



评述

杂交捕获测序技术应用与研究进展

陈明林, 骆佳, 匡卫民*, 于黎*

云南大学省部共建云南生物资源保护与利用国家重点实验室, 生命科学学院, 昆明 650500

* 联系人, E-mail: yuli@ynu.edu.cn; kuangwm0714@ynu.edu.cn

收稿日期: 2024-10-19; 接受日期: 2025-01-24; 网络版发表日期: 2025-02-27

国家重点研发计划(批准号: 2022YFC2602500)、国家自然科学基金(批准号: 32260133)、云南省基础研究项目(批准号: 202301AT070234, 202301AU070214)、云南省教育厅科学研究基金(批准号: 2024Y016)、中央引导地方科技发展资金(批准号: 202407AB110004)、云南省中越濒危非人灵长类保护遗传学国际联合实验室(批准号: 202403AP140011)、云南省兴滇英才支持计划科学家工作室和云南省兴滇英才支持计划青年人才项目资助

摘要 随着测序技术的不断发展, 杂交捕获测序技术通过设计特定的生物素化探针与目标DNA序列进行杂交, 对特定基因或基因组目标区域进行富集和测序, 实现了利用较少的数据量就能对目标区域变异位点进行快速、准确地检测, 大大降低了测序成本. 本文系统综述了杂交捕获测序技术在动物、植物、微生物以及医学研究领域中的广泛应用和研究进展, 揭示了杂交捕获测序技术在古DNA研究、非损伤性样本宿主DNA富集、疾病、病毒和病原体检测等研究领域的巨大潜力. 此外, 文章探讨了杂交捕获在应用和发展过程中遇到的问题与解决方法, 并对杂交捕获技术的优化方向进行了展望, 以期为该技术的进一步应用和发展提供参考.

关键词 杂交捕获测序, 生物素化探针, 高通量测序, DNA富集, DNA文库

随着测序技术的不断发展, 高通量测序技术(high-throughput sequencing)已成为现代生命科学获得大规模遗传信息和基因表达调控等基础遗传数据的重要手段, 因其可以通过高通量、快速、高效且经济的方式获得全基因组数据, 所以目前已被广泛应用于遗传学^[1]、生态学^[2]、进化生物学^[3]以及医学^[4-6]等研究领域^[7]. 然而, 并非所有的研究都关注物种的全基因组, 特定的目标基因组区域也可以提供相对完整和丰富的生物信息, 所以相较于全基因组测序, 靶向测序技术^[8]可直接对样本中的目标基因组区域进行分离测序, 包括特定基因、线粒体DNA、外显子组以及单条染色体等, 目前已经广泛应用于基因型鉴定、SNP分析、病

原体检测等研究^[9,10]. 靶向测序技术降低了DNA样本的复杂性, 利用较少的数据量就能实现对目标区域变异位点的快速检测, 使得靶向测序技术能以更具成本和时间效益的方式应用于复杂DNA样本的大规模研究.

根据获取目标区域方式的不同, 靶向测序技术分为多重PCR靶向测序技术和杂交捕获测序技术. 多重PCR靶向测序技术, 又称多重扩增子测序技术, 是针对目标区域设计多重特异性PCR引物进行扩增富集并进行高通量测序的技术, 这种方法通常适用于几千个变异位点的检测或者1 Mb以内DNA区域的扩增富集^[11,12]. 然而, 杂交捕获测序技术是使用特异性探针和

引用格式: 陈明林, 骆佳, 匡卫民, 等. 杂交捕获测序技术应用与研究进展. 中国科学: 生命科学, 2025, 55: 773–789

Chen M L, Luo J, Kuang W M, et al. Application and research advances of the hybridization capture sequencing technology (in Chinese). *Sci Sin Vitae*, 2025, 55: 773–789, doi: [10.1360/SSV-2024-0303](https://doi.org/10.1360/SSV-2024-0303)

目标基因组区域杂交, 捕获富集DNA片段后进行高通量测序的技术, 其实现了对特定基因或特定基因组区域的富集和测序, 具有高灵敏度和高特异性的特征^[13]. 相较于多重扩增子测序技术, 杂交捕获测序技术可以富集到更多的基因组目标区域, 获得更全面的变异类型, 还降低了非特异性扩增^[9]. 基于杂交捕获测序技术高效的捕获效率, 该测序技术已被广泛应用在许多研究领域.

根据捕获杂交时探针的状态不同, 杂交捕获方法分为固相杂交捕获法(或微阵列杂交法)和液相杂交捕获法^[14](图1). 固相杂交捕获法将带有标记的探针预先固定在固体支持物上, 目标DNA与探针进行杂交反应后, 使杂交目标DNA片段留在固相支持物上, 待杂交后先洗脱未杂交上的DNA片段, 再将与探针杂交的DNA洗脱下来, 最后构建测序文库并进行高通量测序^[15,16]. 相反, 液相杂交捕获法在液相环境中反应, 目

标DNA片段和带有生物素标记的游离探针杂交, 然后通过生物素亲和素的反应使目标DNA片段吸附在带有亲和素的磁珠上, 再通过洗涤和纯化过程去除非目标DNA片段, 最终实现目标DNA捕获^[17]. Mamanova等人^[14]通过测试并比较了液相杂交捕获法和固相杂交捕获法的DNA捕获效率, 发现从液相杂交捕获法中获得的DNA序列均一性和特异性均高于固相杂交捕获法. 此外, 液相杂交捕获法可以用于大批量样本的捕获, 降低了捕获成本. 鉴于液相杂交捕获法易于操作、成本较低的优势, 适用于较大规模样本的DNA捕获和测序, 该技术已经成为目前应用最广泛的捕获技术之一.

目前, 杂交捕获技术已被广泛应用于动物、植物、微生物、病毒和医学等研究领域(图2). 在动物基因组学研究领域中, 杂交捕获技术常用于研究特定基因或基因组区域的变异, 特别是在古DNA样本和非损伤性

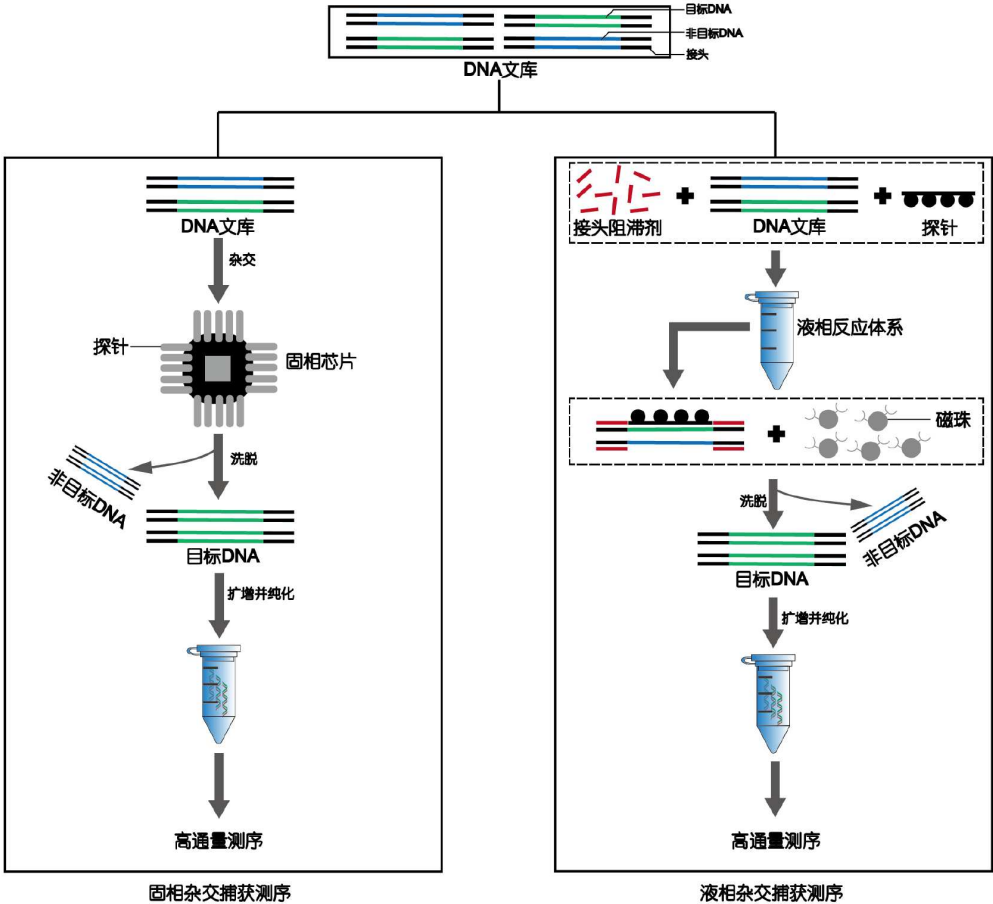


图1 杂交捕获测序技术原理示意图

Figure 1 Schematic illustrations of the hybridization capture sequencing technologies



图 2 杂交捕获测序技术应用图示. 图中的4个部分分别代表杂交捕获测序技术在动物(红色)、植物(绿色)、微生物(紫色)以及医学上(蓝色)的应用

Figure 2 Illustration of the applications of the hybridization capture sequencing technology. The four sections of the circle represent the applications of the hybridization capture sequencing technology in the four areas. Red, animals; green, plants; purple, microorganisms; blue, medicine

样本(如粪便)中的应用. 在植物学研究领域中, 该技术常作为植物基因组测序的替代方案, 进一步进行植物物种的系统发育、遗传多样性以及作物改良等研究. 在微生物学研究领域中, 杂交捕获技术常应用于微生物群落结构分析, 通过捕获和测序特定的微生物基因组区域, 研究生态系统中的物种多样性和功能. 在医学研究领域中, 杂交捕获技术常应用于遗传病检测、肿瘤筛查、病毒和病原体检测中. 本文回顾并综述了杂交捕获技术在这些领域的发展过程和应用现状, 揭示了杂交捕获技术的巨大潜力, 探讨了杂交捕获在应用和发展过程中遇到的问题(表1), 并对杂交捕获技术进行了展望, 以期为该技术的进一步应用和发展提供借鉴和参考.

1 杂交捕获测序技术在动物研究中的应用与进展

1.1 杂交捕获测序技术在动物古DNA中的应用与进展

古DNA样本是研究物种起源, 揭示物种进化和迁

移历史的重要生物遗传材料, 古DNA研究更是一直以来的研究热点^[49]. 因此, 杂交捕获测序技术在古DNA研究中的应用发展迅速. 在早期的研究中, 杂交捕获测序技术聚焦于古代样本的线粒体基因组、特定基因组区域或单条染色体捕获. 例如, Fu等人^[50]从中国北京的天元洞中提取了一个约4万年前的早期现代人的DNA, 利用一个欧洲人的线粒体DNA作为探针捕获了此样本的线粒体基因组, 同时也使用单链生物素化标记的DNA探针对此样本的21号染色体非重复区域以及核基因组多态性位点进行杂交捕获. 最终获得了16.5 kb的线粒体基因组全序列、约30 Mb长度的21号染色体序列和核基因组3000多个多态性位点, 这些古DNA数据为理解早期现代人在欧亚大陆的扩散情况提供了新的遗传学证据, 并且对于探讨现代人的起源与演化提供了重要的遗传学数据. Vilstrup等人^[18]使用MySelect目标富集试剂盒(MYcroarray, USA)对3个已灭绝的马科动物的线粒体基因组进行了杂交捕获, 并获得了完整的线粒体基因组序列, 通过古线粒体基因组序列阐明了现代马、斑马和驴的演化历史. 在成功捕获较小区域DNA序列后, 研究者们尝试在全基因组水平的捕获, 以获得更多、更完整的全基因组数据. 例如, Carpenter等人^[19]研发了一种从古DNA样本中捕获全基因组序列的方法. 研究者利用现代人的DNA通过体外转录形成了生物素化标记的RNA探针, 并利用探针对保加利亚铁器时代和青铜时代古人的牙齿、秘鲁木乃伊骨骼和丹麦青铜时代头发样本的DNA进行了全基因组杂交捕获, 发现古人DNA平均含量从1.2%提高到了59%, 提示全基因组杂交捕获方法可从含有极低水平内源性DNA的样本(0.04%~6.2%)中获取较多古DNA全基因组数据. 此外, 该研究还表明捕获的富集效率受样本初始文库复杂度的影响, 而通过单链DNA建库方法可以保留更多的短片段DNA^[51,52], 从而有效提升文库的复杂度. 同样, Enk等人^[20]利用现代亚洲象的基因组信息, 设计了一套全基因组的生物素化RNA探针, 对已灭绝的猛犸象DNA进行全基因组捕获和富集. 研究结果显示, 相比于捕获前猛犸象DNA含量为4.4%~73.8%, 捕获后猛犸象DNA的含量提高到了78.3%~82.8%, 猛犸象DNA含量提高了1.06~18.65倍. 尽管全基因组杂交捕获对古DNA研究有着巨大的前景, 但捕获效率仍受到诸多因素的影响, 如古DNA含量低、DNA降解严重等. Ziesemer等人^[21]使用MYbaits

表 1 杂交捕获测序技术研究进展

Table 1 Advances in the hybridization capture sequencing technology

应用类型	杂交策略/试剂盒	捕获样本/区域	重要结果	优化建议	参考文献
动物	1.5 µg起始DNA量; MYselect靶向富集试剂盒; 磁珠法回收DNA	马骨骼古DNA样本/线粒体	捕获获取到了完整的线粒体基因组且通过古线粒体基因组序列阐明了现代马、斑马和驴的演化历史	—	[18]
	捕获前0.04%~6.2%的内源性DNA含量; 使用T7接头构建人类基因组DNA文库; 65°C下杂交66 h; 磁珠法回收DNA	人类牙齿、骨骼和头发的古DNA样本/全基因组	内源性DNA含量从1.2%提高到59%, 但仅获得了低覆盖度基因组数据, 无法进行有效的基因型检测	提高样本初始DNA文库复杂度; 使用单链DNA建库可以保留更多的DNA片段	[19]
	MYcroarray (Ann Arbor, MI)	猛犸象遗骸/全基因组	内源性DNA含量提高了1.06~18.65倍, 平均插入长度显著增加, 但重复序列占捕获序列的53.4%~63.6%	在杂交捕获过程中使用靶生物体特异性阻滞剂、减少捕获后的PCR循环	[20]
	使用NEBNext DNA Library Prep Master Set制备DNA文库; MYBait人类全基因组富集试剂盒; 65°C下杂交16 h; 磁珠法回收DNA	牙结石和牙本质/全基因组	内源性DNA含量提高了近4倍, 但低于线粒体基因组富集(2×~139×); 相比捕获前, 捕获后的DNA污染、平均片段长度、冗余率、GC含量均较高, DNA损伤模式没有变化	增加起始DNA的量、优化杂交温度和孵育时间、增加捕获次数	[21]
	捕获前1.8%的内源性DNA含量; 1.5 µg起始DNA含量; Agilent SureSelect靶向富集系统; 探针长度为120 bp; DNA片段长度为200 bp; 65°C下杂交26 h; 二轮捕获; 磁珠法回收DNA	黑猩猩粪便样本/21号染色体、X染色体和线粒体基因组	捕获后成功获得1.5 Mb基因组序列, 且捕获后基因型与同一个体血液样本基本相同, 首次表明了杂交捕获技术应用于非损伤性样本进行DNA富集的可行性	DNA捕获方法与多种测序方式相结合	[22]
	在探针构建步骤中加入双链体特异性核酸酶进行消化; 捕获片段长度为200-400 bp; 65°C下杂交48 h; 磁珠法回收DNA	狒狒粪便样本/全基因组	捕获后内源性DNA含量提高了近40倍, 获得的基因型数据与来自同一个体的高质量、未捕获样本的基因型数据高度一致, 但仅产生低覆盖率数据	—	[23]
	使用覆盖整条线粒体的重叠长片段PCR产物构建探针; 65°C下杂交48 h; 磁珠法回收DNA	动物/植物线粒体基因组、叶绿体基因组	获得完整的线粒体基因组与叶绿体基因组序列, 捕获后线粒体DNA序列平均提高了310倍以上	结合混池杂交捕获方式提高捕获效率	[24~27]
	使用KAPA文库制备试剂盒制备DNA文库; DNA文库片段长度为200 bp; 47°C下杂交36 h; 混池捕获	黑猩猩粪便样本/外显子组	两轮捕获后的富集因子、捕获特异性均增加, 捕获敏感性无明显变化, 而文库复杂度呈下降趋势; 非目标DNA的序列从一轮捕获的34.15%降低到了二轮捕获的7.83%	二轮捕获可提高捕获效率	[28]
	Agilent SureSelect靶向富集系统; DNA文库片段长度为200 bp; 混池捕获	黑猩猩粪便样本/外显子组	一轮捕获后内源性DNA含量为55%(44.4%~66%), 两轮捕获后内源性DNA含量为51.4%(33.8%~63.1%)	内源性DNA含量大于2%的样品, 使用一轮捕获方法优于二轮捕获方法; 增加DNA的起始量可提高杂交捕获的富集率	[29,30]
	IDT xGen Exome Research Panel (version 1.0); DNA片段长度为150 bp; 磁珠法回收DNA	黑猩猩粪便、尿液样本/外显子组	捕获后尿液样本内源性DNA含量为96.4%, 而粪便样本捕获后的内源性DNA含量仅为44.9%	内源性DNA含量较高的非损伤性样本可提高捕获效率	[31]
植物	Agilent SureSelect靶向富集系统	杨树/外显子组	捕获到了20.76 Mb的杨树基因组序列, 但捕获到了较多的重复基因	混池杂交捕获可提高捕获效率	[32]
	Agilent SureSelect靶向富集系统	松树/外显子组	平均捕获到95.47%的目标序列, 但在捕获过程中发现, 覆盖多个外显子的探针捕获效率严重下降	不将探针设计于外显子之间的区域	[33,34]
	Roche Nimblegen SeqCap EZ 富集方案	云杉/外显子组	捕获到18.5%~21%的目标基因组区域, 但仍有41%~44%的序列不属于目标物种	避免低同源目标区域、增加探针密度可提高捕获效率	[35]

(表1续)

应用类型	杂交策略/试剂盒	捕获样本/区域	重要结果	优化建议	参考文献
微生物	HiSpOD软件设计生物素化RNA探针; 65°C下杂交24 h	土壤样本/ <i>mcrA</i> 基因、 <i>rdhA</i> 基因	<i>mcrA</i> 序列占比从直接测序的0.003%提高到了捕获后测序的40%	—	[36,37]
	Roche NimbleGen SeqCap EZ富集方案	湖泊沉积物样本/酶类基因	捕获后的基因组中与有机质降解相关的酶类基因是宏基因组测序的5倍以上.	—	[38]
	KASpOD和PhylArray软件设计探针; 65°C下杂交24 h	土壤样本/16SrRNA基因序列	杂交捕获富集到了仅占群落0.00006%的微生物, 且16SrRNA基因序列的比例从捕获前的0.09%~0.41%提高到了捕获后的58.74%~97.30%	—	[39]
	xGen Lockdown Reagents (IDT); 连续两轮富集	人类粪便样本/16SrRNA基因序列	内源性DNA含量从捕获前小于1%提高到了捕获后的25.4%~44.2%	—	[40]
医学研究	NimbleGen和SureSelect富集系统	人类患者血液样本/遗传疾病相关的致病突变	准确检测到遗传疾病中的致病突变, 捕获后的DNA序列中>99.7%的碱基的覆盖深度超过100×, 且所有变异均显示基于捕获的高通量测序和Sanger测序具有100%的一致性	—	[41,42]
	Agilent-SureSelectQXT和Illumina-Nextera捕获试剂盒	人类血液和口腔细胞样本/口腔细胞中的258个基因	准确检测到相关基因, 且Sanger测序和杂交捕获测序结果高度一致性	—	[43]
	HaloPlex杂交捕获试剂盒	患者血液样本/肌萎缩侧索硬化症的致病基因	捕获后96.15%的目标区域实现了10×的覆盖率, 94.84%的目标区域实现了20×的覆盖率	—	[44]
	Agilent SureSelect靶向富集系统	患者脑脊液样本/人类疱疹病毒基因组	获得了具有高覆盖率(>94%)的病毒全长基因组	—	[45]
	ViroCap探针集; NimbleGen SeqCap EZ富集系统;	患者血液样本/感染脊椎动物的34个DNA和RNA病毒家族	捕获后共检测到32种病毒, 相比捕获前增加了11种, 32个基因组中有12个基因组的基因组覆盖率>90%, 其中8种病毒获得了完整的基因组	—	[46]
	VirCapSeq-VERT探针集; Roche NimbleGen SeqCap RNA富集系统; 47°C下杂交12 h	患者血液样本/207个感染脊椎动物病毒类群的病毒基因组	富集后的病毒序列相比未捕获提高了100~1000倍, 且几乎所有病毒都获得了全长基因组序列(>95%)	—	[47]
	ViroFind探针集, 65°C下杂交18 h	大脑样本/535种可能导致人畜共患的DNA和RNA病毒	与捕获前测序相比, 临床样本中存在的病毒序列提高了127倍, 并在患者体内发现了高度复杂的多瘤病毒种群	—	[48]

WGE kit (Arbor Biosciences, USA)全基因组生物素化RNA探针对24个考古人类牙本质和牙结石样本进行全基因组杂交捕获. 相比于捕获前, 捕获后古人DNA含量提高了近4倍, 但其富集效率远不如对线粒体基因组或特定基因组区域的捕获效率高. 具体而言, 捕获后从牙结石样本中获得了相比捕获前6~139倍的线粒体基因组, 从牙本质样本中获得了相比捕获前2~99倍的线粒体基因组. 为了进一步探究影响全基因组杂交捕获的因素, 作者比较了捕获前后的DNA污染、平

均片段长度、冗余率、GC含量和DNA损伤模式等. 结果发现, 相比捕获前, 捕获后的DNA污染、平均片段长度、冗余率、GC含量均较高. 因此, 作者认为基因组的大小和复杂性、核DNA的拷贝数、基因组中重复片段以及与外源生物(如微生物)同源区域等因素可能影响捕获的富集效率, 建议通过增加起始DNA的量、优化杂交温度和孵育时间、进行多轮捕获富集等方式提高全基因组杂交捕获的富集效率^[53].

综上所述, 杂交捕获测序技术在动物古DNA研究

中具有广泛的应用,极大促进了古基因组学的发展.然而,由于古DNA长期的埋藏和周围环境因素的影响,导致DNA高度碎片化、化学修饰复杂以及内源性DNA含量极少,影响杂交捕获富集效果.此外,古DNA样本极易受到现代人类DNA以及环境中其他生物DNA的污染,在杂交捕获实验过程中,从样本采集、处理到实验环境等各个环节都需要极其严格的防污染措施.因此,优化古DNA提取方法、文库构建方法、提高探针特异性、优化杂交反应条件以及加强实验室污染防控可进一步提高杂交捕获技术在古DNA研究中的效率和准确性.

1.2 杂交捕获测序技术在动物非损伤性样本中的应用与进展

鉴于杂交捕获测序技术能够成功应用于古DNA富集,研究者也将该技术应用于动物非损伤性样本(如粪便、尿液、脱落毛发等)的内源性(或宿主)DNA捕获和富集.例如, Perry等人^[22]使用Agilent SureSelect设计了长度为120 bp的探针,对6只圈养黑猩猩的粪便样本进行内源性DNA杂交捕获.最终成功获得了1.5 Mb基因组目标序列,包括21号染色体、X染色体和线粒体基因组序列,该研究首次表明了杂交捕获技术应用于非损伤性样本进行DNA富集的可行性.同样, Snyder-Mackler等人^[23]利用基因组生物素化全RNA探针针对狒狒的粪便DNA样品进行了杂交捕获,发现相较于捕获前,捕获富集后的宿主DNA含量提高了近40倍.同时,作者对同一个体的粪便DNA样本捕获测序数据和血液样本DNA的数据进行基因型检测和比较分析,结果发现两者基因型一致性高达77%~86.5%,表明了杂交捕获数据的准确性.

然而,由于动物非损伤性样本全基因组富集的成本高且富集效率较低,目前的研究主要集中于线粒体基因组、外显子组以及特定染色体的杂交捕获.基于 Maricic等人^[24]的杂交捕获方法, Liedigk等人^[25]从食蟹猴(*Macaca fascicularis*)的博物馆样本和粪便样本中成功富集到了完整的线粒体基因组.为了进一步提高 Maricic等人^[24]所述杂交捕获方法的捕获效率, van der Valk等人^[26]结合了 Rohland等人^[54]将所有DNA文库进行合并再捕获的混池捕获方法,对黑猩猩和大猩猩的粪便样本进行了线粒体基因组的一轮捕获和二轮捕获.一轮捕获后,线粒体DNA序列从捕获前的0.13%提

高到了29.15%,二轮捕获后提高到了40.5%,平均提高了310倍以上,获得了80%以上样本的完整线粒体基因组.然而,虽然线粒体基因组可以提供一定关于进化的信息,但对物种个体识别、性别鉴定、种群结构、群体历史等解析能力不足.因此,一些研究使用杂交捕获方法从粪便中富集了宿主的外显子组或特定的染色体. Hernandez-Rodriguez等人^[28]使用Roche SeqCap EZ Developer Kit对黑猩猩基因组上的外显子组(57.5 Mb)进行了杂交捕获,探究了一轮杂交捕获和二轮杂交捕获以及多重文库对于外显子捕获的影响.为了评估和比较捕获效率,研究者以富集因子^[55]、捕获敏感性、捕获特异性^[56]以及文库复杂度^[23,57,58]来评估杂交捕获的效率.相比一轮捕获,两轮捕获后的富集因子、捕获特异性均增加,捕获敏感性无明显变化,而文库复杂度呈下降趋势,表明两轮捕获后DNA数据提高,且捕获到了更多的内源性DNA;而且,在相同的数据量下,两轮捕获能获得比一轮捕获更多的基因型数据,非目标DNA的序列也从一轮捕获的34.15%降低到了两轮捕获的7.83%,表明两轮捕获相比一轮捕获获得了更多的有效数据.然而,由于非损伤性样本的DNA质量差异较大^[59],无法判断在不同质量DNA样本的两轮捕获比一轮捕获具有更高的富集率.为此, White等人^[29]使用SureSelect XT人全外显子RNA探针针对黑猩猩粪便内源性DNA进行了杂交捕获,探究了不同内源性DNA含量样本的捕获轮数对DNA杂交捕获结果的影响,发现内源性DNA含量大于2%的样品,相较于二轮捕获(平均内源性DNA含量为51.4%),一轮捕获(平均内源性DNA含量为55%)富集效率更高.然而,该研究只针对粪便内源性DNA含量>2%且总DNA浓度>6 ng/ μ L的样本进行了杂交捕获实验,但大部分粪便样本的宿主DNA含量通常低于1%.因此, Fontsero等人^[30]使用SureSelect XT人全外显子RNA探针针对内源性DNA含量低于1%的黑猩猩粪便样本进行了DNA杂交捕获富集.首先,研究者分别使用1和2 μ g的起始DNA量进行杂交捕获实验,并根据Hernandez-Rodriguez等人^[28]描述的方法评估了捕获效率,结果发现2 μ g起始DNA捕获后的富集因子、捕获敏感度、文库复杂度均高于1 μ g起始DNA的捕获效率,而捕获特异性无明显变化,表明增加DNA的起始量可提高杂交捕获的富集率.另外,研究者根据内源性DNA含量将样本分成3组,其平均内源性DNA含量分别为0.49%、0.69%、0.81%.对这3组

样本的捕获后结果进行综合比较分析后,发现内源性DNA含量更高的样本可以捕获到更多的基因组数据.这项研究为处理内源性DNA含量极低样本的捕获富集提供了参考.基于以上外显子富集的捕获方法,Fontserè等人^[60]使用预先设计的RNA探针从黑猩猩的粪便中捕获21号染色体的非重复区域,且对内源性DNA含量<5%的样本进行了两轮捕获,最终获得了555个平均覆盖率超过0.5×的样本.此外,Ozga等人^[31]使用IDT xGen Exome Research Panel(version 1.0)分别对野生东部黑猩猩的尿液、粪便、牙本质和牙结石生物样本进行了外显子组捕获.捕获后对其内源性DNA含量进行统计比较后发现,尿液DNA的内源性DNA含量(96.4%)显著高于牙本质(75.9%)、粪便(44.9%)和牙结石(38.3%),表明尿液是一种较好的动物非损伤性样本.

综上所述,杂交捕获测序方法在动物非损伤性样本中多应用于线粒体、外显子、单条染色体等较小区域的捕获.而限制其进一步发展的主要原因为非损伤性样本的高复杂性,非损伤性样本DNA通常由低比例的内源性DNA与大量外源微生物DNA组成.因此,如何提高捕获前内源性DNA的含量是提高动物非损伤性样本杂交捕获效率的关键因素.此外,随着研究者对非损伤性样本外显子捕获的探索,混池捕获、多轮捕获等方式均可提高捕获效率.随着杂交捕获测序技术的进一步发展,动物非损伤性样本的杂交捕获效率有待进一步提高.

1.3 杂交捕获测序技术在动物环境DNA中的应用与进展

对环境DNA(environmental DNA, eDNA)进行杂交捕获并测序可获得当地生态系统的物种多样性及生态系统状态信息.例如,Wilcox等人^[61]首次对水生动物环境DNA进行了杂交捕获,为了测试杂交捕获在环境DNA中富集不同水生物种DNA的可行性,作者使用生物素化PCR产物的探针捕获了在实验室制备的已知物种组成的DNA混合物.研究结果表明,捕获富集后可以精准捕获到混合样本中已知每个物种的DNA,证实了杂交捕获方法在环境DNA检测方面的可行性.此外,为了探究杂交捕获方法在环境DNA中的灵敏度,作者还构建了DNA总量分别为0.01、0.1、1、10和100 ng的已知物种组成的DNA混合物文库.经过比较分析发现,当DNA总量低于0.1 ng时,杂交捕获效率急剧下降,

表明起始DNA总量对于捕获效率提高极其重要.但是,当用于捕获的环境DNA起始含量较低时,作者认为改变实验条件可提高捕获效率^[62,63],如使用RNA探针、缩短探针长度、增加探针密度等^[53,64].对于陆生动物来说,由于物种分布不均且难以采集到包含均匀动物DNA的环境DNA样本,故只能通过陆生哺乳动物聚集的水潭、水底沉积物等收集eDNA^[65].例如,Seeber等人^[66]从不同非洲生态系统中的水坑中收集了水和沉积物样本,并从中提取出eDNA,随后采用靶向38种哺乳动物的线粒体基因组的RNA探针针对eDNA进行杂交捕获和测序,最终共鉴定出了16个物种,其中线粒体基因组覆盖率最大可达到99.8%.此外,研究还使用了3种非探针靶向哺乳动物的DNA去模拟混合DNA样本进行了杂交捕获实验,具体而言,将考拉(*Phascolarctos cinereus*)、麝牛(*Ovibos moschatus*)和北极熊(*Ursus maritimus*)的DNA进行等浓度混合,然后对混合DNA文库进行线粒体基因组捕获,结果发现,即使捕获前非靶向探针物种的线粒体基因组覆盖率仅为30%,捕获后仍能达到91%的线粒体基因组覆盖率,表明尽管探针和非靶向探针物种之间存在很大差异,但也可成功进行物种捕获.因此,杂交捕获富集eDNA是监测共享水源陆生哺乳动物物种的有效方法.为了探究古代的生态环境,Murchie等人^[67]开发了一种新的沉积古DNA提取方法,称为冷自旋提取,并结合杂交捕获方法,使用PalaeoChip Arctic-1.0 bait-set(一组用于捕获北极地区已灭绝和现存大型动物线粒体基因组的RNA探针,也包含北极和亚北极植物的叶绿体条形码)对加拿大育空地区更新世-全新世过渡时期的冻土样本进行了杂交捕获.研究结果发现相比商用土壤提取试剂盒提取到的动植物DNA,杂交捕获后获得的植物和动物DNA提高了7.7~19.3倍,利用捕获后的数据,发现了加拿大育空克朗代克地区全新世晚期猛犸象和马的潜在分布证据和全新世早期松树的存在证据.基于相同的方法,Kjær等人^[68]利用PalaeoChip Arctic-1.0 bait-set探针集成功还原了200万年前格陵兰岛最北端的生态圈.为了将环境DNA样本用于特定物种的群体遗传分析,Jensen等人^[69]开发了一套鲸鲨线粒体和核DNA的杂交捕获探针,在鲸鲨生活的海水样本中进行了杂交捕获.结果发现不仅捕获到了大量的鲸鲨线粒体基因组序列,且线粒体基因组序列的平均覆盖率很高,约235×(90×~352×),通过分析线粒体基因序列的

变异位点,发现至少捕获到了5个不同的个体.然而,对核DNA的捕获只富集到少量的鲸鲨DNA序列,反而捕获到了数百万个鲭金枪鱼的DNA序列,因此,在设计探针时应考虑采样点目标物种与非目标物种DNA序列之间的差异性,避免非目标物种DNA对杂交捕获的干扰.

杂交捕获测序技术在环境DNA中的应用在探索特定地区过去与现在的生态多样性方面起着重要作用,其可以更为精确地捕获到更多的物种DNA.近年来,也有研究尝试利用杂交捕获测序技术富集环境DNA中特定物种的基因组序列,但富集到的核基因组数据极少^[69].而造成其效率低下的原因主要为环境样本的特殊性.环境样本中往往含有大量的杂质和多种来源的DNA,如土壤中的腐殖质、水体中的各种有机和无机物质等,这些杂质会干扰杂交捕获的效率和准确性.此外,环境中的DNA容易受到各种因素的影响而发生降解和损伤,直接影响捕获的效率.另外,同物种之间的基因序列存在差异,且环境样本中可能存在未知物种,这增加了引物和探针设计的难度和复杂性,需要大量的生物信息学分析和实验验证来确保其特异性和有效性.然而,随着采样技术以及探针设计精度的进一步提高,利用环境DNA获取特定物种的核基因组数据将成为可能.

2 杂交捕获测序技术在植物研究中的应用与进展

通过杂交捕获测序技术与高通量测序技术相结合是获得植物遗传多样性的有效方法^[14],其中外显子捕获被广泛应用.Zhou等人^[32]首次测试了杂交捕获富集被子植物外显子组时的富集效率.此研究使用SureSelect eArray软件(Agilent Technologies, Santa Clara, USA)设计了长度为120 bp的生物素化RNA探针,对被子植物毛果杨的外显子组进行了杂交捕获和重测序.结果发现捕获到了20.76 Mb(5%)的杨树基因组序列,其中仅有3%的DNA序列没有覆盖到目标区域,且所有样本在目标区域覆盖率均一性都很高.此外,作者指出,在非模式生物中,使用全外显子组杂交捕获时,为了降低定制探针的成本,建议在进行捕获之前,将多个样本的DNA文库进行混合,这样既可以提高捕获效率,又能减少捕获成本.

尽管DNA测序通量迅速增加且成本降低,但因裸子植物基因组较大,对其进行基因组测序仍然是一个挑战^[70,71].因此,许多研究采用杂交捕获方法对裸子植物的编码和调控区域进行测序,以此对裸子植物开展基因组研究.例如,为了进一步探究杂交捕获在富集植物小基因组区域时的捕获效率以及未知内含子区域对捕获的影响,Neves等人^[33]根据15983个火炬松转录本序列在Agilent Technologies设计了一套探针,分别评估了基因组复杂性对捕获效率的影响,混池捕获的富集效率以及探针的可扩展性.在对基因组复杂性对捕获效率影响的研究中,捕获后比对到单倍体互补DNA(cDNA)、单倍体DNA和二倍体DNA的序列比例分别为71%, 87%和83%,表明捕获效率与基因组的复杂性并无直接关联.此外,混池捕获方法通过添加定制条形码降低了捕获成本并提高了测序通量.同时,火炬松探针的扩展性实验表明,杂交捕获方法在表征密切相关物种基因组时具有灵活性,使用相同的探针可以在湿地松中捕获到火炬松相同区域98.9%的序列.另外,作者发现不将探针设计于外显子之间的区域后,测序深度从3.6×提高到了11.2×,表明探针的设计策略可影响捕获的效率.然而,并不是所有的植物外显子捕获都具有较高的捕获效率.基于Neves等人^[33]的方法,Suren等人^[35]在Roche Nimblegen进行了探针设计,并针对云杉和黑松的外显子组区域进行杂交捕获时,发现捕获后平均只有18.5%和21%的测序Reads分别比对到了云杉和黑松的目标基因组区域,总体捕获效率较差.为此,作者分析了可能影响植物外显子组捕获的几种因素,包括探针与目标物种的同源性、探针密度等,发现探针同源性高低和探针密度均与捕获后的基因组覆盖率存在显著正相关性.因此,作者建议应避免对低同源目标区域的捕获,同时增加探针密度以提高外显子区域捕获效率.另外,作者在研究过程中发现小插入缺失/内含子(<10 bp)可能导致杂交捕获成功率较差,因此建议通过围绕小插入缺失设计探针来改进捕获设计,同时减少PCR的循环次数来增加捕获前的DNA文库复杂性.根据Suren等人^[35]的建议,Lesur等人^[34]对捕获探针设计进行了改进.对于长度小于1.5 kb的基因,从外显子中选择一个150 bp的目标序列作为探针,较长的基因则被人工细分为3个区域,并从基因的两个极端区域选择两个150 bp的目标序列作为探针,而这两个序列一个位于外显子内,另一个位于内含

子和外显子的过渡区域. 随后, 利用SureSelect捕获试剂盒对目标区域进行杂交捕获, 平均捕获到95.47%的目标序列, 且所有样本的平均覆盖深度为 $96.81 \times (48.39 \sim 161.67 \times)$. 此外, 杂交捕获方法还被用于从湖泊沉积物岩芯中提取植物沉积古代DNA. 基于Mariac等人^[72]的方法, Schulte等人^[27]设计了覆盖整个落叶松叶绿体基因组的长片段PCR生成的探针, 对来自西伯利亚北部泰米尔地区6700年前的沉积物核心的落叶松种群的叶绿体基因组进行了捕获测序, 并获得了落叶松的完整叶绿体基因组.

综上所述, 由于植物物种基因组的复杂性, 杂交捕获测序是一种基因组测序的替代方案, 并在多种研究中发挥着重要的作用. 虽然部分研究的植物基因组捕获效率较低, 但捕获后均得到了高测序深度的基因组数据, 表明杂交捕获方法在植物个体或种群水平研究中的可行性^[33,35,73,74]. 此外, 探针设计对于植物基因组捕获的富集效率有重要影响, 根据不同的目标区域设计不同的探针是准确富集目标物种DNA的有效手段.

3 杂交捕获测序技术在微生物研究中的应用与进展

杂交捕获测序技术可以从宏基因组样品中选择性富集靶标基因组区域, 这些基因组区域用于目标群落的结构和功能研究^[13]. 例如, Denonfoux等人^[36]首次描述了杂交捕获对土壤等复杂宏基因组样品的富集效果, 该研究以甲烷菌的甲烷辅酶M还原酶亚单位A (*mcrA*)基因为目标, 利用HiSpOD软件^[75]设计了捕获探针, 对来自法国中央高地的Pavin湖沉积物样本进行了杂交捕获. 经过两轮杂交捕获后, *mcrA*序列占比从直接测序的0.003%提高到了捕获后测序的40%, 捕获富集到了长达2.5 kb的DNA片段, 富集效率在41%~100%. 基于Denonfoux等人^[36]的杂交捕获方法, Biderre-Petit等人^[37]利用靶向*rdhA*基因的生物素化RNA探针, 从法国中央高原的帕文湖的水样中捕获到了新的脱卤球菌(*Dehalococcoidia*)还原脱卤酶同源序列, 表明杂交捕获方法可以检测到更多的微生物群落. Manoharan等人^[38]针对与有机质降解相关的酶类基因定制了用于杂交捕获的NimbleGen SeqCap EZ探针, 在土壤样本中富集了与有机质降解相关的酶类基因. 研

究结果表明, 捕获后的基因组中与有机质降解相关的基因是宏基因组测序的5倍以上, 且捕获后的目标基因组在土壤中的分布与基于传统核糖体测序的结果高度相关. 除了针对特殊的基因进行捕获, 更多的研究采用靶向16S rRNA基因的方法探究微生物群落的多样性. Gasc和Peyret^[39]基于16S rRNA序列数据库使用KASpOD^[76]和PhylArray^[77]软件设计了一组针对细菌16S rRNA基因的生物素化RNA探针, 并将其应用于两组样本, 一组样本为由21个细菌和7个古细菌物种组成的人工模拟细菌群落; 另一组样本为从旧化工厂收集的受六氯环己烷污染的土壤样品. 针对这两组样本对16S rRNA基因进行杂交捕获, 测序结果表明, 相比于捕获前的0.41%, 捕获后人工模拟群落的16S rRNA基因的比例提高到了捕获后的97.30%; 而土壤样本中细菌的16S rRNA基因比例从捕获前的0.09%提高到了捕获后的58.74%. 此外, 在模拟群落中, 杂交捕获富集到了仅占群落0.00006%的微生物, 表明捕获灵敏度极高^[78,79]. 除了土壤、水样等组成较为复杂的宏基因组样本, 杂交捕获方法也被用于从生物体中捕获微生物群落. 基于Gasc和Peyret^[39]的方法, Cariou等人^[79]分别使用杂交捕获测序方法与扩增子测序方法对豌豆蚜虫中提取的DNA进行了富集测序, 以比较两种捕获方法在富集昆虫样本微生物群落时的灵敏度. 首先, 对两种方法进行定性比较, 结果发现两种方法均检测到了豌豆蚜虫中本身的微生物群落^[80]. 随后, 对两种方法检测到的微生物群落进行了定量检测, 结果显示两种方法获得的微生物群落丰度相似, 表明杂交捕获测序方法与扩增子测序方法在富集昆虫微生物群落时具有相似的性能. 然而, 由于昆虫微生物群落相对简单, 无法体现杂交捕获在富集微生物群落时的灵敏度. 因此, 虽然扩增子测序方法相对简单, 但在更大更多种类的微生物分类群中进行微生物群落富集时, 杂交捕获方法仍然体现出了更强大的富集能力^[39]. 从临床标本中检测和鉴定细菌DNA是分子微生物学实践中的基本方法, 为了进一步验证杂交捕获的效率, 也为了在物种水平上精确识别操作分类单元, Barrett等人^[40]利用靶向三个操作分类单元(OTU)序列的生物素化RNA探针对人类粪便中提取的DNA进行了杂交捕获. 杂交捕获后, 16S rRNA基因序列从捕获前小于1%提高到了25.4%~44.2%, 最终获得了59个完整的16S rRNA基因序列, 证实了杂交捕获方法的靶向富集效率.

综上所述, 杂交捕获技术可用于富集特定微生物的DNA片段, 帮助准确鉴定和区分不同种类的微生物, 从而深入探究微生物的多样性。然而, 微生物样本来源广泛, 性质差异大。在进行杂交捕获前的样本预处理过程中, 很难充分去除杂质、抑制物等干扰因素, 若处理不当就会降低杂交捕获的效率和效果。此外, 在很多情况下, 微生物的DNA含量极低, 需要杂交捕获技术具备极高的灵敏度才能有效富集目标DNA片段。但目前部分技术手段在面对这种超低含量的微生物DNA时, 还存在一定的提升空间, 难以保证完全精准地捕获到目标DNA。因此, 开发高效的样本预处理技术以及提高探针特异性或是未来微生物基因组捕获的发展方向。

4 杂交捕获测序技术在医学研究中的应用与进展

4.1 杂交捕获测序技术在疾病诊断中的应用与进展

自2009年首次报道以来, 杂交捕获技术已是临床遗传学中用于诊断人类基础疾病的标准方法, 其中外显子组是应用最广泛的捕获靶标^[81,82]。首先, 外显子捕获是检测罕见孟德尔疾病突变的有效方法。例如, Shearer等人^[41]利用SureSelect捕获方法对非综合征性耳聋患者进行了杂交捕获, 并从中发现了致病突变, 且捕获的特异性、灵敏度与Sanger测序相当(>99%)。基于此, 有更多研究者采用Agilent SureSelect捕获试剂盒或其他商用捕获试剂盒检测到了遗传疾病相关的致病突变^[83-85]。为了进一步探究杂交捕获方法在疾病诊断方面的准确性, Baudhuin等人^[42]利用SureSelect定制了靶向一些遗传疾病相关基因的杂交捕获试剂盒, 对未鉴定过的患者DNA样本进行了杂交捕获, 发现捕获后的DNA序列中>99.7%的碱基的覆盖深度超过100×, 且所有变异均显示基于捕获的高通量测序和Sanger测序之间具有100%的一致性。Fridman等人^[43]的研究也支持了这一观点, 作者使用了两种外显子组捕获试剂盒: Agilent-SureSelectQXT和Illumina-Nextera, 将258个基因的Sanger测序结果与杂交捕获测序的结果进行了比较, 发现Sanger测序和杂交捕获测序的真实变异检测敏感性分别为99%和97%~100%, 且Sanger测序和杂交捕获之间存在高度一致性。外显子捕获还是诊断复杂疾病的常用技术手段。例如, Liu等人^[44]使用Halo-

Plex杂交捕获试剂盒对8名肌萎缩侧索硬化症的患者进行了致病基因捕获, 捕获后96.15%的目标区域实现了10×的覆盖率, 94.84%的目标区域实现了20×的覆盖率, 表明了外显子富集的高效率。基于此, 外显子捕获还被用于如自闭症^[86,87]、精神分裂症^[88]、心肌病^[89]等复杂疾病的诊断。近年来, 随着测序技术的广泛应用, 杂交捕获同样被应用于线粒体疾病诊断^[90,91]、癌症筛查^[92-97]等, 表明了杂交捕获技术在疾病诊断方面具有广泛的应用前景。

4.2 杂交捕获测序技术在病毒学研究中的应用与进展

在临床样本中, 每毫升的病毒基因组拷贝数可达数十亿个, 但与宿主核酸相比, 病毒核酸的含量非常低, 而杂交捕获方法可有效富集病毒基因组序列^[98]。例如, 从一系列临床样本(包括血液、唾液、囊泡液、脑脊液和肿瘤细胞系)中, Depledge等人^[45]使用Agilent eArray软件设计了针对水痘带状疱疹病毒、卡波西氏肉瘤相关疱疹病毒和艾巴氏病毒的RNA文库, 并将此RNA文库上传到E-array由Agilent Biotechnologies进行合成为RNA探针。研究者通过这些探针对13个人类疱疹病毒基因组进行了杂交捕获测序, 最终获得了具有高覆盖率(>94%)的病毒全长基因组, 优于聚合酶链式反应(PCR)的方法(88%~92%)。此外, 使用不同病毒家族的捕获探针库进行杂交捕获的方法可以从单个临床样本中提出多个病毒基因组, 这为基于多个基因组序列的诊断检测和宿主病原体相互作用的研究提供了巨大的潜力。此后, 多个研究也将杂交捕获方法应用于丙型肝炎病毒、水痘带状疱疹病毒、人类免疫缺陷病毒等与人类疾病息息相关的病毒基因组检测中^[45,99-101]。此外, Wylie等人^[46]开发了名为ViroCap的探针集, 旨在富集感染脊椎动物的34个DNA和RNA病毒家族(190个病毒属, 337个物种)。为了测试这些探针的富集效率, 作者比较了14个临床样本和8个发热幼儿样本在捕获前后的测序情况。捕获后共检测到32种病毒, 相比捕获前增加了11种, 32个基因组中有12个基因组的基因组覆盖率为>90%, 其中8种病毒获得了完整的基因组。另外, 虽然ViroCap无法检测新的病毒序列, 但这些探针可以识别到核苷酸序列同源性低至58%的变异病毒序列。同年, Briese等人^[47]描述了一个靶向207个感染脊椎动物病毒类群的病毒基因组捕获探针

(VirCapSeq-VERT), 该探针富集后的病毒序列相比未捕获提高了100~1000倍, 且几乎所有病毒都获得了全长基因组序列(>95%)。此外, VirCapSeq-VERT可以通过与较大基因组片段中的短保守序列杂交来检测新型病毒。Chalkias等人^[48]描述了一种名为ViroFind的类似方法, 这个探针集包括了535种可能导致人畜共患的DNA和RNA病毒。此研究使用ViroFind对5名进行性多灶性白质脑病患者进行了多瘤病毒DNA的杂交捕获, 与捕获前测序相比, 临床样本中存在的病毒序列提高了127倍, 在患者体内发现了高度复杂的多瘤病毒种群, 表明ViroFind在研究病毒基因组结构的优势。此外, 由于杂交捕获方法的优势性, 杂交捕获方法还被用于内源性逆转录病毒, 而考拉逆转录病毒(KoRV)是已知唯一一种正在侵入其宿主物种的生殖系逆转录病毒。Tsangaras等人^[102]使用北方考拉和南方考拉的DNA制备了靶向完整考拉逆转录病毒基因组的捕获探针, 并对博物馆中的考拉DNA样本进行了杂交捕获。最终成功获得了历史样本的考拉逆转录病毒全长高覆盖率序列, 且考拉逆转录病毒序列占总序列的2.5%~41%。

4.3 杂交捕获测序技术在病原体研究中的应用与进展

从疾病媒介或动物宿主中快速准确地检索人类病原体的全基因组序列能够对病原体流行病学和进化动力学进行精细、全面的研究, 但病原体基因组相对于人类的基因组非常小, 且无法快速、高效地将病原体DNA从含有大量人类DNA的复杂样本中分离出来, 而杂交捕获方法凭借其可从复杂样本中快速捕获和检测目标DNA片段的能力被广泛应用于获取病原体基因组信息。例如, Bright等人^[103]使用间日疟原虫基因组DNA作为模板进行体外转录构建了全基因组生物素化RNA捕获探针, 并利用此探针对人类患者血液样本DNA进行了全基因组杂交捕获, 捕获后间日疟原虫的基因组DNA从不到0.5%提高到了平均55%(20%~80%), 且全基因组覆盖率高达93%。随后, Smith^[104]通过疟原虫全基因组生物素化DNA探针对患者血液样本DNA进行疟原虫全基因组杂交捕获, 捕获后平均71.3%的DNA序列为疟原虫DNA。然而, 此研究发现在富含AT的基因组区域富集效果不佳, 导致捕获后数据的均一性较差。因此, 此两项研究证明了杂交捕获方

法在研究复杂样本病原体DNA的优势和可行性。此外, 杂交捕获技术还被应用于从非人类样本中富集病原体基因组。例如, Carpi等人^[105]使用定制探针对稚虫和蝗虫的DNA文库进行了混池杂交捕获富集, 成功地富集并测序到了30个伯氏疏螺旋体基因组, 平均基因组覆盖率为99.5%, 基因组测序深度为132 $\times$ 。此外, 研究还发现, 即使一个池的文库数量达到20个, 仍具有高捕获效率和基因组覆盖率, 表明混池捕获的方法可以在很大程度上降低每个样品的捕获反应成本, 降低了对媒介传播和人畜共患传染病的种群基因组学研究的成本要求。

利用杂交捕获技术, 研究人员还从古代人类或动物遗骸中提取病原体DNA, 以揭示古代疾病的流行病学和遗传学特征, 探究古代社会的健康状况和与病原体的相互作用。例如, Wu等人^[106]筛选了来自中国新疆青铜器时代泉二沟墓地的14具古代个体的病原体DNA, 根据已发表的参考序列设计了长度为100 bp的捕获探针。随后, 研究者利用这些探针对这些个体进行了杂交捕获, 最终肠炎沙门氏菌DNA百分比捕获前提高了41~228倍, 平均全基因组覆盖深度为0.09 $\times$ ~4.2 $\times$, 基因组覆盖率比例为3.69%~86.83%。这项研究表明, 至少在3000年前, 肠道沙门氏菌感染了欧亚大陆东部的人类, 并为病原体沿着原丝绸之路传播提供了第一个古老的DNA证据。为了提高杂交捕获在古代病原体DNA中的捕获效率, Clavel等人^[107]开发了一种新的古老病原体样本的DNA提取方法。研究者首先对古代牙齿样本进行了常规提取^[108], 但提取到的鼠疫耶尔森氏杆菌的DNA小于0.93%。因此, 研究者选择对提取出的DNA进行杂交捕获。虽然捕获后提高了鼠疫耶尔森氏杆菌DNA的含量, 但其中包含的大部分序列由PCR重复序列组成(37.72%~98.03%)。基于此, 研究者们通过调整在37 $^{\circ}$ C下对古牙髓粉样本进行预消化的时间, 以1 h为梯度进行了17个DNA提取实验, 发现在预消化1 h后构建的大多数文库包含更多的鼠疫耶尔森菌序列, 释放了大约99%的古老鼠疫耶尔森氏杆菌的DNA。此外, 为了验证这种提取方式对杂交捕获的影响, 研究者对在此条件下提取的DNA进行了杂交捕获, 发现预消化1小时样本的PCR重复最低, 表明这种提取方法提高了古老鼠疫耶尔森氏杆菌在捕获前起始DNA的总量。

综上所述, 杂交捕获测序技术在医学研究中的应

用包括疾病诊断、病毒基因组的获取以及病原体的检测。由于疾病与人类的生存息息相关, 杂交捕获测序技术在医学研究中的发展迅速, 不仅针对不同疾病开发出了不同的探针集, 甚至基于此技术使病毒DNA序列的检测更为灵敏。然而, 随着医学研究的不断深入, 新的基因变异或病毒序列不断被发现, 探针需持续更新优化, 这个过程是极其复杂且耗时的。此外, 在检测一些微量病原体或者早期肿瘤DNA变化时, 对杂交捕获技术的灵敏度要求极高, 目前部分技术手段在这种超低含量目标DNA检测方面还存在一定的提升空间, 难以做到完全精准地捕获和检测。因此, 持续优化并合成可以检测更多疾病、病毒以及病原体的探针集可能是医学杂交捕获方面的重要难题及发展趋势。

5 结语与展望

基因组测序技术在现代生物学研究中发挥重要的作用, 从Sanger测序到高通量测序, 再到第三代长读长测序技术, 技术的每一次进步都极大地推动了生命科学领域的研究和发展。如何将基因组测序技术应用于更多的复杂样本和研究是推进基因组时代向前发展的必要基础。杂交捕获测序技术不仅可以从内源性DNA含量极低的样本中获得大量内源性DNA序列、目标基因以及常规测序无法检测到的区域, 还可从复杂样本中获得感兴趣的基因组区域或检测到稀有物种的DNA。鉴于这种方法具有高通量、特异性强和具有可扩展性的特点, 目前该技术已广泛应用于动物、植物、微生物以及医学等研究领域, 极大促进了这些领域的发展。

尽管杂交捕获测序技术已有了广泛的应用, 该技术仍然有很多局限性。例如, 捕获实验的成本相对较高, 限制了一些实验室对大规模样本的捕获应用; 复杂样本中的抑制剂、外源DNA干扰极大地影响了捕获的效率和准确性; 对于一些低丰度靶点或基因变异类型的捕获灵敏度还需提升。针对这些问题, 许多研究在现有的研究基础上进行了一些方法学上的改进^[109~113], 使杂交捕获的富集效率得到了不同程度的提高。因此, 杂交捕获测序技术有望在多个方向实现

突破。首先, 诸多与杂交捕获相适应的试剂盒与测序技术相继出现, 如微量建库试剂盒、微量测序技术以及全基因组杂交捕获试剂盒, 这些创新成果将为相关领域的捕获提供高效的工具支撑。其次, 在探针设计方面, 借助生物信息学与机器学习技术的深度融合, 设计特异性更强、覆盖度更广的探针, 同时应建立自动化的探针更新系统, 确保技术始终紧跟科学研究与临床需求的步伐。此外, 样本处理技术应朝着标准化、高效化与高质量化的方向发展, 开发出更温和、更精准的核酸提取与处理方法, 有效减少核酸降解, 并提高样本质量, 以降低捕获时的受干扰程度。同时, 新型封闭试剂的研发将进一步降低非特异性结合, 提高杂交捕获的准确性。另外, 增加起始DNA的量、进行多文库混池杂交、减少建库捕获过程中的PCR循环次数、进行二轮捕获或多轮捕获等均是提高捕获效率的有效实验方案, 后续可进一步设计系统的实验以评估多种实验参数对不同样本的影响。最后, 多种技术结合是当前技术向前发展的必要趋势。例如, 杂交捕获可结合纳米技术, 改进探针的设计并放大靶标信号, 或将开发超灵敏的杂交捕获检测技术。杂交捕获技术还可与微流控芯片技术结合, 将实现小型化、高通量、自动化的检测技术, 进一步降低成本并提高捕获检测的效率。

拓宽杂交捕获测序技术的应用范围可极大促进捕获技术的革新与发展。杂交捕获测序技术因其高灵敏度有望应用于新兴传染病的快速诊断和监测。在应对突发公共卫生事件时, 能够快速开发针对新型病原体的检测试剂盒, 及时掌握疫情动态, 为疫情防控决策提供依据。杂交捕获技术可与单细胞测序技术相结合, 对单个细胞进行基因检测, 有助于深入了解细胞异质性, 揭示疾病发生发展的细胞层面机制, 为精准治疗提供更细致的分子信息。杂交捕获技术还可与基因编辑技术联合应用, 以加速肿瘤等疾病药物的研发进程。

综上所述, 杂交捕获技术在未来的发展若能进一步降低测序成本, 扩大基因组捕获范围, 优化探针合成策略, 解决重复序列带来的低捕获效率问题, 杂交捕获技术将在科学研究中发挥更重要的作用。

参考文献

- 1 Friel J, Bombarely A, Fornell C D, et al. Comparative analysis of genotyping by sequencing and whole-genome sequencing methods in diversity

- studies of *Olea europaea* L. [Plants](#), 2021, 10: 2514
- 2 Ojaveer H, Galil B S, Carlton J T, et al. Historical baselines in marine bioinvasions: implications for policy and management. [PLoS One](#), 2018, 13: e0202383
 - 3 Tang Y, Chen J Y, Ding G H, et al. Analyzing the gonadal transcriptome of the frog *Hoplobatrachus rugulosus* to identify genes involved in sex development. [BMC Genomics](#), 2021, 22: 552
 - 4 Zhong Y, Xu F, Wu J, et al. Application of next generation sequencing in laboratory medicine. [Ann Lab Med](#), 2021, 41: 25–43
 - 5 Chen C K, Liao J, Li M S, et al. Urine biopsy technologies: cancer and beyond. [Theranostics](#), 2020, 10: 7872–7888
 - 6 García J C, Bustos R H. The genetic diagnosis of neurodegenerative diseases and therapeutic perspectives. [Brain Sci](#), 2018, 8: 222
 - 7 Ansorge W J. Next-generation DNA sequencing techniques. [New Biotechnol](#), 2009, 25: 195–203
 - 8 Mardis E R. Next-generation DNA sequencing methods. [Annu Rev Genomics Hum Genet](#), 2008, 9: 387–402
 - 9 Gasc C, Peyretailade E, Peyret P. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. [Nucleic Acids Res](#), 2016, 44: 4504–4518
 - 10 Gaudin M, Desnues C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. [Front Microbiol](#), 2018, 9: 2924
 - 11 Krishnakumar S, Zheng J, Wilhelmy J, et al. A comprehensive assay for targeted multiplex amplification of human DNA sequences. [Proc Natl Acad Sci USA](#), 2008, 105: 9296–9301
 - 12 Turner E H, Ng S B, Nickerson D A, et al. Methods for genomic partitioning. [Annu Rev Genom Hum Genet](#), 2009, 10: 263–284
 - 13 Ribiere C, Beugnot R, Parisot N, et al. Targeted gene capture by hybridization to illuminate ecosystem functioning. [Methods Mol Biol](#), 2016, 1399: 167–182
 - 14 Mamanova L, Coffey A J, Scott C E, et al. Target-enrichment strategies for next-generation sequencing. [Nat Methods](#), 2010, 7: 111–118
 - 15 Albert T J, Molla M N, Muzny D M, et al. Direct selection of human genomic loci by microarray hybridization. [Nat Methods](#), 2007, 4: 903–905
 - 16 Okou D T, Steinberg K M, Middle C, et al. Microarray-based genomic selection for high-throughput resequencing. [Nat Methods](#), 2007, 4: 907–909
 - 17 Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. [Nat Biotechnol](#), 2009, 27: 182–189
 - 18 Vilstrup J T, Seguin-Orlando A, Stiller M, et al. Mitochondrial phylogenomics of modern and ancient equids. [PLoS One](#), 2013, 8: e55950
 - 19 Carpenter M L, Buenrostro J D, Valdiosera C, et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. [Am J Hum Genet](#), 2013, 93: 852–864
 - 20 Enk J M, Devault A M, Kuch M, et al. Ancient whole genome enrichment using baits built from modern DNA. [Mol Biol Evol](#), 2014, 31: 1292–1294
 - 21 Ziesemer K A, Ramos - Madrigal J, Mann A E, et al. The efficacy of whole human genome capture on ancient dental calculus and dentin. [Am J Phys Anthropol](#), 2019, 168: 496–509
 - 22 Perry G H, Marioni J C, Melsted P, et al. Genomic-scale capture and sequencing of endogenous DNA from feces. [Mol Ecol](#), 2010, 19: 5332–5344
 - 23 Snyder-Mackler N, Majoros W H, Yuan M L, et al. Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. [Genetics](#), 2016, 203: 699–714
 - 24 Maricic T, Whitten M, Pääbo S, et al. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. [PLoS One](#), 2010, 5: e14004
 - 25 Liedigk R, Kolleck J, Böker K O, et al. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). [BMC Genomics](#), 2015, 16: 222
 - 26 van der Valk T, Lona Durazo F, Dalén L, et al. Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. [Mol Ecol Resour](#), 2017, 17: e111–e121
 - 27 Schulte L, Bernhardt N, Stoof - Leichenring K, et al. Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. [Mol Ecol Resour](#), 2021, 21: 801–815
 - 28 Hernandez - Rodriguez J, Arandjelovic M, Lester J, et al. The impact of endogenous content, replicates and pooling on genome capture from faecal samples. [Mol Ecol Resour](#), 2018, 18: 319–333

- 29 White L C, Fontseré C, Lizano E, et al. A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. [Mol Ecol Resour](#), 2019, 19: 609–622
- 30 Fontseré C, Alvarez - Estape M, Lester J, et al. Maximizing the acquisition of unique reads in noninvasive capture sequencing experiments. [Mol Ecol Resour](#), 2021, 21: 745–761
- 31 Ozga A T, Webster T H, Gilby I C, et al. Urine as a high-quality source of host genomic DNA from wild populations. [Mol Ecol Resour](#), 2021, 21: 170–182
- 32 Zhou L, Holliday J A. Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. [BMC Genomics](#), 2012, 13: 703
- 33 Neves L G, Davis J M, Barbazuk W B, et al. Whole-exome targeted sequencing of the uncharacterized pine genome. [Plant J](#), 2013, 75: 146–156
- 34 Lesur I, Alexandre H, Boury C, et al. Development of target sequence capture and estimation of genomic relatedness in a mixed oak stand. [Front Plant Sci](#), 2018, 9: 996
- 35 Suren H, Hodgins K A, Yeaman S, et al. Exome capture from the spruce and pine giga-genomes. [Mol Ecol Resour](#), 2016, 16: 1136–1146
- 36 Denonfoux J, Parisot N, Dugat-Bony E, et al. Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. [DNA Res](#), 2013, 20: 185–196
- 37 Biderre-Petit C, Dugat-Bony E, Mege M, et al. Distribution of Dehalococcoidia in the anaerobic deep water of a remote meromictic crater lake and detection of Dehalococcoidia-derived reductive dehalogenase homologous genes. [PLoS One](#), 2016, 11: e0145558
- 38 Manoharan L, Kushwaha S K, Hedlund K, et al. Captured metagenomics: large-scale targeting of genes based on ‘sequence capture’ reveals functional diversity in soils. [DNA Res](#), 2015, 22: 451–460
- 39 Gasc C, Peyret P. Hybridization capture reveals microbial diversity missed using current profiling methods. [Microbiome](#), 2018, 6: 61
- 40 Barrett S R, Hoffman N G, Rosenthal C, et al. Sensitive identification of bacterial DNA in clinical specimens by broad-range 16S rRNA gene enrichment. [J Clin Microbiol](#), 2020, 58: e01605-20
- 41 Shearer A E, DeLuca A P, Hildebrand M S, et al. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. [Proc Natl Acad Sci USA](#), 2010, 107: 21104–21109
- 42 Baudhuin L M, Lagerstedt S A, Klee E W, et al. Confirming variants in next-generation sequencing panel testing by sanger sequencing. [J Mol Diagnostics](#), 2015, 17: 456–461
- 43 Fridman H, Bormans C, Einhorn M, et al. Performance comparison: exome sequencing as a single test replacing Sanger sequencing. [Mol Genet Genomics](#), 2021, 296: 653–663
- 44 Liu Z J, Li H F, Tan G H, et al. Identify mutation in amyotrophic lateral sclerosis cases using HaloPlex target enrichment system. [Neurobiol Aging](#), 2014, 35: 2881.e11–2881.e15
- 45 Depledge D P, Palser A L, Watson S J, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. [PLoS One](#), 2011, 6: e27805
- 46 Wylie T N, Wylie K M, Herter B N, et al. Enhanced virome sequencing using targeted sequence capture. [Genome Res](#), 2015, 25: 1910–1920
- 47 Briese T, Kapoor A, Mishra N, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. [mBio](#), 2015, 6: e01491-15
- 48 Chalkias S, Gorham J M, Mazaika E, et al. ViroFind: a novel target-enrichment deep-sequencing platform reveals a complex JC virus population in the brain of PML patients. [PLoS One](#), 2018, 13: e0186945
- 49 Martiniano R, Garrison E, Jones E R, et al. Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. [Genome Biol](#), 2020, 21: 250
- 50 Fu Q, Meyer M, Gao X, et al. DNA analysis of an early modern human from Tianyuan Cave, China. [Proc Natl Acad Sci USA](#), 2013, 110: 2223–2227
- 51 Meyer M, Kircher M, Gansauge M T, et al. A high-coverage genome sequence from an archaic denisovan individual. [Science](#), 2012, 338: 222–226
- 52 Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. [Proc Natl Acad Sci USA](#), 2013, 110: 15758–15763
- 53 Cruz-Dávalos D I, Llamas B, Gaunitz C, et al. Experimental conditions improving in-solution target enrichment for ancient DNA. [Mol Ecol Resour](#), 2017, 17: 508–522

- 54 Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res*, 2012, 22: 939–946
- 55 Gupta T, Marlow F L, Ferriola D, et al. Microtubule actin crosslinking factor 1 regulates the balbiani body and animal-vegetal polarity of the zebrafish oocyte. *PLoS Genet*, 2010, 6: e1001073
- 56 Jones M R, Good J M. Targeted capture in evolutionary and ecological genomics. *Mol Ecol*, 2016, 25: 185–202
- 57 Chen Y, Negre N, Li Q, et al. Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat Methods*, 2012, 9: 609–614
- 58 Daley T, Smith A D. Predicting the molecular complexity of sequencing libraries. *Nat Methods*, 2013, 10: 325–327
- 59 Taberlet P, Luikart G. Non-invasive genetic sampling and individual identification. *Biol J Linn Soc*, 1999, 68: 41–55
- 60 Fontserè C, Kuhlwil M, Morcillo-Suarez C, et al. Population dynamics and genetic connectivity in recent chimpanzee history. *Cell Genomics*, 2022, 2: 100133
- 61 Wilcox T M, Zarn K E, Piggott M P, et al. Capture enrichment of aquatic environmental DNA: a first proof of concept. *Mol Ecol Resour*, 2018, 18: 1392–1401
- 62 Bragg J G, Potter S, Bi K, et al. Exon capture phylogenomics: efficacy across scales of divergence. *Mol Ecol Resour*, 2016, 16: 1059–1068
- 63 Pajmans J L A, Fickel J, Courtiol A, et al. Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Mol Ecol Resour*, 2016, 16: 42–55
- 64 Ávila-Arcos M C, Cappellini E, Romero-Navarro J A, et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci Rep*, 2011, 1: 74
- 65 Rees H C, Maddison B C, Middleditch D J, et al. REVIEW: the detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *J Appl Ecol*, 2014, 51: 1450–1459
- 66 Seeber P A, McEwen G K, Löber U, et al. Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Mol Ecol Resour*, 2019, 19: 1486–1496
- 67 Murchie T J, Kuch M, Duggan A T, et al. Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quat Res*, 2021, 99: 305–328
- 68 Kjær K H, Pedersen M W, De Sanctis B, et al. A 2-million-year-old ecosystem in greenland uncovered by environmental DNA. *Nature*, 2022, 612: 283–291
- 69 Jensen M R, Sigsgaard E E, Liu S, et al. Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. *Mol Ecol Resour*, 2021, 21: 690–702
- 70 Baker M. *De novo* genome assembly: what every biologist should know. *Nat Methods*, 2012, 9: 333–337
- 71 Leitch I. Nuclear DNA C-values complete familial representation in gymnosperms. *Ann Bot*, 2001, 88: 843–849
- 72 Mariac C, Scarcelli N, Pouzadou J, et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol Ecol Resour*, 2014, 14: 1103–1113
- 73 Mascher M, Richmond T A, Gerhardt D J, et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J*, 2013, 76: 494–505
- 74 Zhou L, Bawa R, Holliday J A. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). *Mol Ecol*, 2014, 23: 2486–2499
- 75 Dugat-Bony E, Missaoui M, Peyretailade E, et al. HiSpOD: probe design for functional DNA microarrays. *Bioinformatics*, 2011, 27: 641–648
- 76 Parisot N, Denonfoux J, Dugat-Bony E, et al. KASpOD—a web service for highly specific and explorative oligonucleotide design. *Bioinformatics*, 2012, 28: 3161–3162
- 77 Militon C, Rimour S, Missaoui M, et al. PhylArray: phylogenetic probe design algorithm for microarray. *Bioinformatics*, 2007, 23: 2550–2557
- 78 Fierer N, Barber-Åhn A, Laughlin D C. Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities. *Front Microbiol*, 2014, 5: 614
- 79 Cariou M, Ribière C, Morlière S, et al. Comparing 16S rDNA amplicon sequencing and hybridization capture for pea aphid microbiota diversity analysis. *BMC Res Notes*, 2018, 11: 461
- 80 Gauthier J P, Outreman Y, Mieuzet L, et al. Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS One*, 2015, 10: e0120664
- 81 Choi M, Scholl U I, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA*, 2009, 106: 19096–19101

- 82 Ng S B, Turner E H, Robertson P D, et al. Targeted capture and massively parallel sequencing of 12 human exomes. [Nature](#), 2009, 461: 272–276
- 83 Nectoux J, de Cid R, Baulande S, et al. Detection of TRIM32 deletions in LGMD patients analyzed by a combined strategy of CGH array and massively parallel sequencing. [Eur J Hum Genet](#), 2015, 23: 929–934
- 84 Rousseau-Nepton I, Okubo M, Grabs R, et al. A founder *AGL* mutation causing glycogen storage disease type IIIa in Inuit identified through whole-exome sequencing: a case series. [CMAJ](#), 2015, 187: E68–E73
- 85 Lee J W. Mutations in PDGFRB and NOTCH3 are the first genetic causes identified for autosomal dominant infantile myofibromatosis. [Clin Genet](#), 2013, 84: 340–341
- 86 Griesi-Oliveira K, Acab A, Gupta A R, et al. Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons. [Mol Psychiatry](#), 2015, 20: 1350–1365
- 87 Poultney C S, Goldberg A P, Drapeau E, et al. Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. [Am J Hum Genet](#), 2013, 93: 607–619
- 88 Guipponi M, Santoni F A, Setola V, et al. Correction: exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes. [PLoS One](#), 2015, 10: e0141630
- 89 Pérez-Serra A, Toro R, Campuzano O, et al. A novel mutation in lamin A/C causing familial dilated cardiomyopathy associated with sudden cardiac death. [J Cardiac Fail](#), 2015, 21: 217–225
- 90 Gai X, Ghezzi D, Johnson M A, et al. Mutations in FBXL4, encoding a mitochondrial protein, cause early-onset mitochondrial encephalomyopathy. [Am J Hum Genet](#), 2013, 93: 482–495
- 91 Calvo S E, Compton A G, Hershman S G, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. [Sci Transl Med](#), 2012, 4: 118ra110
- 92 Sikkema-Raddatz B, Johansson L F, de Boer E N, et al. Targeted next-generation sequencing can replace sanger sequencing in clinical diagnostics. [Hum Mutat](#), 2013, 34: 1035–1042
- 93 Drilon A, Wang L, Arcila M E, et al. Broad, hybrid capture–based next-generation sequencing identifies actionable genomic alterations in lung adenocarcinomas otherwise negative for such alterations by other genomic testing approaches. [Clin Cancer Res](#), 2015, 21: 3631–3639
- 94 Xie J, Lu X, Wu X, et al. Capture-based next-generation sequencing reveals multiple actionable mutations in cancer patients failed in traditional testing. [Molec Gen Gen Med](#), 2016, 4: 262–272
- 95 Rozenblum A B, Ilouze M, Dudnik E, et al. Clinical impact of hybrid capture–based next-generation sequencing on changes in treatment decisions in lung cancer. [J Thoracic Oncol](#), 2017, 12: 258–268
- 96 Xu M D, Liu S L, Feng Y Z, et al. Genomic characteristics of pancreatic squamous cell carcinoma, an investigation by using high throughput sequencing after in-solution hybrid capture. [Oncotarget](#), 2017, 8: 14620–14635
- 97 Clark T A, Chung J H, Kennedy M, et al. Analytical validation of a hybrid capture–based next-generation sequencing clinical assay for genomic profiling of cell-free circulating tumor DNA. [J Mol Diagnostics](#), 2018, 20: 686–702
- 98 Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. [Nature](#), 2011, 469: 539–542
- 99 Metsky H C, Matranga C B, Wohl S, et al. Zika virus evolution and spread in the Americas. [Nature](#), 2017, 546: 411–415
- 100 Munyuza C, Ji H, Lee E R. Probe capture enrichment methods for HIV and HCV genome sequencing and drug resistance genotyping. [Pathogens](#), 2022, 11: 693
- 101 Thomson E, Ip C L C, Badhan A, et al. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. [J Clin Microbiol](#), 2016, 54: 2470–2484
- 102 Tsangaras K, Siracusa M C, Nikolaidis N, et al. Hybridization capture reveals evolution and conservation across the entire koala retrovirus genome. [PLoS One](#), 2014, 9: e95633
- 103 Bright A, Tewhey R, Abeles S, et al. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. [BMC Genomics](#), 2012, 13: 262
- 104 Smith M. An in-solution hybridisation method for the isolation of pathogen DNA from human DNA-rich clinical samples for analysis by NGS. [TOGENJ](#), 2012, 5: 18–29
- 105 Carpi G, Walter K S, Bent S J, et al. Whole genome capture of vector-borne pathogens from mixed DNA samples: a case study of *Borrelia burgdorferi*. [BMC Genomics](#), 2015, 16: 434

- 106 Wu X, Ning C, Key F M, et al. A 3,000-year-old, basal *S. enterica* lineage from Bronze Age Xinjiang suggests spread along the Proto-Silk Road. *PLoS Pathog*, 2021, 17: e1009886
- 107 Clavel P, Louis L, Sarkissian C D, et al. Improving the extraction of ancient *Yersinia pestis* genomes from the dental pulp. *iScience*, 2023, 26: 106787
- 108 Gamba C, Hanghøj K, Gaunitz C, et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol Ecol Resour*, 2016, 16: 459–469
- 109 Liu Y, Miao B, Li W, et al. Bronze Age cheese reveals human-Lactobacillus interactions over evolutionary history. *Cell*, 2024, 187: 5891–5900
- 110 Lu W X, Zhu M, Chen Y, et al. A novel approach to improving hybrid capture sequencing targeting efficiency. *Mol Cell Probes*, 2019, 46: 101424
- 111 Meng Z, Wang S, Yu L, et al. A novel fast hybrid capture sequencing method for high-efficiency common human coronavirus whole-genome acquisition. *mSystems*, 2024, 9: e01222-23
- 112 Sundararaman B, Vershinina A O, Hershauer S, et al. A method to generate capture baits for targeted sequencing. *Nucleic Acids Res*, 2023, 51: e69
- 113 Jiménez - Mena B, Flávio H, Henriques R, et al. Fishing for DNA? Designing baits for population genetics in target enrichment experiments: guidelines, considerations and the new tool supeRBaits *Mol Ecol Resour*, 2022, 22: 2105–2119

Application and research advances of the hybridization capture sequencing technology

CHEN MingLin, LUO Jia, KUANG WeiMin* & YU Li*

State Key Laboratory for Conservation and Utilization of Bio-Resource in Yunnan, School of Life Sciences of Yunnan University, Kunming 650500, China

* Corresponding authors, E-mail: yuli@ynu.edu.cn; kuangwm0714@ynu.edu.cn

With the development of sequencing technology, the hybridization capture sequencing technology can enrich and sequence-specific genes or genomic target regions by designing specific biotinylated baits to hybridize with the target DNA sequences. This method achieves rapid and accurate detection of the variant loci of the target regions with a smaller amount of data and greatly reduces sequencing costs. In this paper, we systematically reviewed the extensive application and research advances of the hybridization capture sequencing technology in animal, plant, microorganism, and medical study areas. This method highlights the great potential in ancient DNA research, host DNA enrichment of non-invasive samples, the detection of diseases, viruses, and pathogens. In addition, the problems and solutions encountered in the application and development of the hybridization capture sequencing are discussed, and the optimization direction of this technology is prospected. This review provides a reference for further application and development of the hybridization capture sequencing technology.

hybridization capture sequencing, biotinylated baits, high-throughput sequencing, DNA enrichment, DNA library

doi: [10.1360/SSV-2024-0303](https://doi.org/10.1360/SSV-2024-0303)