

高效自动化 HD-Marker 文库构建流程的建立及评估*

杨志辉¹, 吕佳¹, 刘平平^{1,2**}, 张天琦¹, 黄晓文¹, 倪萍¹, 包振民^{1,3,4}, 王师^{1,2,4}

(1. 中国海洋大学海洋生命学院, 海洋生物遗传学与育种教育部重点实验室, 山东 青岛 266003;

2. 青岛海洋科学与技术试点国家实验室, 海洋生物学与生物技术功能实验室, 山东 青岛 266237;

3. 青岛海洋科学与技术试点国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东 青岛 266237;

4. 中国海洋大学三亚海洋研究院, 热带海洋生物种质资源开发与种业工程实验室, 海南 三亚 572000)

摘要: 针对 HD-Marker (High-throughput, sequencing-based GoldenGate approach) 技术应用于育种研究时面临的短时间内快速完成大批量样品基因分型的挑战, 本研究基于自动化移液工作站建立了 HD-Marker 文库自动化构建体系并评估了此方法在文库构建中的可靠性。本研究基于自动化移液工作站建立了 HD-Marker 文库自动化构建流程, 并评估了此方法在文库构建中的可靠性。结果表明扩增产物具有较高均匀性, 浓度范围在 23.1~38.7 ng/ μ L 之间; 90.63% 的文库可获得清晰文库条带; 文库位点捕获效率均在 97% 以上, 分型位点比例能达到 100%, 与标准 WGS 文库数据相比, 位点分型准确性也在 97%~100% 之间。此自动化文库构建流程可并行处理 96 个样品的文库制备, 整体运行时间仅为人工操作的 1/7。本研究首次实现了标准化的自动化 HD-Marker 基因分型文库制备, 为大规模育种群体的基因分型效率的提高提供依据, 并为全基因组选择育种的实施提供了具有实用价值的高效技术平台。

关键词: 自动化移液工作站; HD-Marker; 文库构建; 基因分型; 全基因组选择育种

中图分类号: Q-31

文献标志码: A

文章编号: 1672-5174(2022)05-066-10

DOI: 10.16441/j.cnki.hdxh.20210168

引用格式: 杨志辉, 吕佳, 刘平平, 等. 高效自动化 HD-Marker 文库构建流程的建立及评估[J]. 中国海洋大学学报(自然科学版), 2022, 52(5): 66-75.

Yang Zhihui, Lv Jia, Liu Pingping, et al. Establishment and evaluation of an efficient and automated HD-Marker library construction procedure[J]. Periodical of Ocean University of China, 2022, 52(5): 66-75.

随着现代分子生物学技术的快速发展, 分子育种技术已成为遗传改良的重要手段^[1]。全基因组选择育种^[2]利用全基因组范围的遗传标记计算基因组育种值, 在分子水平对优良品质进行选育, 大大加快了育种进程, 已成为动植物育种领域的一大研究热点。近年来, 作为动植物中开展全基因组选择育种工作的重要技术支撑, 高通量基因分型技术呈迅猛发展势头。随着二代测序(Next generation sequencing, NGS)平台的建立及测序成本的不断降低, 全基因组重测序技术已成为动植物群体遗传学及育种研究的有效方法, 但处理大量样品且物种基因组较大时费用仍然是非常昂贵的^[3]。简化基因组测序技术^[4-7]通过限制性内切酶酶切基因组 DNA, 降低基因组复杂度, 实现基因组内低成本大规模单核苷酸多态性(Single nucleotide polymorphisms, SNPs)标记的分型^[8]。但此类技术仅限于酶切标签上位点的检测, 使其更适用于全基因组范围

的非靶向 SNPs 标记开发和应用。然而基因相关位点, 如位于候选基因或候选基因附近的 SNP 位点, 可能与表型性状变异有更大关联, 在全基因组选择育种中具有更为重要的价值。随着越来越多物种基因组信息的获取及丰富的 SNP 标记资源的积累, 对靶向位点分型可以更好地实现全基因组选择育种的研究需求。

基因芯片技术是伴随着“人类基因组计划”发展而来的技术, 其最早应用于 DNA 的测序。该技术基于核酸杂交的基本原理, 通过检测芯片中固定的已知探针堆与待测样品的杂交谱重组出样品序列^[9]。基因芯片技术广泛应用于农畜牧业的遗传育种研究中, 如群体进化分析^[10]、优良性状基因的挖掘^[11]、育种值的估算^[12]等。目前在牛、猪、鸡、山羊、绵羊等家畜以及水稻、玉米、马铃薯、小麦等植物中, 相应的基因芯片都已被大量开发, 这极大推动了相关育种工作。但对非模式生物如扇贝等水产生物而言, 芯片开发的研究相对

* 基金项目: 中央高校基本科研业务费专项(202064008); 三亚崖州湾科技城管理局项目(SKJC-2020-02-009)资助

Supported by the Fundamental Research Funds for the Central Universities(202064008); the Project of Sanya Yazhouwan Science and Technology City Management Foundation(SKJC-2020-02-009)

收稿日期: 2021-04-12; 修订日期: 2021-05-10

作者简介: 杨志辉(1992—), 男, 博士生, 主要从事贝类分子遗传学研究。E-mail: meswyzh@163.com

** 通讯作者: E-mail: liupingping@ouc.edu.cn

较少, 主要由于其缺乏商业化芯片, 芯片的定制和生产费用比较昂贵^[13]。其次芯片技术位点固定^[13], 位点选择灵活性不高, 无法满足个性化要求。最后芯片技术所需的昂贵的专用仪器也是限制其推广应用的一大原因^[14]。液相分子杂交技术 HD-Marker (High-throughput, sequencing-based GoldenGate approach) 可在单个管内同时实现 1 万多个靶位点的基因分型^[15], 近期已提高其通量至 86 000。该技术通过杂交-延伸-连接等反应获取靶位点的碱基信息, 后续通过 PCR 扩增将捕获后的完整序列富集, 然后进行高通量测序。HD-Marker 将液相杂交技术的位点选择灵活性和高通量测序技术通量高、成本低的优势有效结合, 无需固化探针, 可自由选择覆盖基因组几百至几万个标记位点进行靶向分型, 为非模式生物的低成本靶位点基因分型提供了一种兼容不同通量级别的理想技术手段^[15]。但育种工作者仍然面临着短期内需快速完成大批量样本基因分型工作的挑战。自动化是解决大批量样本基因分型过程中手工操作耗时耗力问题的有效途径。Hamilton、Eppendorf、Tecan、Beckman 等几家主要的工作站生产厂家不仅在核酸提取、药物筛选等应用上提供全自动化的解决方案, 还可以搭建自动化文库构建平台, 且它们的自动化移液工作站在法医学、临床以及 NGS 服务平台等实验室获得了广泛的应用^[16-18]。

本研究旨在建立标准化的、自动化的、高效高通量的 HD-Marker 文库构建流程, 以满足短时快速对大规模样本并行基因分型的需求。本研究依托自动化移液系统 (Freedom EVO 100-8), 根据 HD-Marker 技术的流程按照实验需求配置了工作台的桌面, 测定和调节了重要移液参数并且编写了自动化实验程序, 并以虾夷扇贝 (*Patinopecten yessoensis*) 100 个位点的探针池为原料, 从扩增效果均匀性、实验成功率、文库捕获率及分型

准确性等方面验证了自动化文库构建流程的可靠性。

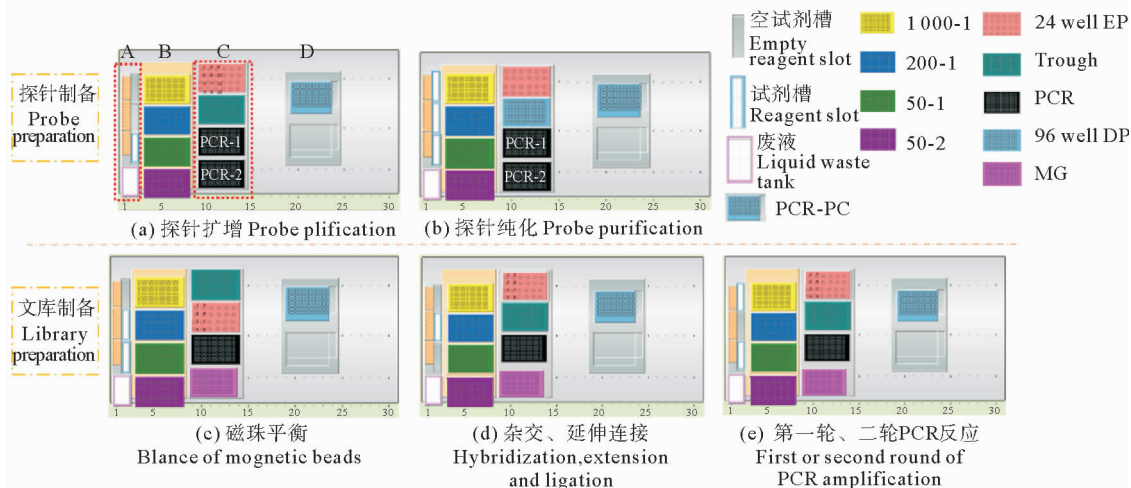
1 材料与方法

1.1 实验材料

自动化移液系统 Freedom EVO 100-8; 真空抽滤模块, 包括真空抽滤器主机、真空抽滤模块底座、真空抽滤 96 孔板适配器; 4 位加样针载架 (可回放至深孔板结构的吸头盒中); 4 位微板载架。主要耗材包括 96 孔 PCR 板 (生工生物工程股份有限公司, 货号 F600960-0010)、1.5 mL EP 管 (青岛赛尚科贸有限公司, 货号 LXGHY-15)、100 mL Through 试剂槽 (Tecan, 货号 10613049)、12 通道试剂槽 (生工生物工程股份有限公司, 货号 F619714-0001)。

1.2 实验方法

1.2.1 自动化文库构建工作台的桌面配置 桌面配置主要涉及原料槽、枪头以及样品 PCR 管等的合理摆放。应遵循的基本原则为: (1) 使步骤之间板位的放置协同性尽可能高, 无需对桌面布局产生较大的更改; (2) 使加样机械臂的移液操作基本遵循最少路径, 提高移液的效率。图 1 所示根据 HD-Marker 实验流程确立的几个主要实验步骤的桌面布局。桌面布局主要由四大部分组成, 从左到右依次为: 第一部分 (A) 为 100 mL 试剂槽组, 共有 3 个位置, 由上而下标记为 A1~A3, 用于存放用量较大的原料; 第二部分 (B) 为加样针载架, 共有 4 个位置, 由上而下标记为 B1~B4, 用于存放大中小三种规格的枪头盒; 第三部分 (C) 为试剂配制区域, 共有 4 个位置, 由上而下标记为 C1~C4, 用于放置 EP 管原料、母液配置试剂槽以及 96 孔 PCR 管等; 第四部分 (D) 为用于 PCR 产物纯化的组件, 主要配备真空抽滤 96 孔板适配器。在自动化流程中, 主要针对第一部分 (A) 和第三部分 (C) 的桌面进行更改。



(A~D: 分别代表桌面布局的第一至四部分; 1 000/200-1: 1 000/200 μ L 的加样针载架; 50-1 和 50-2: 50 μ L 的加样针载架 1 和 2; 24 well EP: 24 孔离心管架; Trough: 母液试剂配置槽; PCR: 96 孔 PCR 板; 96 well DP: 96 深孔板; MG: 磁力架; Filter: PCR 产物纯化的组件)

图 1 主要实验步骤的桌面布局

Fig.1 Layout of the desktop for main steps

1.2.2 重要移液参数的测定和调节

(1) 自动化液体处理系统吸放液针头的标准空间定位。为了使8通道加样机械臂能够准确的探测到液面的位置或者PCR管、EP管的底部,必须对使用的耗材进行空间定位,确定位置参数 X 、 Y 、 Z 。需要进行定位的耗材主要包括96孔PCR板(规格:200 μL)、12通道试剂槽、100 mL试剂槽、磁力架上放置的96孔PCR板(规格:100 mL)。主要步骤如下,加载枪头后将机械臂移至96孔PCR板的1号管上方,使用微调功能调节针头至96孔板A排1列空位的正中间,将针头下探到管底上方1~2 mm的高度获得合适的 Z 轴坐标后,单击“<”使当前调节的值覆盖原来的 Z 轴坐标值,与 X 、 Y 轴的调节以及其他耗材的调节也遵循以上的方式。

(2) 移液参数的调节。为了使移液操作精确,液体处理的配置设置也需要反复进行优化和调节。其中主要调节的部分包括:吸液速度和喷液速度、延迟时间及浸入液面的高度等。

(3) 容器吸液安全余量的确定。通常为了保证所吸取的液体足量,吸液会设置在 $Z_{\max}$ 位置-2 mm左右的距离,用自动化液体处理系统吸取定量液体后,测定管内剩余的2 mm安全距离内剩余的体积即为安全余量,不同的耗材于2 mm距离处的安全余量分别进行测定。

(4) 移液准确度和精度的测定。采用去离子水作为移液液体,并用天平称量的方式对自动化液体处理系统的移液精准度进行测定。环境温度、水温均为20 $^{\circ}\text{C}$,电子天平内相对湿度60%~90%。测定从试剂槽中吸取定量体积到96孔板之间的移液精准度。利用自动化液体处理系统进行5、10、100、1 000 μL 4个体积量的移液操作,并称重,重复15次。

1.2.3 自动化探针制备方法的建立

(1) 探针扩增流程的建立。通过一次流程,实现192管PCR母液配置及分液操作,10 min内完成操作。整个实验过程中涉及到的原料管板位放置及体系如下($\times 8$ 代表8通道的枪头, $\times 12$ 代表循环12次)。

① PCR反应所需要的 H_2O 放置在100 mL试剂槽(图1标记为试剂槽);Oligo模板、引物Fwd-Primer(10 $\mu\text{mol/L}$)、引物Rev-Primer(10 $\mu\text{mol/L}$)分别放置于24微孔板EP的1~3位,OneTaqmix(NEB)放置5~8位(见图1)。

② 在12通道母液槽(Trough)的第一列和第二列中进行母液的配置,用平行8根枪头从100 mL试剂槽位置吸取 H_2O 体积5 760 μL (720 $\mu\text{L} \times 8$),移液到第一列,循环两次,第二次喷液设置在第二列,用平行3根枪头分别从24 well EP的1、2、3位置分别吸取Oligo模板体积80 μL 、引物Fwd-Primer体积480 μL 、引

物Rev-Primer体积480 μL ,循环2次,步骤同上;用平行4根枪头从24 well EP的5~8位置吸取体积OneTaqmix体积5 760 μL (900 $\mu\text{L} \times 4$),循环两次,步骤同上。

③ 配置好母液后进行分液操作,每次调取8根枪头从母液槽的第一列吸取母液体积480 μL (60 $\mu\text{L} \times 8$),分别加到96孔PCR板(图1(a)标记为PCR-1)的第一列每孔,循环12次按列分配母液。母液槽第二列中的母液分配至第二张96孔PCR板(图1(a)标记为PCR-2)。

④ 取出两个96孔PCR板,共计192管,贴上PCR膜后随即放入PCR仪中,完成扩增反应。

(2) 探针纯化流程的建立。待PCR反应结束后,对192管PCR产物进行纯化操作,采用的是真空抽滤法同时进行96个样品的纯化。整个实验过程中涉及到的原料管板位放置及体系如下($\times 8$ 代表8通道的枪头, $\times 12$ 代表循环12次)。

① 纯化操作所需要的PM缓冲液、PE缓冲液、EB洗脱液分别放置在100 mL试剂槽位置(见图1(b)),用于样品与PM混合的收集版放置于4位微板载架的site2(图1(b)标记为96 well DP),两份96孔PCR产物分别放置在4位微板载架的site3和site4(见图1(b)标记为PCR-1及PCR-2)。真空抽滤96孔板适配器放置在真空抽滤模块底座的site1(图1(b)标记为PCR-PC)。

② 产物与PM缓冲液的混匀在96 well EP中进行,首先用平行8根枪头将PCR-1中所有液体转移至96孔深孔板,PCR-1及PCR-2的每两列转移至96 well EP的同一列。

③ 用平行8根枪头转移PM缓冲液5 760 μL (360 $\mu\text{L} \times 8 \times 12$)至96 well EP进行混匀操作,然后平行转移至PCR-PC中,开启压力设置,施加压强至350 mmPa,计时6 min;降压至0。再进行一次压强为600 mmPa的加压,计时3 min结束后释放掉压力,纯化柱中的液体抽滤到废液桶中。

④ 转移PE的步骤与步骤(3)类似,需要修改PE缓冲液体积为86 400 μL (900 $\mu\text{L} \times 8 \times 12$),PE加到收集管以后,开启压力,压力值及时间设置同上,重复2次。

⑤ 最后一遍PE抽滤后,静置5 min,使滤膜充分干燥,此时需要手动将收集管下方放置回收管。

⑥ 转EB的步骤与步骤(3)类似,需要修改EB缓冲液的体积为5 280 μL (55 $\mu\text{L} \times 8 \times 12$)。EB加到收集管静置10 min,使DNA充分的溶解以后,开启压力,压力值及时间设置同上,重复2次。

1.2.4 自动化文库制备方法的建立

(1) 磁珠平衡。通过一次流程,实现96管平衡磁珠的准备,磁珠的洗涤和平衡等操作在10 min内即可完成。整个实验过程中涉及到的原料管、96孔板的放置及体系如下($\times 8$ 代表8通道的枪头, $\times 12$ 代表循环

12 次)。

①平衡磁珠的缓冲液($0.5 \times \text{SSC buffer}$)放置在 100 mL 试剂槽(图 1(c)标记为试剂槽),链霉素亲和磁珠放置在 24 孔 EP 的 1~3 位。96 孔 PCR 板分别放置在 4 位微板载架的 site3(图 1(c)标记为 PCR),磁力架放置在 4 位微板载架的 site4(图 1(c)标记为 MG)。

②调取 8 根枪头吸取磁珠混悬液,吸液操作之前进行吹洗操作,使磁珠呈均匀悬浮状态,保证每一次的吸取都等量,吸取磁珠体积为 $20 \mu\text{L} \times 8$,喷液在 96 孔 PCR 板,循环 12 次,完成 96 孔的磁珠分配。

③磁珠分配后,利用 Roma 抓扳手将 96 孔 PCR 板移至磁力架上,静置 5 min,使磁液分离,分离后移除上清。此处需要设置移液两次,使上清液充分移除。

④Roma 抓扳手将磁力架上的 96 孔 PCR 板移回至 4 位微板载架的 site3 位置,进行二轮磁珠清洗,后续操作步骤同上,平衡好的磁珠最终回位到 site3,继续进行下一轮反应。

(2)杂交反应。杂交反应的体系主要在平衡好的磁珠及杂交液中进行,整个实验过程中涉及到的原料管、96 孔板的放置及体系如下($\times 8$ 代表 8 通道的枪头, $\times 12$ 代表循环 12 次)。

①杂交液放置在 100 mL 试剂槽 site2 (图 1(d)标记为试剂槽),制备好探针放置于 4 位微板载架的 site2 位(标记为 Trough),上一步平衡好的磁珠放置在 site3 板位(图 1(d)标记为 PCR),样品 DNA 保存于 96 孔 PCR 板中,放置在 4 位微板载架的 site4 位(图 1(d)标记为 MG)。

②调取 8 根枪头吸取杂交液体积为 $30 \mu\text{L} \times 8 \times 12$ ($\times 8$ 代表 8 通道的枪头, $\times 12$ 代表循环 12 次),移至 site3 板位 96 孔 PCR 板中。

③调取 8 根枪头吸取探针体积为 $10 \mu\text{L} \times 8 \times 12$,移至 site3 板位 96 孔 PCR 板中。

④调取 8 根枪头吸取样品 DNA 体积为 $10 \mu\text{L} \times 8 \times 12$,移至 site3 板位 96 孔 PCR 板中,混匀。

⑤将 96 孔板进行封膜置于 PCR 仪中杂交过夜。

⑥延伸反应。杂交反应完成后,涉及到一些上清的移除和冲洗操作与平衡磁珠的过程类似,配置延伸连接反应的母液可以参照探针扩增,其中涉及到需要更改的体积变量($\times 8$ 代表 8 通道的枪头, $\times 12$ 代表循环 12 次):连接缓冲液体积为 $2\,212.8 \mu\text{L}$ ($23.05 \mu\text{L} \times 8 \times 12$),连接酶体积为 $120 \mu\text{L}$ ($1.25 \mu\text{L} \times 8 \times 12$),dNTP 体积为 $24 \mu\text{L}$ ($0.25 \mu\text{L} \times 8 \times 12$)。

(4)一轮、二轮 PCR。对延伸连接后的模板进行两轮 PCR,可以参照探针扩增的程序编译模板,其中需要对反应体系中所涉及到的移液体积及分液体积进行修改,主要的体积变量为引物 primer1 体积为 $38.4 \mu\text{L}$

($0.4 \mu\text{L} \times 8 \times 12$)、引物 primer2 体积为 $38.4 \mu\text{L}$ ($0.4 \mu\text{L} \times 8 \times 12$)、dNTP 体积为 $115.2 \mu\text{L}$ ($1.2 \mu\text{L} \times 8 \times 12$)、Phusion 酶体积为 $38.4 \mu\text{L}$ ($0.4 \mu\text{L} \times 8 \times 12$)、 H_2O 体积为 $38.4 \mu\text{L}$ ($0.4 \mu\text{L} \times 8 \times 12$)。完成母液分配后,将 96 孔板进行封膜置于 PCR 仪中进行扩增。富集完成的产物参照探针纯化流程进行纯化,后随机选取 8 个文库根据 Qubit 定量结果将样品等量混合后使用 Illumina HiSeq X-ten 平台(PE150)进行高通量测序。

2 结果与分析

2.1 HD-Marker 文库自动化构建流程的搭建

2.1.1 吸放液针标准空间定位校准及移液参数的调节

通过细致定位和微调,样品容器的水平位置及 Z 轴相关参数都已建立(见表 1),确保了枪头能够在 Z_{travel} 的参数范围内平稳无阻碍的移动。对 Z 参数的设置除了包括 Z_{travel} (Z 移动),还包括 Z_{travel} 和 Z_{start} 的调节。其中 Z_{Dispense} 定位校准的作用是保证喷液时液体能够准确的打入目标孔位,但又不会使液体溅入到相邻孔位,通常在该位置管边缘向下 3~5 mm 处。另外 Z_{start} 的位置是配合页面探测功能使用的,即从该位置启用页面探针功能,吸液时根据页面的高度同步下降,有效的避免枪头外壁挂载液体造成浪费,此类吸液模式适用于内切酶、聚合酶等黏稠并且成本较高的试剂。其他的吸液操作多采用 Z_{max} 位置吸液,这样能够有效的保证吸取的液体足量,并且减少由于探测模式产生的误报情况。

每次移液所涉及到的吸液和喷液参数基本都有其特定的标准,该参数的设置与移液体积、液体质地、吸液喷液的位置等都息息相关。总体原则是:(1)基本 1 s 内完成吸液,对于体积较小的试剂,一般不宜太快;经测试对于 $10 \mu\text{L}$ (最小吸液体积)体积来说,吸液速度一般不宜超过 $40 \mu\text{L/s}$; (2)吸入液体后需要有一定的等待延迟时间,通常为 300~1 000 ms,如果吸取黏稠的液体需要适当的延长等待时间,保证吸取的液体足量; (3)喷液速度的设置通常在 $400 \sim 600 \mu\text{L/s}$,平稳有效的喷液速度的调试非常重要,如果喷液速度过慢,产生的压力不足以释放掉枪头内所有液体会造成移液的不精准,但如果喷液速度过快容易产生液体的喷溅造成污染; (4)不同液体质地其吸液和喷液需要根据具体情况采取合适的吸液高度。

2.1.2 各容器液体吸液的安全余量测定 尽管 Tecan 自动化液体处理系统在吸液时可以选择页面探测模式或者吸液页面跟随的模式,但实际操作中面对如酶、SDS 等黏稠或者容易起泡的液体时并不灵敏,容易产生报错。因此在吸取液体时,多数情况会选择在最低下探高度(Z_{max})的位置进行吸液,而此时若液体不足会

影响反应体系的准确性。为保证吸取液体足量准确，对不同液体装载容器进行安全余量的测定是至关重要的。经测试，1.5 mL EP 管，最低安全余量 25 μL ；12

孔母液槽中，单槽最低安全余量为 300 μL ；母液试剂配置槽 100 mL，最低安全余量 3 mL；96 深孔板等用于液体全部转移时，无需测量吸液安全余量。

表 1 自动化液体处理系统针头位置定位校准结果

Table 1 Needle position correction result of automated liquid handling system

样品容器 Consumables	生产厂家 Manufacturer	货号 Article number	X	Y	Z _{travel}	Z _{start}	Z _{max}
12 通道试剂槽 12 Channels-reagent reservoirs	生工生物工程股份有限公司	F619714-0001	345	234	456	700	1 090
100 mL 试剂槽 100 mL Reagent troughs	Tecan	10613049	435	346	238	300	1 110
96 孔 PCR 板 96 well-PCR plate	生工生物工程股份有限公司	F600960-0010	345	578	340	400	1 098
磁力架 Magnetic frame	深圳市博尔熙科技发展有限公司	MAG-96-11	234	987	546	600	1 087
磁力架+PCR 板	—	—	654	465	543	500	1 100

2.1.3 移液精准度测定 为了评价移液操作的可靠性，本文作者对 5、10、100 和 1 000 μL 的移液体积进行了准确度和精度的检测。其中移液体积为 1 000 μL 时，各加样针组间相对误差范围在 0.1%~0.9%之间，变异系数平均为 0.86%；移液体积为 100 μL 时，各加样针组间相对误差范围在 2.0%~5.6%之间，变异系数平均为 4.19%；移液体积为 10 μL 时，各加样针组间相对误差范围在 2.2%~4.7%之间，变异系数平均为 5.43%；移液体积为 5 μL 时，各加样针组间相对误差

范围在 1.2%~6.2%之间，变异系数平均为 8.09%。结果显示，Freedom EVO 100-8 自动化液体处理系统对于 10 μL 以上的移液体积的处理，其准确度及精密度都较为理想。但在 5 μL 体积下，其整体准确度及精密度稍显不足。因此在后续进行实验步骤设置时，会充分考虑到精准移液的体积限制，尽可能使移液体积均控制在 10 μL 以上。表 2 所示的是自动化液体处理系统的 8 个加样针的移液相对误差和变异系数。

表 2 自动化液体处理系统移液准确度和精密度

Table 2 Accuracy and precision of fluid transfer using automated liquid handling system

%

加样针 Sampling needle	5 $\mu\text{L}(n=12)$		10 $\mu\text{L}(n=12)$		100 $\mu\text{L}(n=12)$		1 000 $\mu\text{L}(n=12)$	
	相对误差 Relative error	变异系数 Coefficient of variation	相对误差 Relative error	变异系数 Coefficient of variation	相对误差 Relative error	变异系数 Coefficient of variation	相对误差 Relative error	变异系数 Coefficient of variation
Tip1	4.1	5.7	3.8	5.3	2.7	3.3	0.7	0.9
Tip2	2.5	9.1	3.9	4.5	4.4	4.2	0.1	0.8
Tip3	4.0	11.5	3.3	5.1	3.8	3.8	0.6	0.9
Tip4	4.2	7.5	2.5	5.3	3.1	4.6	0.9	1.0
Tip5	1.2	7.6	4.7	5.8	5.1	4.1	0.2	0.8
Tip6	6.2	6.6	2.2	5.6	5.6	4.8	0.5	0.9
Tip7	1.8	9.2	3.1	5.5	2.0	4.5	0.5	0.9
Tip8	4.5	7.5	4.2	6.3	2.2	4.2	0.3	0.7

2.2 HD-Marker 文库自动化构建流程的评估

2.2.1 探针扩增及纯化结果 利用自动移液工作站同步进行 96 管探针的扩增, 将母液分配至单管, 且液面的高度均匀一致, 并在 10 min 内完成了所有样品的 PCR 体系制备。待 PCR 反应结束后, 利用工作站真空抽滤法同时进行了 96 管 PCR 产物的纯化。电泳图检测结果显示所有管内均没有非特异性扩增(见图 2), 并

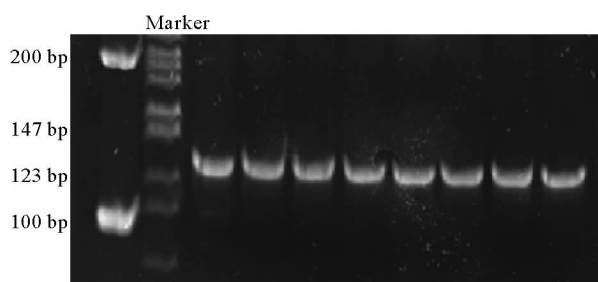


图 2 探针扩增检测电泳图

Fig.2 Polyacrylamide gel electrophoresis results of PCR product of oligo pools

且纯化后的产物具有均匀一致的亮度(见图 3); 浓度检测结果显示, 所有管内的浓度范围在 23.1~38.7 ng/ μ L 之间。96 管 PCR 纯化产物按照列或行分析时, 单因素方差分析结果表明行或列间无显著差异, 浓度分布基本均匀一致(见图 4), 表明自动化 PCR 扩增以及纯化的均匀性。

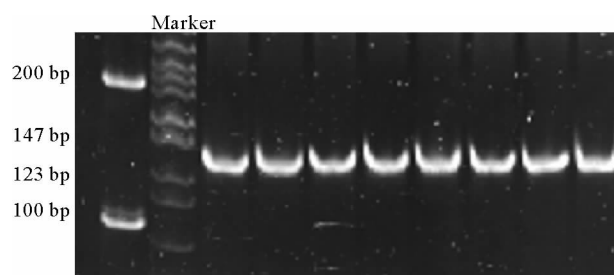
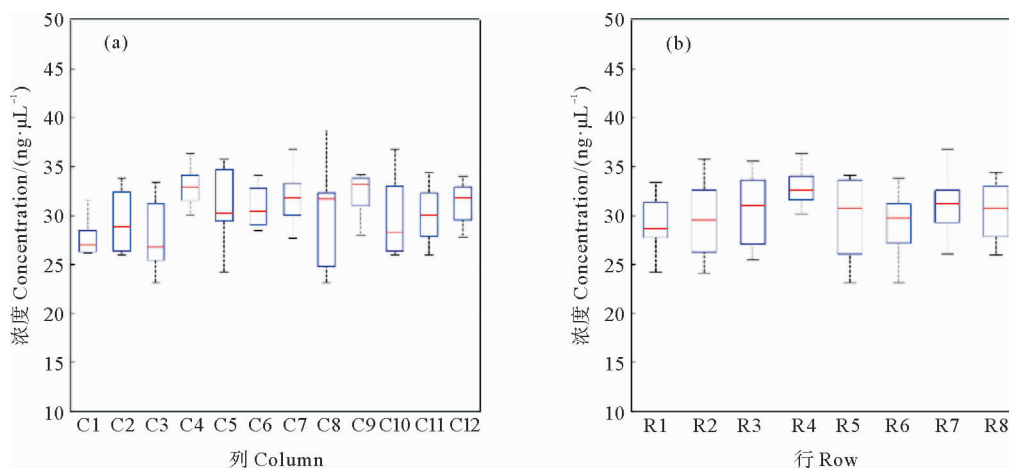


图 3 纯化产物 PAGE 检测图

Fig.3 Polyacrylamide gel electrophoresis results of purified product



((a)12 列之间的浓度分布箱型图; (b) 8 行之间的浓度分布箱型图。 (a) Boxplot of mean product concentration per column; (b) Boxplot of product concentration per row.)

图 4 探针纯化后列之间和行之间的浓度箱型图

Fig.4 Box diagrams between columns and rows after purification

2.2.2 自动化文库制备结果 利用调试好的自动化分型流程, 制备了 32 个虾夷扇贝 (*P. yessoensis*) DNA 样品的 HD-Marker 文库。其中成功建库的样品 29 个, 剩下的 3 个样品的文库无法满足高通量测序的要求, 自动化的建库成功率为 90.63%。最终获得的 29 个文库浓度分布在 8.9~10.6 ng/ μ L 之间, 且文库浓度均匀, 质量符合测序要求。

随机选取 8 个成功建库的样品进行 PE150 测序, 从特异性、捕获率及分型准确率等几个指标评价自动化制备 HD-Marker 文库测序结果的可靠性。结果显示, 在所有样品中特异性均在 80.0% 以上, 位点捕获率

均在 97% 以上, 分型位点比例能达到 100%。与标准 WGS 文库数据相比, 位点分型准确性也在 97%~100% 之间(见表 3)。

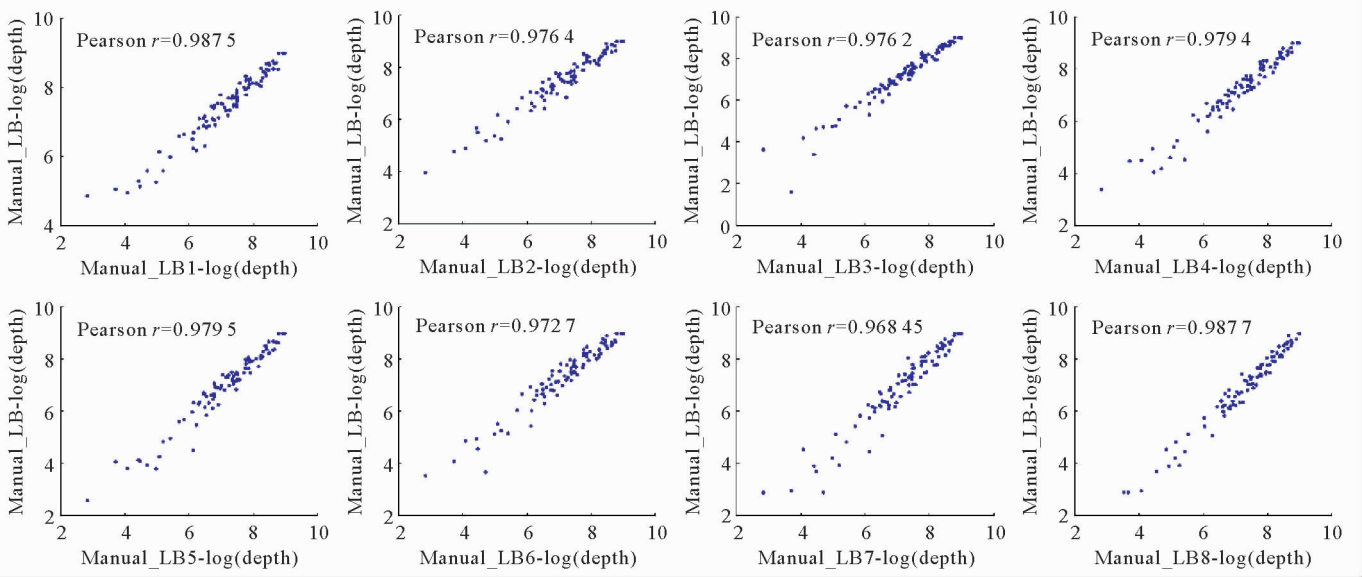
2.3 自动化制备文库与人工手动建库的比较

与人工构建文库的测序结果相比较, 自动化文库与人工构建的文库在位点测序深度(覆盖某位点的 reads 数目)上具有较高的一致性, 8 个自动化文库的 Pearson 系数^[19]均能达到 0.96(见图 5)。综上所述, 利用自动化移液工作站制备的 HD-Marker 基因分型文库能够获得较为稳定可靠的测序结果。

表 3 自动化分型文库的测序结果
Table 3 Sequencing result of automated genotyping library

文库 Library	测序 reads Raw reads	比对率 ¹ Mapping rate/%	位点捕获率 ² Capture rate/%	分型准确性 ³ Genotyping accuracy/%
LB1	458 806	80.6	97	97.0
LB2	538 646	80.0	100	98.0
LB3	582 047	86.6	99	99.0
LB4	513 708	89.9	98	97.0
LB5	495 595	89.8	100	100.0
LB6	461 025	85.7	100	98.0
LB7	550 432	87.9	99	99.0
LB8	536 249	85.0	98	100.0

注:1. 比对至靶位点处的 reads 与测序总 reads 的比例; 2. 捕获位点与总位点比例; 3. 与全基因组重测序技术分型一致的位点数与总位点数的比例。
Note: 1.Ratio of mapped reads to total reads; 2.Ratio of captured loci to total loci; 3.Ratio of loci consistent with the whole genome resequencing to the total loci.



(Manual_LB-log(depth):人工制备文库测序深度的 log 值;LB1-LB8:8 个自动化制备的文库。Manual_LB-log(depth):Log value of sequencing depth of manual library; LB1-LB8:8 automated libraries.)

图 5 自动化制备的文库与人工建库相比测序深度的一致性

Fig.5 The correlation of sequencing depth between automated library and manual library

当样品数量较大(>96 个样品)时基于自动化工作站的自动化文库制备所需时间仅为手动的 1/7(见表 4),尽管分步之间的衔接仍需人工干预,但对人工需求相比全人工手动建库大大降低。其次,机器预编译的程序原则上不会存在漏加、错加和少加等情况,构建的文库具有较高的均匀性,特别适用于大批量样品的基

因分型文库的构建。但相对于人工手动建库,自动化文库制备所需成本较高,约为手动建库的 1.1 倍,但总体而言,自动化文库的构建能够有效的减少人力,并且获得均一稳定的建库结果,在短时间内快速实现大量样品的高通量测序文库构建。

表 4 自动化制备文库与手动建库的比较

Table 4 The comparison between automated and manual construction of HD-Marker library

步骤 Step	自动化制备文库 Automated library construction/min	人工手动制备文库 Manual library construction/min
扩增(样品 192)Amplification (sample 192)	8	150
纯化(样品 96)Purification (sample 96)	28	480
酶切(样品 96)Digestion (sample 96)	5	30
平衡磁珠(样品 96)Magnetic beads balancing (sample 96)	40	120
杂交(样品 96)Hybridization (sample 96)	12	120
延伸(样品 96)Extension (sample 96)	40	180
一轮及二轮 PCR(样品 96) First and second PCR amplification (sample 96)	30	150
总计 Total	163	1 230
稳定性 Stability	相对稳定、不同样品间差异较小, 几乎不存在错加、多加等情况	存在误差、或者错加、 多加、少加等情况
成本 Cost	较高,约为人工 手动建库的 1.1 倍	相对自动化低 (可参考 Lv 等 ^[15])

3 讨论

短时高效获取大量样品的 SNP 分型信息是实现动植物全基因组选择育种的关键要素。除模式生物、人及畜牧动物等可依靠基因芯片快速获取分型信息^[9-12], 大量生物尤其非模式生物主要通过二代测序平台的检测分析获取分型结果。但其首要的步骤就是构建高质量的测序文库,而测序文库的构建涉及大量手工操作,尤其处理大规模样品是一项十分耗时耗力的工作。本研究首次建立了基于自动化移液工作站的自动化高通量 HD-Marker 基因分型文库构建流程,并对其进行了评估。

本研究建立的自动化文库构建流程一次反应可以同时运行 96 个样品的文库构建,整体运行时间仅为人工操作的 1/7。为解决大量样品纯化效率低的问题,本研究建立了基于 96 孔真空抽滤法的 PCR 产物纯化方式,加样和纯化步骤全程可自动化完成,改变以往单个离心柱回收进行纯化的方式,能够大幅度提高探针制备的效率。通过电泳图和浓度纯度的结果,判断出不同管之间产物的得率无显著差异,说明依托自动化移液工作站建立的真空抽滤法适用于 HD-Marker 的探针制备环节。另外,文库制备的全部流程反应中均包含链霉素磁珠,因此磁珠的平衡、磁珠与液体的混匀以及磁液分离等操作都是影响实验的关键步骤。因此,本文作者对涉及到磁珠操作时的吸喷液速度及次数进行了优化调整,既能保证磁珠充分悬浮于液体中,又能

得到较好的磁液分离效果。

本研究还比较了使用自动化移液工作站构建基因分型文库与人工手动建库的优缺点,发现自动化移液工作站构建基因分型文库有诸多优点:(1)准确:自动化工作站按照设定好的程序自动进行实验,不会出现漏加或者错加等实验错误。(2)稳定:重复进行多次实验,结果均一性程度高。(3)高效:自动化工作站设定好程序,可以无人值守工作,提高实验室工作效率。(4)实验多元化:全自动化移液工作站通常都整合丰富多样的实验设备,如振荡器 Te-Shake,磁珠提取模块 Te-MagS,真空提取模块 Te-Vac 等,在药物和抗体筛选^[20-21]、核酸蛋白提取^[22]、免疫组化^[23]、酶免实验^[24]等方向应用广泛。但自动化移液工作站的使用也存在一定的不足,针对性的建议如下:(1)原装进口枪头的成本较高,后续可在某些可循环利用的实验环节将枪头丢弃操作优化为放回枪头盒中,或者通过一些不与样品 DNA 直接接触的方式进一步降低耗材使用的成本;(2)机械吸液无法像人工充分吸取试剂槽底部的试剂,因此在原料的损耗上相比人工要多,但可以通过试剂的回收再利用降低试剂的损耗。(3)本研究使用的 Tecan Freedom EVO 自动化液体处理系统配备 1 mL 注射器,但从本研究结果中可看出在低于 10 μ L 体积的定量中,准确性及精密度不够理想,因此流程建立时需尽可能规避这部分误差。

总体来说,基于 TECAN 工作站建立的自动化 HD-Marker 文库构建流程,可依托预编译的控制程序

实现标准化自动化地 HD-Marker 基因分型文库构建, 并且能够获得均匀稳定的建库质量, 保证整体测序的质量。可以有效地代替人工操作, 避免和减少人为操作导致的实验误差, 并提高批量样品并行处理的效率, 在动植物的遗传育种研究中具有重要的应用价值。

参考文献:

- [1] Ribaut J M, de Vicente M C, Delannay X. Molecular breeding in developing countries: Challenges and perspectives [J]. *Current Opinion in Plant Biology*, 2010, 13(2): 213-218.
- [2] Meuwissen T H, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps[J]. *Genetics*, 2001, 157(4): 1819-1829.
- [3] Davey J W, Hohenlohe P A, Etter P D, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing[J]. *Nature Reviews Genetics*, 2011, 12(7): 499-510.
- [4] Baird N A, Etter P D, Atwood T S, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers[J]. *PLoS One*, 2008, 3(10): e3376.
- [5] Elshire R J, Glaubitz J C, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species[J]. *PLoS One*, 2011, 6(5): e19379.
- [6] Wang S, Meyer E, McKay J K, et al. 2b-RAD: A simple and flexible method for genome-wide genotyping[J]. *Nature Methods*, 2012, 9(8): 808-810.
- [7] Wang S, Liu P, Lv J, et al. Serial sequencing of isolength RAD tags for cost-efficient genome-wide profiling of genetic and epigenetic variations[J]. *Nature Protocols*, 2016, 11(11): 2189-2200.
- [8] Andrews K R, Good J M, Miller M R, et al. Harnessing the power of RADseq for ecological and evolutionary genomics[J]. *Nature Reviews Genetics*, 2016, 17(2): 81-92.
- [9] Pease A C, Solas D, Sullivan E J, et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis[J]. *Proceedings of the National Academy of Sciences*, 1994, 91(11): 5022-5026.
- [10] Cañas-álvarez J J, González-Rodríguez A, Munilla S, et al. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip[J]. *Journal of Animal Science*, 2015, 93(11): 5164-5174.
- [11] Zhu B, Li Q, Liu R, et al. Genome-wide association study of H/L traits in chicken[J]. *Animals (Basel)*, 2019, 9(5): 260.
- [12] Junqueira V S, Cardoso F F, Oliveira M M, et al. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models[J]. *Journal of Animal Breeding and Genetics*, 2017, 134(1): 14-26.
- [13] Albrechtsen A, Nielsen F C, Nielsen R. Ascertainment biases in SNP chips affect measures of population divergence[J]. *Molecular Biology and Evolution*, 2010, 27(11): 2534-2547.
- [14] Jiang Z, Wang H, Michal J J, et al. Genome wide sampling sequencing for snp genotyping: Methods, challenges and future development[J]. *International Journal of Biological Sciences*, 2016, 12(1): 100-108.
- [15] Lv J, Jiao W, Guo H, et al. HD-Marker: A highly multiplexed and flexible approach for targeted genotyping of more than 10 000 genes in a single-tube assay[J]. *Genome Research*, 2018, 28(12): 1919-1930.
- [16] Stangegaard M, Frøslev T G, Frank-Hansen R, et al. Automated extraction of DNA from blood and PCR setup using a Tecan Freedom EVO liquid handler for forensic genetic STR typing of reference samples[J]. *Journal of Laboratory Automation*, 2011, 16(2): 134-140.
- [17] 匡金枝, 聂同钢, 杨智, 等. 法医物证 DNA 自动化检验技术体系的研究[J]. *中国法医学杂志*, 2009, 24(3): 164-167.
Kuang J Z, Nie T G, Yang Z, et al. A study of automation workstation for the DNA extraction of forensic biological evidence[J]. *Chinese Journal of Forensic Medicine*, 2009, 24(3): 164-167.
- [18] Vogeser M, Spöhrer U. Automated processing of whole blood samples for the determination of immunosuppressants by liquid chromatography tandem-mass spectrometry[J]. *Clinical Chemistry and Laboratory Medicine*, 2006, 44(9): 1126-1130.
- [19] Sedgwick P. Pearson's correlation coefficient[J]. *British Medical Journal*, 2012, 345: e4483.
- [20] Ye Q Z, Nan F J, Hu L Y, et al. High-throughput screening for human collagenase1 inhibitors[J]. *Acta Pharmacologica Sinica*, 2000, 21(11): 1021-1026.
- [21] Wierling P S, Bogumil R, Knieps-Grünhagen E, et al. High-throughput screening of packed-bed chromatography coupled with SELDI-TOF MS analysis: Monoclonal antibodies versus host cell protein[J]. *Biotechnology and Bioengineering*, 2007, 98(2): 440-450.
- [22] 严江伟, 王顺霞, 任贺, 等. 应用自动化工作站提取常见生物样本 DNA[J]. *中国法医学杂志*, 2007, 2(22): 73-75.
Yan J W, Wang S X, Ren H, et al. Study on the application of the automation workstation in DNA extraction of biological samples[J]. *Chinese Journal of Forensic Medicine*, 2007, 2(22): 73-75.
- [23] Rumballe B, Georgas K, Little M H. High-throughput paraffin section *in situ* hybridization and dual immunohistochemistry on mouse tissues[J]. *Cold Spring Harbor Protocols*, 2008(8): pdb. prot5030.
- [24] 张鹏, 魏巍, 李玉燕, 等. 使用全自动酶免工作站进行酶联免疫吸附试验的效果评估[J]. *家禽科学*, 2020(7): 10-14.
Zhang P, Wei W, Li Y Y, et al. Effect evaluation of enzyme-linked immunosorbent assay with automatic ELISA workstation[J]. *Poultry Science*, 2020(7): 10-14.

Establishment and Evaluation of an Efficient and Automated HD-Marker Library Construction Procedure

Yang Zhihui¹, Lv Jia¹, Liu Pingping^{1,2}, Zhang Tianqi¹,
Huang Xiaowen¹, Ni Ping¹, Bao Zhenmin^{1,3,4}, Wang Shi^{1,2,4}

(1. Key Laboratory of Marine Genetics and Breeding, Ministry of Education, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China; 2. Laboratory for Marine Biology and Biotechnology, Pilot Qingdao National Laboratory for Marine Science and Technology(Qingdao), Qingdao 266237, China; 3. Laboratory for Marine Fisheries Science and Food Production Processes, Pilot Qingdao National Laboratory for Marine Science and Technology(Qingdao), Qingdao 266237, China; 4. Laboratory of Tropical Marine Germplasm Resources and Breeding Engineering, Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China)

Abstract: In view of the challenge of rapidly completing genotyping of large quantities of samples in a short period of time with the application of HD-Marker in breeding research, we established a standardized automated HD-marker (High-throughput, sequencing-based GoldenGate approach) library preparation procedure based on Tecan Freedom EVO workstation and evaluated the reliability of this method in library construction. we established a standardized automated HD-marker library preparation procedure based on Tecan Freedom EVO workstation and evaluated the reliability of this method in library construction. The results indicated the high uniformity of amplification products with the concentration ranging from 23.1 ng/ μ L to 38.7 ng/ μ L. In all the tested samples, 90.63% of them could obtain clear library bands, indicating the high successful rate for library construction. The analysis of the sequencing library revealed a high capture rate (more than 97%), calling rate (100%) and genotyping accuracy (97%~100%, compared with the result of whole genome resequencing). The automatic library construction system can simultaneously run 96 libraries, while only takes 1/7 of the manual operation time. Altogether, the automatic HD-Marker library preparation procedure has been successfully implemented for the first time, providing an efficient and valuable genotyping platform for implementation of genome selection.

Key words: automatic liquid handling workstation; HD-Marker; library construction; genotyping; genomic selection

责任编辑 高 蓓