

DOI: 10.13376/j.cbls/2016015

文章编号: 1004-0374(2016)01-0112-06

## 线粒体基因组的高通量测序策略

杨 婧, 黄 原\*

(陕西师范大学生命科学院, 西安 710062)

**摘要:** 综述了高通量测序技术在线粒体全基因组测序中的策略, 利用该技术对线粒体全基因组进行序列测定的方法可以归纳为两种, 一种是先对目标 mtDNA 进行富集, 包括 mtDNA 的提取纯化, 目标区域 PCR 扩增法以及特异性探针杂交富集法(可分为基于微阵列和基于 PCR 探针的杂交富集法), 然后对富集出的线粒体 DNA 进行高通量测序; 另一种是先从待测样本的基因组高通量数据中挖掘出线粒体基因组序列信息, 之后利用诱饵序列或者近缘物种的线粒体全基因组参考序列, 使用软件 MITObim 对其进行组装。此外, 还给出了线粒体高通量测序的优化流程图和介绍了混合样品的线粒体高通量测序策略。

**关键词:** 线粒体基因组; 高通量测序; mtDNA 富集; 组装

**中图分类号:** Q78 **文献标志码:** A

## Strategy of high-throughput sequencing technology in the mitochondrial genome sequencing

YANG Jing HUANG Yuan\*

(College of Life Science, Shaanxi Normal University, Xi'an 710062, China)

**Abstract:** This paper reviews the application of high-throughput sequencing technologies in the mitochondrial genomes. Two methods using high-throughput sequencing technology are included. The first method needs to enrich mtDNA from the sample, including extraction and purification of mtDNA, PCR amplicon-based, specific-probe hybridization (microarray-based and PCR-generation probe), then the enriched mtDNA was sequenced. For the second method, mitochondrial genome sequences were extracted from genome at first, then they were assembled using bait sequence or the mitochondrial genome sequence of related species as a reference, with software MITObim. In addition, the optimization flow chart of mitogenome sequencing using high-throughput sequencing technology was given in this paper. Finally, this paper also introduces the sequencing strategy of mixed sample.

**Key words:** mitochondrial genome; high-throughput sequencing; mtDNA enrichment; assemble

线粒体基因组具有独特的特征, 包括结构简单、编码区域高度保守、母系遗传、进化速率快、重组率低、拷贝数高等<sup>[1]</sup>, 成为分子标记的首选。线粒体 DNA (mitochondrial DNA, mtDNA) 作为一种很实用的分子标记已被广泛地应用于种群遗传学、进化生物学、谱系地理学和系统发生学等方面的研究<sup>[2-3]</sup>。此外, 线粒体基因组与单个基因的短序列相比, 不仅包含单个基因含有的遗传信息, 同时也可提供基因组水平的特征, 如基因重排<sup>[4]</sup>。随着序列数据的大量增长, 线粒体基因组已经成为研究基因组进化模式和机制的良好材料。

传统的线粒体基因组测序技术主要是基于 PCR 扩增产物的 Sanger 测序方法, 其中通过长 PCR 产物结合引物步移法 (primer walking) 是目前小规模线粒体基因组测序的常用方法<sup>[5]</sup>, 此方法的缺点是许多类群缺乏通用引物信息, 使得扩增线粒体基因组部分片段变得困难。基因组中特殊的结构, 如重复序列、高 A+T 含量和二级结构也会导致扩增的失败。

收稿日期: 2015-08-06; 修回日期: 2015-09-14

基金项目: 国家自然科学基金项目 (31172076, 31372192)

\*通信作者: E-mail: yuanh@snnu.edu.cn

高通量测序技术与一代 Sanger 测序技术相比而言, 测序通量获得了大幅度的提升, 已开始被应用于核基因组、转录组和 microRNA 等方面的研究。在线粒体全序列测定方面, 由于动物线粒体基因组较小 (大多数在 15~18 kb), 使得采用高通量测序技术测定动物线粒体基因组的性价比太低而限制了其

应用。近年来, 随着测序费用的下降和多种新策略的应用, 研究者们已经可以成功地应用高通量测序技术在短时间内获得大量动物线粒体基因组序列。本文主要综述了高通量测序技术在线粒体全基因组测序中的策略, 见表 1。

## 1 高通量线粒体基因组测序平台

表1 线粒体基因组高通量测序策略的比较

		方法	优点	缺点
单个样本的线粒体高通量测序策略	mtDNA富集	mtDNA提取	降低后续组装的难度,	需要新鲜的样本
		长PCR扩增	提高精确性	长PCR产物在部分类群中难扩增
		特异性探针杂交富集法		需要制备探针
	总DNA测序	使用诱饵(bait)作为组装起始 比对近缘种, 筛选线粒体基因组数据, 从头组装	实验操作简单, 只需提取总基因组	核基因组中存在线粒体假基因, 组装较困难; 测序覆盖度低。
混合样本的线粒体高通量测序策略	使用条形码或标签	通过条形码或标签在组装时将每个样本分开	大大提高了测序的性价比	标签文库制备繁琐, 花费较高
	不使用条形码或标签	利用多个装配软件将不同来源的序列区分开来		需要较高的生物信息分析能力

目前存在的高通量测序技术主要有 Illumina Solexa 合成测序、Roche 454 焦磷酸测序、ABI SOLiD 连接法测序以及 ThermoFisher 半导体测序 (Ion Torrent) 等<sup>[6]</sup>。其中, Illumina HiSeq、Roche 454 FLX、Illumina MiSeq 和 Ion Torrent PGM 等平台都被应用到了线粒体基因组高通量测序中<sup>[7-9]</sup>。然而, 前两种平台用于线粒体高通量测序都各有其不足。HiSeq 平台的数据输出量过大 (可达 600 Gb), 且需花费长达十多天之久的测序运行时间, 较适用于动植物全基因组测序、转录组测序和表观组学的研究, 并不适用于细胞器基因组的测序。454 FLX 易受到高分子均聚物的影响, 会直接影响线粒体基因组的装配精确性, 然而, 其读长长, 被广泛地应用于微生物群落多样性分析、复杂样品的宏基因组学研究和外显子测序等。最近 Illumina 公司推出的 MiSeq 测序平台和 ThermoFisher 的 Ion Torrent PGM 测序平台较适用于线粒体基因组的测序。MiSeq 运行时间短 (5~65 h), 测序读长可达 2 × 300 bp, 且单次运行可产生 0.3~15 Gb 的数据量 (足以满足线粒体基因组测序要求), 可用于快速高效的线粒体基因组测序<sup>[10-11]</sup>。Ion Torrent PGM 基于半导体测序原理, 从文库构建到数据产出只需 2 d, 也被广泛地应用于线粒体基因组测序中。由于线粒体基因组的 A+T 含量较高, 构建的文库属于碱基不平衡文库, 需注意在测序时做碱基平衡处理。

## 2 线粒体基因组的高通量测序策略

目前线粒体基因组高通量测序策略主要可分为两种: mtDNA 富集测序和总 DNA 测序。mtDNA 富集测序是指首先富集出样本的 mtDNA, 再对其进行高通量测序和拼接的方法; 而总 DNA 测序是指从样本的全基因组测序数据集中挖掘和组装线粒体基因组的方法<sup>[12]</sup>。

### 2.1 mtDNA富集测序

#### 2.1.1 线粒体DNA的提取纯化及测序

MtDNA 的提取方法较多, 比较常用的是氯化铯超速离心法、柱层析法以及碱变性法。然而, 由于 mtDNA 只占细胞总 DNA 的一小部分, 并且 mtDNA 的提取需要新鲜的组织、器官或者个体, 因此, 获得 mtDNA 需要花费大量时间, 而且产量极低。针对这个问题, Quispe-Tintaya 等<sup>[8]</sup>设计了一种从哺乳动物细胞中快速分离 mtDNA 的可靠方法, 可以得到 ~2 000 倍富集的 mtDNA, 之后使用 Ion Torrent PGM 系统对其进行测序。

#### 2.1.2 长PCR扩增及测序

基于长 PCR 的 mtDNA 富集法, 首先, 通过通用引物扩增出待测样品的保守短片段, 根据这些短片段设计出物种专一性的长 PCR 引物, 并将 mtDNA 扩增为相互重叠的几段长片段。然后, 将这些长 PCR 产物通过物理方法修剪成大小合适的

片段用于高通量测序<sup>[13-17]</sup>。此方法获得的测序数据用于后续线粒体装配时较简单且准确率高,然而,由于引物的限制,使得长 PCR 的获得较为困难。

### 2.1.3 线粒体DNA特异性探针杂交富集法

这是近年来发展起来的一种快速富集 mtDNA 用于高通量测序的新技术,其中使用较广泛的是基于微阵列的 mtDNA 杂交富集方法。这项技术首先利用微阵列技术合成大量寡核苷酸探针,使之能与线粒体基因组互补结合,从而富集到 mtDNA 片段,然后对富集得到的 mtDNA 片段进行高通量测序<sup>[18]</sup>。例如, Lippold 等<sup>[19]</sup>使用了微阵列杂交的方法捕获普氏野马 (*Equus przewalski*) 的 mtDNA。然而,由于基于微阵列芯片的杂交富集法需要商业化的探针制备,因此,通过 PCR 产生的探针用于序列捕获的技术(sequence capture using PCR-generated probes, SCPP)应运而生。该技术也可用于同时对多个样本(25~100)进行 mtDNA 富集测序,已应用于人类和雀形目的线粒体基因组高通量测序<sup>[20-21]</sup>。如果可以通过以上方法成功地富集出 mtDNA 用于高通量测序,可以简化后续的组装并提高装配的精确度。

## 2.2 总DNA测序

总 DNA 测序法是从样本基因组的高通量测序数据中筛选和装配线粒体基因组的方法<sup>[22]</sup>。此方法的优势在于样本的总 DNA 易获得,使用酒精保存的标本即可提取。缺点是由于核基因组中存在大量不具功能的线粒体基因组片段以及线粒体假基因,使得线粒体基因组的装配变得复杂。使用该方法需通过样本的总基因组和线粒体基因组的大小确定合适的测序数据量,确保线粒体基因组有足够的覆盖度。

截止到目前,有多个软件可用于基于基因组数据的线粒体基因组装配,主要基于以下两种原理:(1)将诱饵(bait)序列作为组装的起始,通过重叠碱基将线粒体基因组进行组装。Bait 序列是通过 Sanger 测序获得的样品的一段短序列(200~1 000 bp)<sup>[23]</sup>;(2)从 NCBI (National Center for Biotechnology Information) 的 Organelle Genome Resources 数据库中获得近缘物种的线粒体全基因组序列,将样本的基因组数据与之比对,从中筛选出线粒体来源的 reads,之后使用装配软件(Newbler、SOAPdenovo、CLC Genomics Workbench 等)进行从头组装(*de novo assemble*)<sup>[24-25]</sup>。

Mitochondrial baiting and iterative mapping (MITObim, <http://github.com/chrishah/MITObim>)<sup>[26]</sup>是常用的线粒体基因组的组装方法,它只需要与样

本亲缘关系近的物种的 bait 序列作为组装起始即可,在组装过程中利用迭代的思想得到目标样本的线粒体全基因组。MITObim 在运行时间、内存消耗和精确性方面表现出很强的优越性,它可以全自动化地从来自于全基因组的高通量数据中重新获得线粒体基因组,其精确度超过 99.5%,运行时间少于 24 h。此方法还适用于宏基因组以及混合物种样品的研究,也适用于其他细胞器基因组甚至是核 DNA 样品的研究。

由于 mtDNA 提取、长 PCR 扩增、杂交探针富集法等在内的 mtDNA 富集法在实验操作上都存在一定难度,随着高通量测序及生物信息学分析技术的发展,越来越多的研究者们都选择使用总 DNA 直接测序法来获得线粒体基因组序列,本文对此方法进行了优化总结<sup>[27]</sup>,见图 1。

## 2.3 通过其他高通量测序数据获得线粒体基因组序列

在外显子组测序数据中,约有 1%~5% 的 reads 可以比对到线粒体基因组序列,因此,可以获得平均覆盖度约为 100\* 的线粒体基因组序列<sup>[28-29]</sup>。MitoSeek (<http://github.com/riverlee/MitoSeek>) 可以从外显子组测序数据中提取线粒体 DNA 的信息<sup>[30]</sup>。随着转录组数据的大量扩增,可以从线粒体转录组数据中挖掘单顺反子和多顺反子转录本,还可为线粒体转录起始和终止位点确定提供重要信息,解析线粒体转录本转录加工过程和成熟机制<sup>[31-32]</sup>。

## 3 利用混合样品策略进一步提高测序通量

高通量测序技术应用于线粒体基因组测序的优势在于可以在短时间内有效地获得大量线粒体基因组数据,研究者们通常采用在单次测序运行中,甚至是测序仪的单个通道中同时对多个样本的 DNA 进行测序。混合样品测序必须考虑到的因素包括混合样本的数量与测序仪的数据输出量以及样本之间的遗传距离<sup>[33]</sup>,难度在于对不同来源的样本序列进行分离。因此,需在构建 DNA 测序文库时对不同来源的样本添加条形码(barcodes)、标记序列(tags)或者 MIDs (multiplexing identifiers),测序生成的 reads 通过这些标签而分开,之后各个样本进行单独地线粒体基因组组装<sup>[34]</sup>。实际上,使用商业化的条形码可以同时 2~96 个不同样本进行测序,如 Ion Xpress Barcode kit (Life Technologies) 和 Nextera XT Index kit (Illumina) 等<sup>[35]</sup>。King 等<sup>[36]</sup>使用 Nextera XT DNA Sample Preparation Kit (Illumina, San Diego,

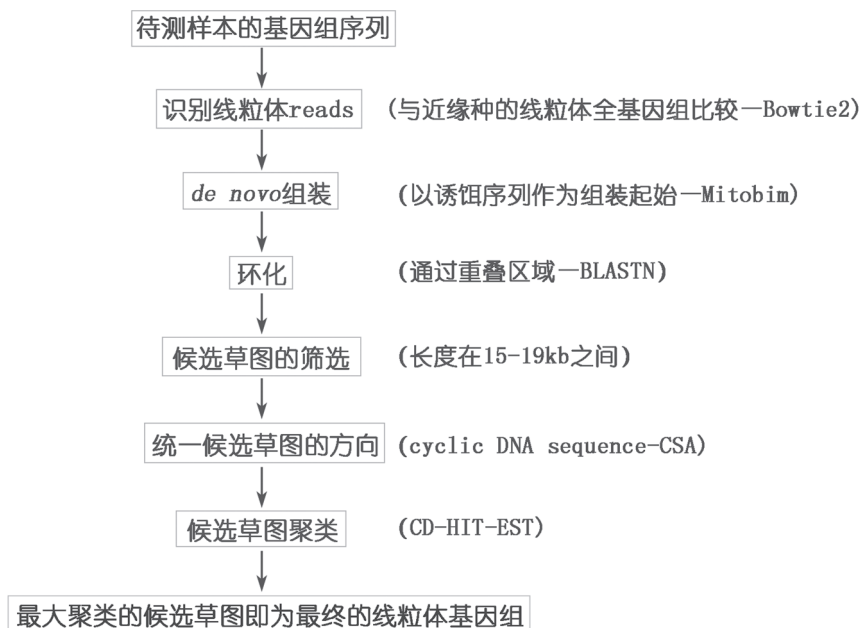


图1 线粒体高通量测序策略(基于基因组)优化流程图

CA) 和 MiSeq™ 平台同时对 96 个人的线粒体基因组 (添加标签) 进行了测定, 平均测序深度为 5 500\*。

平行标签测序技术 (parallel tagged sequencing, PTS) 是一种基于 454 高通量测序平台, 且适用于多样品混合的分子条形码技术, 可在单次运行中测定 300 个平均长度在 17 kb 左右的线粒体全基因组序列 (约 20 倍覆盖率)。PTS 不仅可以用于 shotgun 方法构建的 DNA 文库, 还可用于 PCR 产物构建的测序文库<sup>[37]</sup>。此外, Bybee 等<sup>[38]</sup> 发明了另一种称为 targeted amplicon sequencing (TAS) 的混合样品标签技术, 他们通过两个步骤的 PCR 将 10 bp 的标签连接到 PCR 产物上。此方法生成的含有标签的扩增子可以直接用于高通量测序<sup>[39]</sup>。

在多个样本的高通量测序中引入标签的操作简单, 然而, 需要构建每个样品单独的测序文库, 增加了建库花费。为了降低建库成本, 有些研究者不使用条形码将样品标记起来, 而是在测序之后的组装过程中将不同来源的序列区分开来<sup>[40]</sup>。此方法的缺点是需要较高的生物信息学组装技术, 并且在前期混合样品的亲缘关系选择具有一定局限性。Tang 等<sup>[33]</sup> 在后续组装中利用 *cox1* 基因将 49 种昆虫的线粒体基因组成功地分离组装。

#### 4 线粒体高通量测序数据的保存

下一代测序技术和多样的生物信息学分析手段

使得从几乎所有真核生物的总 DNA 中测序和组装线粒体基因组变得快速、简单和经济。截至到 2015 年 2 月, NCBI 中记录了大于 5 300 个物种的线粒体全基因组信息, 相比于前几年, 线粒体基因组数量呈指数式增长<sup>[41]</sup>。随之出现了一些在线数据库来储存和描述这些数据, 例如 NCBI 的 Organelle Genome Resources (<http://www.ncbi.nlm.nih.gov/genome/organelle/>) 和 MITOMAP (<http://www.mitomap.org/MITOMAP>)。除了原始序列外, 这些数据库还包括其他方面的信息: 转录后编辑、单核苷酸多态性 (SNP)、非常规的遗传密码、内含子插入位置等。这些基因组信息有利于比较基因组的研究, 使得研究者们更容易解析线粒体遗传数据。此外, 高通量测序技术提供了大量未被研究的线粒体遗传数据, 许多期刊要求作者在发表文章前向公共资源提交序列, 例如 GenBank 的 SRA (sequence read archive, <http://www.ncbi.nlm.nih.gov/sra/>), 目前 SRA 已经包含数十亿细胞器来源的 reads, 为研究者们提供了更多的可以挖掘的资源。

#### 5 存在问题及展望

高通量测序技术可以快速地测定多个样本的线粒体基因组全序列, 已广泛地应用于大量后生动物 (包括线虫动物<sup>[15]</sup>、刺胞动物<sup>[42]</sup>、软体动物<sup>[43]</sup>、节肢动物<sup>[4]</sup> 和脊椎动物<sup>[44-45]</sup> 等) 以及被子植物<sup>[16]</sup> 中。此外, 线粒体高通量测序还有助于挖掘单核苷酸多

态性 (SNP)、简单重复序列 (SSR)、线粒体转录组等信息<sup>[46]</sup>。许多疾病的发生与线粒体功能失常密切相关,包括神经退行性病变、糖尿病、肥胖、肿瘤等,应用高通量测序技术可以快速有效地获得线粒体基因组分子诊断的数据<sup>[47]</sup>。目前应用高通量测序技术测定线粒体基因组的主要问题是更加有效地富集 mtDNA,并在测序完成后进行正确的组装。在线粒体基因组富集方面,基于线粒体 DNA 特异性探针杂交富集法是最有发展前途的方法。在线粒体基因组组装方面,由于存在重复区域以及 AT 富集区,且真实的线粒体序列和核拷贝的线粒体 DNA (NUMT) 存在极大的相似性,使得 reads 的装配更加困难<sup>[48]</sup>。但是,随着已知线粒体基因组物种数量的增加和大量相关软件的开发,相信很快可以克服这些困难。

#### [参 考 文 献]

- [1] Wu QL, Li Q, Gu Y, et al. The complete mitochondrial genome of *Taeniogonalos taihorina* (Bischoff) (Hymenoptera: Trigonalyidae) reveals a novel gene rearrangement pattern in the Hymenoptera. *Gene*, 2014, 543: 76-84
- [2] Ballard JW, Rand DM. The population biology of mitochondrial DNA and its phylogenetic implication. *Annu Rev Ecol Evol Syst*, 2005, 36: 621-42
- [3] Simon C, Buckley TR, Frati F, et al. Incorporating molecular evolution into phylogenetic analysis and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu Rev Ecol Evol Syst*, 2006, 37: 545-79
- [4] Yuan ML, Wei DD, Wang BJ, et al. The complete mitochondrial genome of the citrus red mite *Panonychus citri* (Acari: Tetranychidae): high genome rearrangement and extremely truncated tRNAs. *BMC Genomics*, 2010, 11: 597
- [5] 沙淼, 林立亮, 李雪娟, 等. 线粒体基因组测序策略和方法. *应用昆虫学报*, 2013, 50: 293-7
- [6] Emblem A, Karlsen BO, Evertsen J, et al. Mitogenome polymorphism in a single branch sample revealed by SOLiD deep sequencing of the *Lophelia pertusa* coral genome. *Gene*, 2012, 506: 344-9
- [7] Miller AD, Good RT, Coleman RA, et al. Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and *de novo* genome assembly for the critically endangered orange-bellied parrot, *Neophema chrysogaster*. *Mol Biol Rep*, 2013, 40: 35-42
- [8] Quispe-Tintaya W, White RR, Popov VN, et al. Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. *Biotechniques*, 2013, 55: 133-6
- [9] Biswal DK, Ghatani S, Shylla JA, et al. An integrated pipeline for next generation sequencing and annotation of the complete mitochondrial genome of the giant intestinal fluke, *Fasciolopsis buski* (Lankester, 1857) Looss, 1899. *Peer J*, 2013, 1: e207
- [10] Gan HM, Schultz MB, Austin CM. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evol Biol*, 2014, 14: 19
- [11] Williams ST, Foster PG, Littlewood DT. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene*, 2014, 533: 38-47
- [12] Ye F, Samuels DC, Clark T, et al. High-throughput sequencing in mitochondrial DNA research. *Mitochondrion*, 2014, 17: 157-63
- [13] Jex AR, Hall RS, Littlewood DT, et al. An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Res*, 2010, 38: 522-33
- [14] Johansson ML, Sremba AL, Feinberg LR, et al. The mitochondrial genomes of *Euphausia pacifica* and *Thysanoessa raschii* sequenced using 454 next-generation sequencing with a phylogenetic analysis of their position in the Malacostracan family tree. *Mol Biol Rep*, 2012, 39: 9009-21
- [15] Jex AR, Hu M, Littlewood DT, et al. Using 454 technology for the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). *BMC Genomics*, 2008, 9: 11
- [16] Uribe-Convers S, Duke JR, Moore MJ, et al. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. *Appl Plant Sci*, 2014, 2: 1300063
- [17] Fendt L, Zimmermann B, Daniaux M, et al. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics*, 2009, 10: 139
- [18] Maricic T, Whitten M, Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, 2010, 5: e14004
- [19] Lippold S, Matzke NJ, Reissmann M, et al. Whole mitochondrial genome sequencing of domestic horses reveals incorporation extensive wild horse diversity during domestication. *BMC Evol Biol*, 2011, 11: 328
- [20] Maricic T, Whitten M, Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, 2010, 5: e14004
- [21] Peñalba JV, Smith LL, Tonione MA, et al. Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol Ecol Resour*, 2014, 14: 1000-10
- [22] Hao W, Fan S, Hua W, et al. Effective extraction and assembly methods for simultaneously obtaining plastid and mitochondrial genomes. *PLoS One*, 2014, 9: e108291
- [23] Timmermans MJ, Dodsworth S, Culverwell CL, et al.

- Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematic. *Nucleic Acids Res*, 2010, 38: e197
- [24] Paszkiewicz K, Studholme DJ. *De novo* assembly of short sequence reads. *Brief Bioinform*, 2010, 11: 457-72
- [25] Fang Y, Wu H, Zhang T, et al. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS One*, 2012, 7: e37164
- [26] Hahn C, Bachmann L, Chevreur B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res*, 2013, 41: e129
- [27] Huang YF, Midha M, Chen TH, et al. Complete taiwanese macaque (*Macaca cyclopis*) mitochondrial genome: reference-assisted *de novo* assembly with multiple k-mer strategy. *PLoS One*, 2015, 10: e0130673
- [28] Diroma MA, Calabrese C, Simone D, et al. Extraction and annotation of human mitochondrial genomes from 1000 genomes whole exome sequencing data. *BMC Genomics*, 2014, 15: S2
- [29] Picardi E, Pesole G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat Methods*, 2012, 9: 523-4
- [30] Guo Y, Li J, Li CL, et al. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics*, 2013, 29: 1210-1
- [31] Nabholz B, Jarvis ED, Ellegren H. Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. *Mol Phyl Evol*, 2010, 57: 466-70
- [32] Grimes BT, Sisay AK, Carroll HD, et al. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics*, 2014, 15: 31
- [33] Tang M, Tan M, Meng G, et al. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res*, 2014, 42: e166
- [34] Wang T, Pradhan K, Ye K, et al. Estimating allele frequency from next-generation sequencing of pooled mitochondrial DNA samples. *Front Genet*, 2011, 2: 51
- [35] Seo SB, Zeng X, King JL, et al. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion Torrent™ PGM™. *BMC Genomics*, 2015, 16: S4
- [36] King JL, LaRue BL, Novroski NM, et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci Int Genet*, 2014, 12: 128-35
- [37] Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. *Nat Protoc*, 2008, 3: 267-78
- [38] Bybee SM, Bracken-Grissom H, Haynes BD, et al. Targeted amplicon sequencing (TAS): a scalable next-generation approach to multilocus, multitaxa phylogenetics. *Genome Biol Evol*, 2011, 3: 1312-23
- [39] Alexander A, Steel D, Slikas B, et al. Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. *Genome Biol Evol*, 2012, 5: 113-29
- [40] McComish BJ, Hills SF, Biggs PJ, et al. Index-free *de novo* assembly and deconvolution of mixed mitochondrial genomes. *Genome Biol Evol*, 2010, 2: 410-24
- [41] Smith DR. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief Funct Genomics*, 2015 [Epub ahead of print]
- [42] Smith DR, Kayal E, Yanagihara AA, et al. First complete mitochondrial genome sequence from a Box Jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol Evol*, 2012, 4: 52-8
- [43] Groenenberg DS, Pirovano W, Gittenberger E, et al. The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. *BMC Genomics*, 2012, 13: 114
- [44] Bagatharia SB, Joshi MN, Pandya RV, et al. Complete mitogenome of Asiatic lion resolves phylogenetic status within *Panthera*. *BMC Genomics*, 2013, 14: 572
- [45] Lloyd RE, Foster PG, Guille M, et al. Next generation sequencing and comparative analyses of *Xenopus* mitogenomes. *BMC Genomics*, 2012, 13: 496
- [46] McCormack JE, Hird SM, Zellmer AJ, et al. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phyl Evol*, 2013, 66: 526-38
- [47] Dinwiddie DL, Smith LD, Miller NA, et al. Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics*, 2013, 102: 148-56
- [48] Li M, Schroeder R, Ko A, et al. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res*, 2012, 40: e137