJ, Hao YN, Gao XT, Li YF, Ding Y, Wang JY, Lei TY, Xu CY, Wang JP (2023). Two-step model of paleohexaploidy, ancestral genome reshuffling and plasticity of heat shock response in Asteraceae. *Hortic Res* **10**, uhad073.
- Lander ES, Waterman MS (1988). Genomic mapping by

- fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239.
- Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD** (2019). Plant DNA C-values database (Release 7.1). <https://cvalues.science.kew.org/>. 2019-04.
- Li A, Liu A, Du X, Chen JY, Yin M, Hu HY, Shrestha N, Wu SD, Wang HQ, Dou QW, Liu ZP, Liu JQ, Yang YZ, Ren GP** (2020). A chromosome-scale genome assembly of a diploid alfalfa, the progenitor of autotetraploid alfalfa. *Hor-tic Res* **7**, 194.
- Li T, Tang SJ, Li W, Zhang SB, Wang JL, Pan DF, Lin ZL, Ma X, Chang YN, Liu B, Sun J, Wang XF, Zhao MJ, You CQ, Luo HF, Wang MJ, Ye XG, Zhai JX, Shen ZB, Du HL, Song XW, Huang G, Cao XF** (2023). Genome evolution and initial breeding of the Triticeae grass *Leymus chinensis* dominating the Eurasian steppe. *Proc Natl Acad Sci USA* **120**, e2308984120.
- Liang FP, Wen XN, Gao HY, Zhang Y** (2018). Analysis of chloroplast genomes features of Asteraceae species. *Genomics Appl Biol* **37**, 5437–5447. (in Chinese)
梁凤萍, 文祥宁, 高赫一, 张颖 (2018). 菊科植物叶绿体基因组特征分析. 基因组学与应用生物学 **37**, 5437–5447.
- Liu B, Yan J, Li WH, Yin LJ, Li P, Yu HX, Xing LS, Cai ML, Wang HC, Zhao MX, Zheng J, Sun F, Wang ZZ, Jiang ZY, Ou QJ, Li SB, Qu L, Zhang QL, Zheng YP, Qiao X, Xi Y, Zhang Y, Jiang F, Huang C, Liu CH, Ren YW, Wang S, Liu HW, Guo JY, Wang HH, Dong H, Peng CL, Qian WQ, Fan W, Wan FH** (2020). *Mikania micrantha* genome provides insights into the molecular mechanism of rapid growth. *Nat Commun* **11**, 340.
- Liu T, Liu YP, Fu G, Lü T, Liu F, Zhang Y, Su DD, Wang YN, Zheng CY, Su X** (2022). Estimation of genome size for *Psammochloa villosa* by flow cytometry and K-mer analysis. *Acta Pratacult Sin* **31**(12), 133–145. (in Chinese)
刘涛, 刘玉萍, 富贵, 吕婷, 刘峰, 张雨, 苏丹丹, 王亚男, 郑长远, 苏旭 (2022). 基于流式细胞法和K-mer分析法检测沙鞭基因组大小. 草业学报 **31**(12), 133–145.
- Ma PJ, Zhou JY, Sun HG, Ma DD, Gao F, Zhou YJ, Zhang GF** (2018). Estimation of jojoba genome's size by flow cytometry and K-mer analysis. *J Beijing Norm Univ (Nat Sci)* **54**, 733–737. (in Chinese)
马鹏举, 周佳熠, 孙会改, 马丹丹, 高飞, 周宜君, 张根发 (2018). 基于流式细胞术和K-mer分析的好好芭基因组大小估测. 北京师范大学学报(自然科学版) **54**, 733–737.
- Ma ZL, Liu F, Zhang Y, Su X** (2020). Research progress and germplasm collection of *Xanthopappus subacaulis* which is endemic species from the Qinghai-Tibet Plateau. *J Qinghai Norm Univ (Nat Sci)* **36**(2), 41–44. (in Chinese)
马子兰, 刘峰, 张阳, 苏旭 (2020). 青藏高原特有植物黄缨菊研究现状及种质资源收集. 青海师范大学学报(自然科学版) **36**(2), 41–44.
- Mao CQ, Sha XF, Huang J, Tao S, Peng F, Li Q, Zhang C, Yuan C** (2023). Genome survey and characteristic analysis of *Ligusticum chuanxiong*. *Chin Tradit Herb Drugs* **54**, 907–914. (in Chinese)
毛常清, 沙秀芬, 黄静, 陶珊, 彭芳, 李群, 张超, 袁灿 (2023). 川芎基因组survey测序及其特征分析. 中草药 **54**, 907–914.
- Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang XQ, Zhang QS, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Šimková H, Staňková H, Vrána J, Chan S, Muñoz-Amatriaín M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao SJ, Chapman B, Dai F, Han Y, Li H, Li X, Lin CY, McCooke JK, Tan C, Wang PH, Wang SB, Yin SY, Zhou GF, Poland JA, Bellgard MI, Borisjuk L, Houben A, Doležel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang GP, Braumann I, Spannagl M, Li CD, Waugh R, Stein N** (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433.
- McEvoy SL, Meyer RS, Hasenstab-Lehman KE, Williams CM** (2024). The reference genome of an endangered Asteraceae, *Deinandra increscens* subsp. *villosa*, endemic to the Central Coast of California. *G3 (Bethesda)* **14**, jkae117.
- Mishiba KI, Ando T, Mii M, Watanabe H, Kokubun H, Hashimoto G, Marchesi E** (2000). Nuclear DNA content as an index character discriminating taxa in the genus *Petunia sensu* Jussieu (Solanaceae). *Ann Bot* **85**, 665–673.
- Morabito C, Aiese Cigliano R, Maréchal E, Rébeillé F, Amato A** (2020). Illumina and PacBio DNA sequencing data, *de novo* assembly and annotation of the genome of *Aurantiochytrium limacinum* strain CCAP_4062/1. *Data Brief* **31**, 105729.
- Natali L, Cossu RM, Barghini E, Giordani T, Buti M,**

- Mascagni F, Morgante M, Gill N, Kane NC, Rieseberg L, Cavallini A** (2013). The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genomics* **14**, 686.
- Ou SJ, Jiang N** (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* **176**, 1410–1422.
- Peng YH, Lai Z, Lane T, Nageswara-Rao M, Okada M, Jasieniuk M, O'Geen H, Kim RW, Sammons RD, Rieseberg LH, Stewart CN Jr** (2014). *De novo* genome assembly of the economically important weed horseweed using integrated data from multiple sequencing platforms. *Plant Physiol* **166**, 1241–1254.
- Qiu F, Ungerer MC** (2018). Genomic abundance and transcriptional activity of diverse *gypsy* and *copia* long terminal repeat retrotransposons in three wild sunflower species. *BMC Plant Biol* **18**, 6.
- Scaglione D, Reyes-Chin-Wo S, Acquadro A, Froenicke L, Portis E, Beitel C, Tirone M, Mauro R, Lo Monaco A, Mauromicale G, Faccioli P, Cattivelli L, Rieseberg L, Michelmore R, Lanteri S** (2016). The genome sequence of the outbreeding globe artichoke constructed *de novo* incorporating a phase-aware low-pass sequencing strategy of F₁ progeny. *Sci Rep* **6**, 19427.
- Shen Q, Zhang LD, Liao ZH, Wang SY, Yan TX, Shi P, Liu M, Fu XQ, Pan QF, Wang YL, Lv ZY, Lu X, Zhang FY, Jiang WM, Ma YN, Chen MH, Hao XL, Li L, Tang YL, Lv G, Zhou Y, Sun XF, Brodelius PE, Rose JKC, Tang KX** (2018). The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Mol Plant* **11**, 776–788.
- Shi HL** (2011). Study on the resistance indexes of four kinds of desert plants in Qinghai Lake area. *J Anhui Agric Sci* **39**, 14227–14229. (in Chinese)
史惠兰 (2011). 青海湖地区4种荒漠植物抗逆性指标的研究. *安徽农业科学* **39**, 14227–14229.
- Staton SE, Burke JM** (2015). Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. *BMC Genomics* **16**, 623.
- Sun HN, Li MY, Liu DY, Wang X, Tian L, Wu FF** (2021). Chromosome number identification and karyotype analysis of *Coreopsis* L. *Acta Agrestia Sin* **29**, 2477–2485. (in Chinese)
孙浩男, 李明阳, 刘冬云, 王鑫, 田琳, 吴芳芳 (2021). 不同种(品种)金鸡菊的染色体数目鉴定及核型分析. *草地学报* **29**, 2477–2485.
- Tian YQ, Zhuang LD, Xu HH** (2008). A new screening method for photo-activated insecticidal plants. *J Huazhong Agric Univ* **27**, 370–372. (in Chinese)
田永清, 庄流东, 徐汉虹 (2008). 光活化杀虫植物的1种新型筛选方法. *华中农业大学学报* **27**, 370–372.
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC** (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204.
- Wang YJ, Liu JQ, Miede G** (2007). Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. *Ann Bot* **99**, 311–322.
- Wang YL, Zang E, Zhang H, Liu ZX, Lan YF, He S, Hao WL, Cao YL** (2022). Genome sizes and characteristics of cliff plants *Opisthopappus taihangensis* and *O. longilobus* on Taihang Mountains. *Guihaia* **42**, 1582–1589. (in Chinese)
王祎玲, 臧恩, 张昊, 刘志霞, 兰亚飞, 何珊, 郝伟丽, 曹艳玲 (2022). 崖壁植物太行菊与长裂太行菊全基因组大小及特征分析. *广西植物* **42**, 1582–1589.
- Xia CJ, Li YG, Xia S, Pang W, Chen CL** (2024). Flow cytometric analysis and sorting in plant genomics. *Chin Bull Bot* **59**, 774–782. (in Chinese)
夏春皎, 李运广, 夏姝, 庞伟, 陈春丽 (2024). 植物基因组学中的流式细胞分析及分选技术. *植物学报* **59**, 774–782.
- Xie M, Yu J, Guo SL** (2018). Latitudinal variation patterns of plant nuclear DNA amount on a global scale and their environmental adaptation significance: a case study with Asteraceae. *Acta Ecol Sin* **38**, 3453–3461. (in Chinese)
解梦, 于晶, 郭水良 (2018). 植物核DNA含量在全球尺度上的纬度变异式样及其气候适应意义——以菊科植物为例. *生态学报* **38**, 3453–3461.
- Xin HB, Ji FF, Wu J, Zhang SY, Yi CJ, Zhao SW, Cong RC, Zhao LJ, Zhang H, Zhang Z** (2023). Chromosome-scale genome assembly of marigold (*Tagetes erecta* L.): an ornamental plant and feedstock for industrial lutein production. *Hortic Plant J* **9**, 1119–1130.
- Xu Z, Wang H** (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–W268.
- Yang W, Si YY, Xu RW, Chen XH** (2023). Characterization of microsatellites and polymorphic marker development in ragworm (*Tylorrhynchus heterochaetus*) based on genome survey data. *South China Fish Sci* **19**(5), 123–133. (in Chinese)

- 杨尉, 司圆圆, 许瑞雯, 陈兴汉 (2023). 基于基因组survey数据的疣吻沙蚕微卫星特征分析及多态标记开发. 南方水产科学 **19**(5), 123–133.
- Yang YH, Zhang YX, Zhang T, Xu XY, Sun YM, Wang XM, Tong HY, Yuan HY (2023). Fluorescence microscopy and transcriptome analysis on self-pollinated stigmas of self-incompatible and self-compatible clones of *Stevia rebaudiana*. *J Plant Resour Environ* **32**(4), 25–32. (in Chinese)
- 杨永恒, 张永侠, 张婷, 徐晓洋, 孙玉明, 王小敏, 佟海英, 原海燕 (2023). 甜菊自交不亲和及自交亲和无性系自花授粉柱头的荧光显微观察和转录组分析. 植物资源与环境学报 **32**(4), 25–32.
- Yang YY, Li SN, Xing YP, Zhang ZG, Liu T, Ao WLJ, Bao GH, Zhan ZL, Zhao R, Zhang TT, Zhang DC, Song YY, Bian C, Xu L, Kang TG (2022). The first high-quality chromosomal genome assembly of a medicinal and edible plant *Arctium lappa*. *Mol Ecol Resour* **22**, 1493–1507.
- Zhang FJ, Chen FD, Schwarzacher T, Heslop-Harrison JS, Teng NJ (2023). The nature and genomic landscape of repetitive DNA classes in *Chrysanthemum nankingense* shows recent genomic changes. *Ann Bot* **131**, 215–228.
- Zhang JD, Feng M (2023). A plant sample optimal pre-treatment for flow cytometric analysis. *Chin Bull Bot* **58**, 285–297. (in Chinese)
- 张晋丹, 冯旻 (2023). 一种提升流式细胞术分析效果的前处理方法. 植物学报 **58**, 285–297.
- Zhang L (2014). Studies on Secondary Metabolites and Their Biological Activity of Three Medicinal Plants. PhD dissertation. Lanzhou: Lanzhou University. pp. 1–160. (in Chinese)
- 张丽 (2014). 三种药用植物次生代谢物结构及其生物活性研究. 博士学位论文. 兰州: 兰州大学. pp. 1–160.
- Zhang L, Chen CJ, Chen J, Zhao QQ, Li Y, Gao K (2014). Thiophene acetylenes and furanosesquiterpenes from *Xanthopappus subacaulis* and their antibacterial activities. *Phytochemistry* **106**, 134–140.
- Zhang X, Ping PY, Hutvagner G, Blumenstein M, Li JY (2021). Aberration-corrected ultrafine analysis of miRNA reads at single-base resolution: a K-mer lattice approach. *Nucleic Acids Res* **49**, e106.
- Zhang Y, Ma ZL, Xu SS, Su X, Li MY (2022). Phylogeography of *Xanthopappus subacaulis* (Asteraceae), an endemic species from the northeastern of the Qinghai-Tibet Plateau. *Bull Bot Res* **42**, 565–573. (in Chinese)
- 张阳, 马子兰, 徐珊珊, 苏旭, 李梅英 (2022). 青藏高原东北部黄缨菊的谱系地理学. 植物研究 **42**, 565–573.
- Zhao L, Zhu YH, Wang M, Han YG, Ma LG, Feng WS, Zheng XK (2021). Estimation of *Rehmannia glutinosa* genome size based on flow cytometry and genome survey analysis. *Chin Tradit Herb Drugs* **52**, 821–826. (in Chinese)
- 赵乐, 朱昀昊, 王敏, 韩永光, 马利刚, 冯卫生, 郑晓珂 (2021). 基于流式细胞术和基因组survey分析的地黄基因组研究. 中草药 **52**, 821–826.
- Zhao RR (2023). Genome Survey Analysis and Genetic Marker Identification of the White-Spotted Bamboo Shark *Chiloscyllium Plagiosum*. Master's thesis. Zhoushan: Zhejiang Ocean University. pp. 1–82. (in Chinese)
- 赵蕊蕊 (2023). 条纹斑竹鲨基因组survey分析及遗传标记开发. 硕士学位论文. 舟山: 浙江海洋大学. pp. 1–82.
- Zheng CY (2023). Development of SSR Molecular Markers and Analyses of Genetic Diversity in *Xanthopappus Subacaulis* (Asteraceae). Master's thesis. Xining: Qinghai Normal University. pp. 1–64. (in Chinese)
- 郑长远 (2023). 黄缨菊SSR分子标记开发及遗传多样性研究. 硕士学位论文. 西宁: 青海师范大学. pp. 1–64.
- Zhou HQ (2014). Analysis of the Relationship Between Genomic GC Content and Patterns of Base Usage, Codon Usage and Amino Acid Usage in Prokaryotes. Master's thesis. Chengdu: University of Electronic Science and Technology of China. pp. 1–55. (in Chinese)
- 周慧琦 (2014). 基因组GC含量与碱基、密码子和氨基酸使用偏好的关系. 硕士学位论文. 成都: 电子科技大学. pp. 1–55.
- Zhou P, Zhang Q, Li J, Li F, Huang J, Zhang M (2023). A first insight into the genomic background of *Ilex pubescens* (Aquifoliaceae) by flow cytometry and genome survey sequencing. *BMC Genomics* **24**, 270.

Genome Size and Characteristics Analysis of *Xanthopappus subacaulis* Based on Flow Cytometry and Genome Survey

Jiarui Jin^{1,2}, Yuping Liu^{1,2,3}, Xu Su^{1,2,3*}, Tao Liu^{4*}, Mingjun Yu^{1,2}, Qian Yang^{1,2}, Rongju Qu^{1,2}
Penghui Zhang^{1,2}, Zhaxi Cairang^{1,2}, Cuojia Nan¹, Leyi Zhou¹

¹School of Life Sciences, Qinghai Normal University, Xining 810008, China; ²Key Laboratory of Biodiversity Formation Mechanism and Comprehensive Utilization of the Qinghai-Xizang Plateau in Qinghai Province, Qinghai Normal University, Xining 810008, China; ³Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China

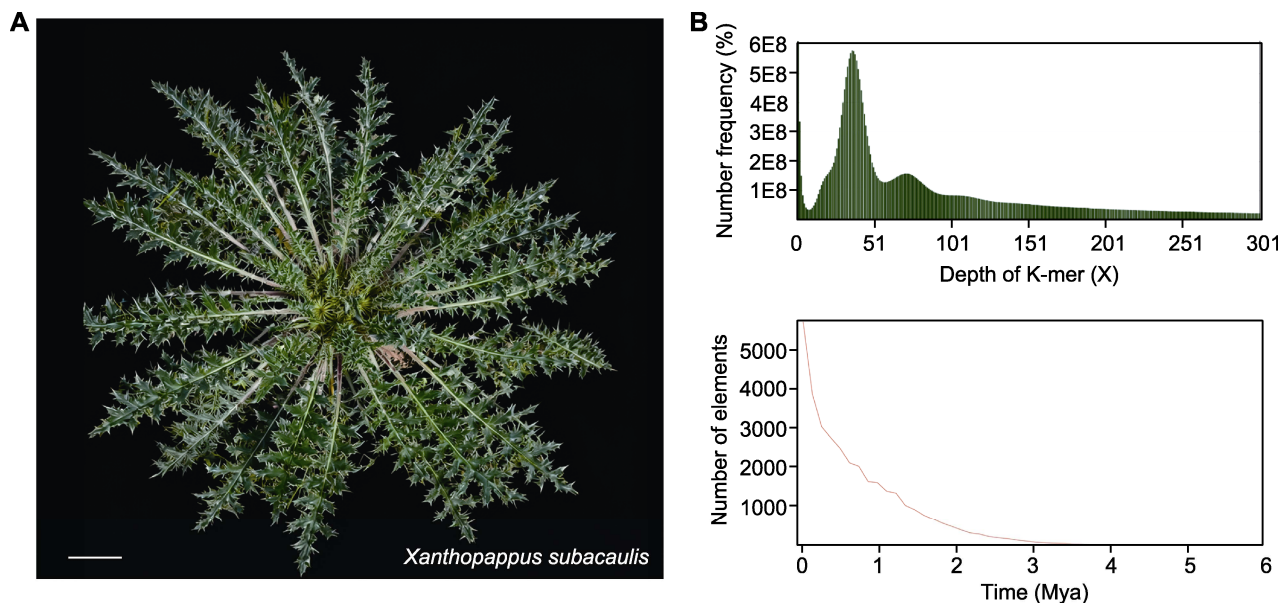
⁴School of Ecology and Environmental Science, Qinghai Institute of Technology, Xining 810016, China

INTRUDUCTION: *Xanthopappus subacaulis*, endemic to the Qinghai-Xizang Plateau, is a perennial medicinal plant from the genus *Xanthopappus* of the family Asteraceae, with important economic, ecological and medicinal values. However, genomic information for this species remains limited, hindering further genetic studies and resource utilization. Determining an appropriate sequencing strategy for its whole genome is a key prerequisite for subsequent genomic studies.

RATIONALE: In order to determine the appropriate sequencing strategy for the whole genome of *X. subacaulis*, we analyzed and evaluated its genome size, heterozygosity, repeat and GC content using flow cytometry and genome survey analysis based on BGI sequencing.

RESULTS: Flow cytometry analyses using *Opisthopappus longilobus* and *Solanum lycopersicum* as reference genomes indicated that *X. subacaulis* was a diploid, with an estimated genome size of 1.94 G and 1.75 G respectively, and a DNA-C value of 0.99 pg. We generated approximately 100.3 G of clean short read sequencing data, with a GC content of 38.5%. K-mer analysis indicated that the genome size of *X. subacaulis* was 2 198.50 Mb, with a heterozygosity rate of 0.69%, and repeat content of 80.15%. The analysis of the long terminal repeat retrotransposons (LTR-RTs) indicated that the LTR/Copia was the most abundant LTR family, accounting for 30.72% of the whole genome, while the Gypsy family and the unknown LTRs accounted for 33.66% and 16.54%, respectively. Moreover, their peak insertion time began approximately three million years ago (Mya), with a marked amplification occurring within the last 1 Mya. These results suggested that the large-scale insertion of LTR elements was (most) likely one of the important factors leading to the genomic complexity of *X. subacaulis*.

CONCLUSION: This study clarifies the key genomic characteristics of *X. subacaulis*, which provides valuable reference data resources for subsequent genetic map construction and functional gene mining of *X. subacaulis*, and also lays a foundation for determining its whole-genome sequencing strategy.



Growth morphology and genomic characteristics of *Xanthopappus subacaulis*. (A) Growth morphology of *X. subacaulis* (bar=5 cm); (B) Depth and frequency distribution of K-mer and insertion time analysis of long terminal repeat retrotransposons (LTR-RTs)

Key words *Xanthopappus subacaulis*, flow cytometry, C value, genome survey, long terminal repeat retrotransposons (LTR-RTs)

Jin JR, Liu YP, Su X, Liu T, Yu MJ, Yang Q, Qu RJ, Zhang PH, Cairang ZX, Nan CJ, Zhou LY (2025). Genome size and characteristics analysis of *Xanthopappus subacaulis* based on flow cytometry and genome survey. *Chin Bull Bot* **60**, 888–900.

* Authors for correspondence. E-mail: xusu8527972@126.com; 532226527@qq.com

(责任编辑: 朱亚娜)

通讯作者简介

苏旭, 博士后, 教授, 博士生导师, 青海师范大学生命科学学院副院长, 青海省青藏高原生物多样性形成机制与综合利用重点实验室主任, 教育部基础教育教学指导委员会委员; 青海青年五四奖章获得者, 青海省“昆仑英才·高端创新创业人才”项目培养领军人才, 首届青海省高教领域“昆仑英才·教学名师”, 青海省自然科学与工程技术学科带头人, 青海省高等院校“135高层次人才培养工程”拔尖学科带头人, 青海省省级骨干教师。长期从事高山植物系统发育、物种形成及适应进化机制等研究。主持国家自然科学基金3项; 科技基础项目2项以及省部级科研项目10余项。在国内外学术期刊上发表论文70余篇; 编写学术专著5部。授权发明专利5件; 获青海省自然科学三等奖和第二届青海省科学成果创新驱动奖各1项。