

利用宏基因组测序构建土壤动物线粒体基因组数据库

张峰

南京农业大学植保学院

2019.10.14

分类



系统分类和进化

比较形态
整合分类
生物地理
性状演化
系统发育基因组学



生物多样性

环境DNA/宏条形码
宏线粒体基因组
宏基因组
土壤食物网



比较基因组学

害虫/模式物种
重要生物学性状

确定土壤环境中的物种多样性是一项耗时、耗力、耗资的工作！

- 部分类群可能难以通过常规采样方法捕获到；
- 需要花费工作人员大量的时间和精力；
- 挑样过程中可能会产生遗漏；
- 形态学鉴定存在的不足（鉴定结果粗糙、错误）。



精确性？

可重复性？

成本？



主要内容

- 一、项目背景介绍
- 二、技术路线
- 三、已完成工作
- 四、结果与分析
- 五、讨论

一、介绍：项目概况

指南二、中国东部农区土壤动物多样性调查

工作内容：对我国东部农区（东北平原、华北平原、长江中下游平原和珠江三角洲）农田中蚯蚓、螨类、跳虫和线虫等重要类群土壤动物的分布和群落特征进行调查，收集其生存现状基础数据；系统采集东部农区农田及其紧邻的自然环境（林地、草地或湿地）中的真土生、半土生和表土生土壤动物样品并制作标本，数字化编目已有标本及物种信息；建设东部农区土壤动物标本库、物种资源数据库及信息共享平台，整理有害物种清单，编制分布图集。

一、介绍：重点任务量

类群	物种数量 (种)	标本数量 (号)	DNA条形码数量 (条)
蚯蚓	250~400	100,000	30,000
土壤螨类	350~400	150,000	30,000
土壤跳虫	250~400	150,000	30,000
土壤线虫	200~400	100,000	30,000
其他土壤动物	250~400	100,000	30,000
总计	>1200	600,000	150,000

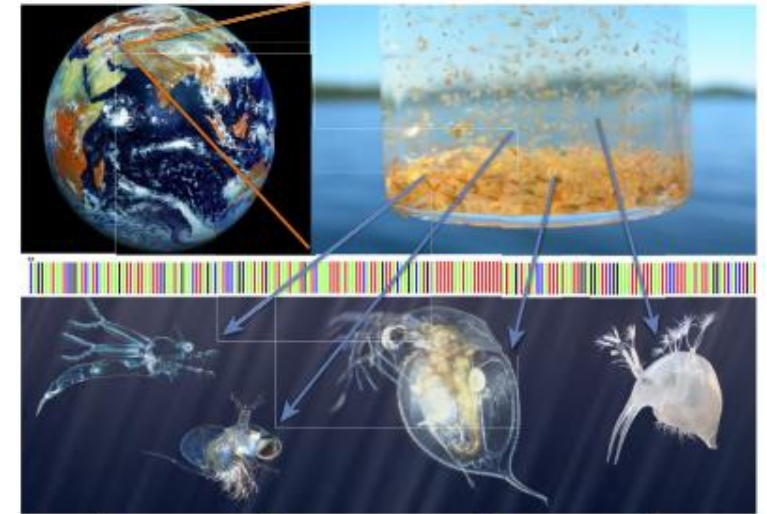
传统方案成本：15万*50元/条=750万元

一、介绍：NGS在生物多样性中的应用



Connecting Earth observation to high-throughput biodiversity data

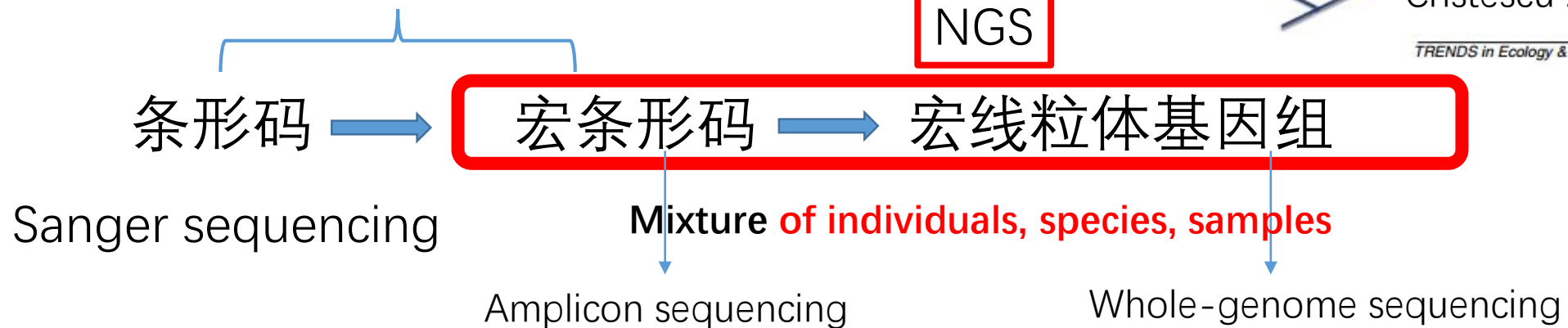
Alex Bush^{1,2,3}, Rahel Sollmann⁴, Andreas Wilting⁵, Kristine Bohmann^{6,7}, Beth Cole⁸, Heiko Balzter^{8,9}, Christopher Martius¹⁰, András Zlinszky¹¹, Sébastien Calvignac-Spencer¹², Christina A. Cobbold¹³, Terence P. Dawson¹⁴, Brent C. Emerson^{15,7}, Simon Ferrier³, M. Thomas P. Gilbert^{6,16}, Martin Herold¹⁷, Laurence Jones¹⁸, Fabian H. Leendertz¹², Louise Matthews¹³, James D. A. Millington¹⁴, John R. Olson¹⁹, Otso Ovaskainen^{20,21}, Dave Raffaelli²², Richard Reeve¹³, Mark-Oliver Rödel²³, Torrey W. Rodgers²⁴, Stewart Snape²⁵, Ingrid Visseren-Hamakers²⁶, Alfried P. Vogler^{27,28}, Piran C. L. White²², Martin J. Wooster^{14,29} and Douglas W. Yu^{1,7*}



Cristescu 2015

TRENDS in Ecology & Evolution

Target specific gene regions by PCR



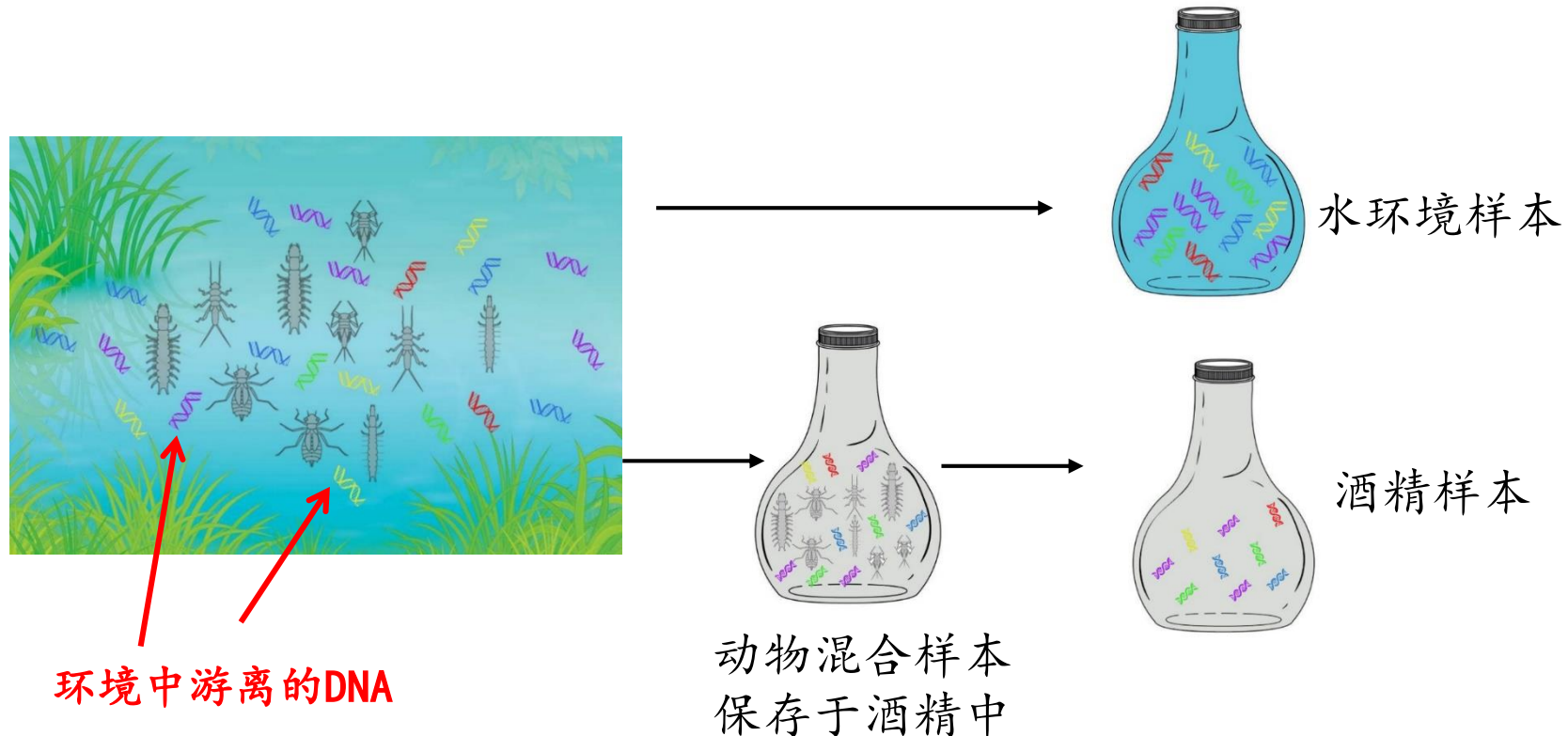
宏条码 vs 宏线粒体基因组

	宏条形码/eDNA	宏线粒体基因组
测序区域	COI, 16/18S... (扩增子)	全基因组
目标组装区域	COI, 16/18S...	线粒体基因组
起始DNA量	低 (< 1ng)	高 (> 200ng)
PCR	+	-
检测成功率	低-高	高
物种丰度评估准确性	低	高
成本(RMB)	20-400 (能混样)	400

环境DNA宏条形码技术 (eDNA Metabarcoding)

利用宏条形码技术对环境样本（如土壤、水、粪便等）中分离的DNA进行扩增和高通量测序，从而达到同时对样本中多个物种（或高级分类单元）鉴定。

快速、有效、低成本、更高的分类水平

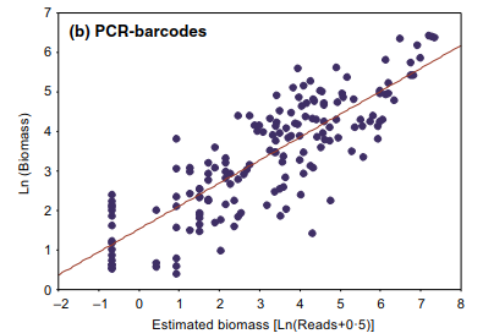
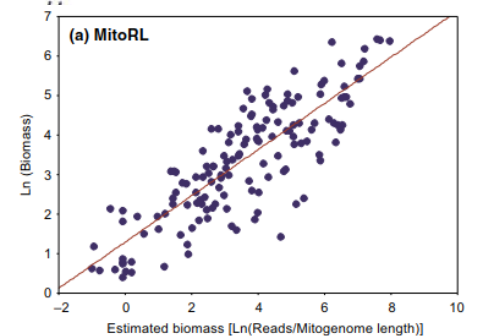
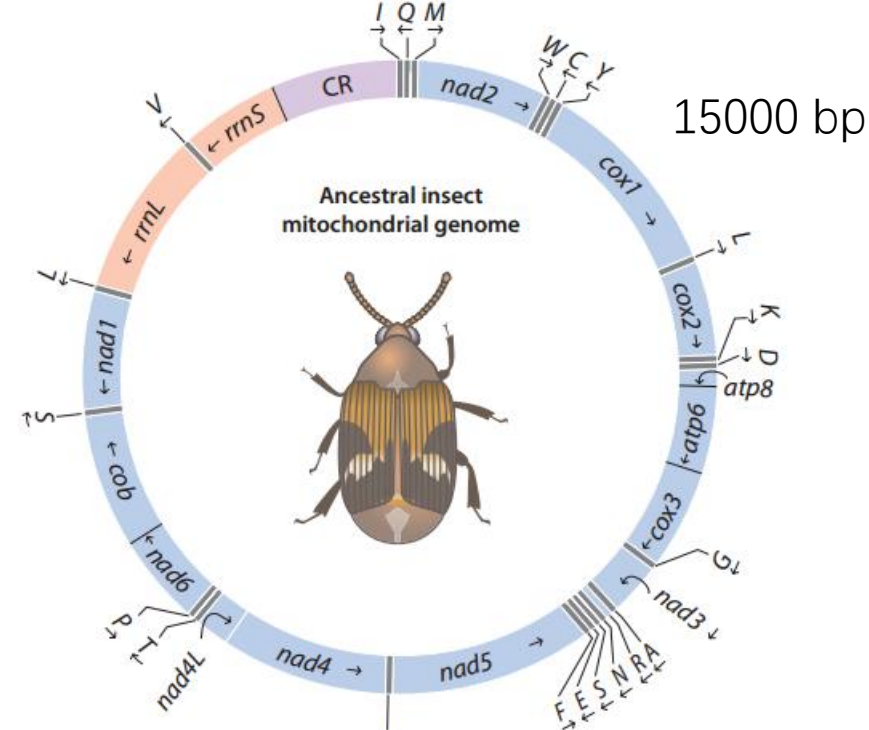


一、介绍：宏线粒体基因组

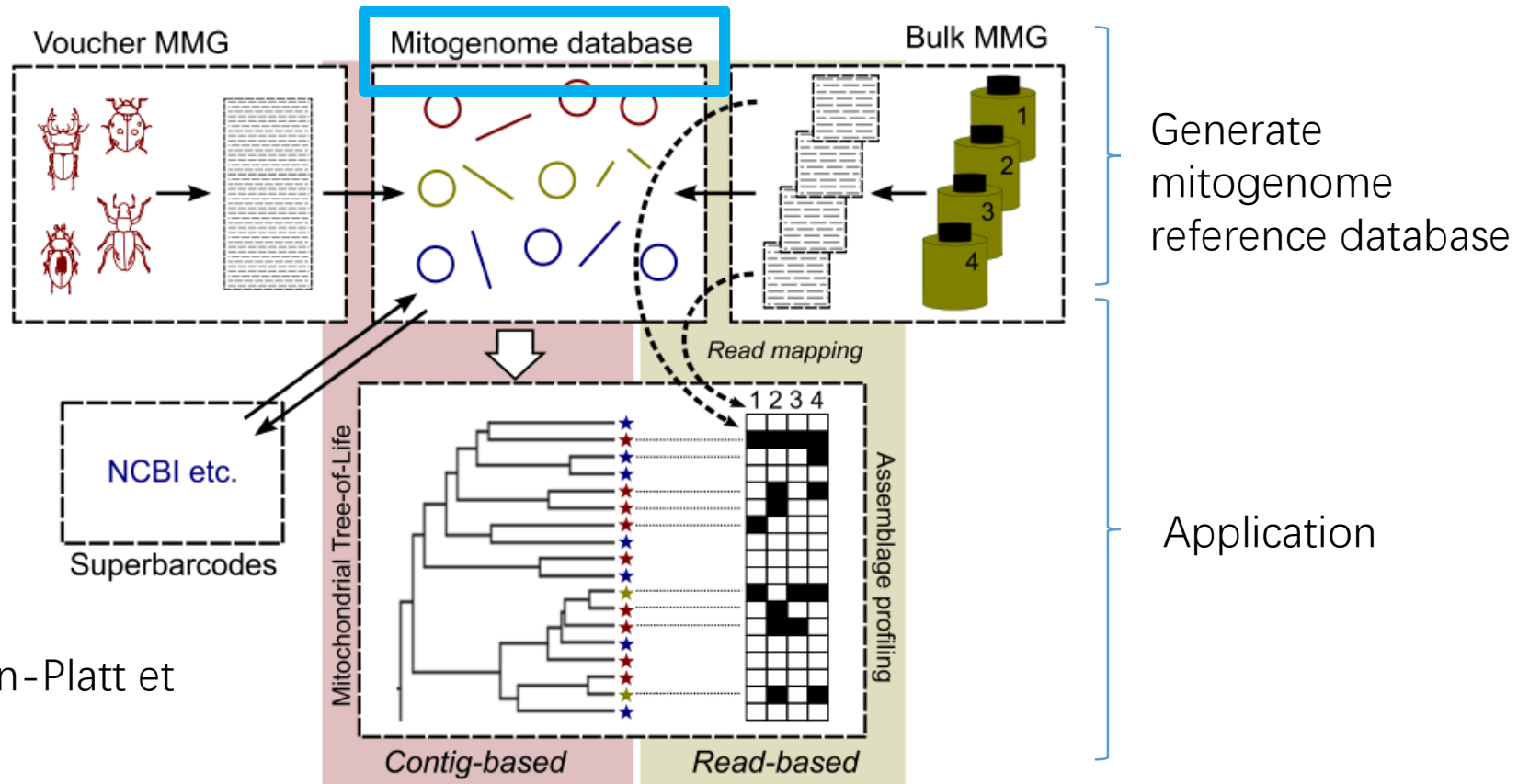
- MMG is a specific form of **metagenome skimming**, targeting the mitochondrial fraction of **bulk specimen sequencing** (Tang et al. 2014, Crampton-Platt et al. 2015).

从低深度全基因组测序数据中组装高拷贝线粒体片段

避免了宏条形码技术中PCR产生的一系列弊端；
检测灵敏度更高；
获取高精度的物种相对丰度/生物量。



一、介绍：MMMG的应用

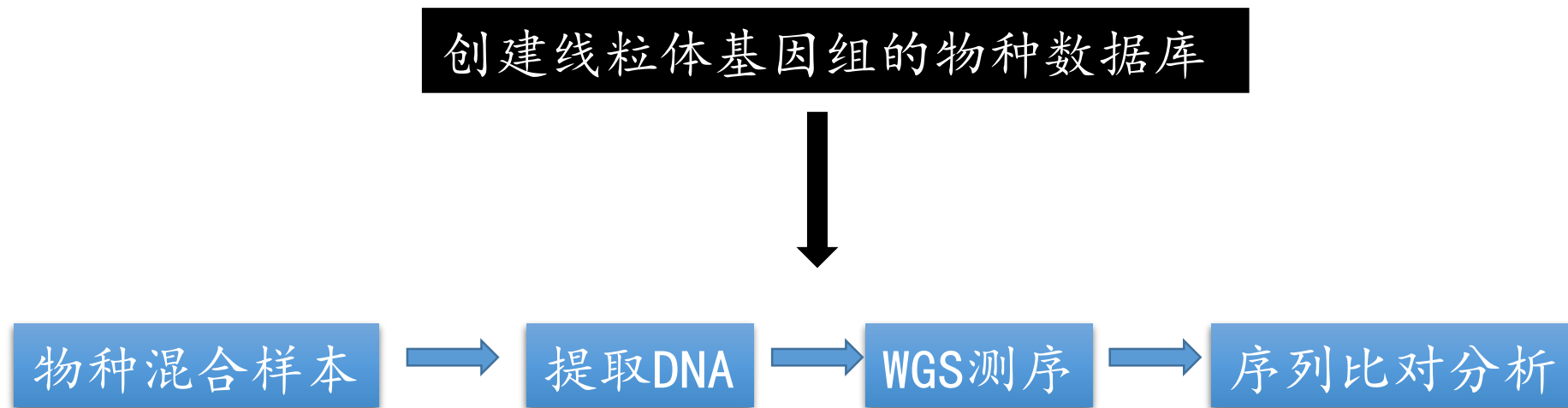


Crampton-Platt et al. 2015

Phylogeny Species composition (abundance/biomass)

Spatial and temporal turnover in complex ecosystems
(phylogenetic community assembly)

宏线粒体基因组的流程



目前更适用于长期固定样地和高精度多样性评估

二、技术路线:

Preprint File available

Advancing mitochondrial metagenomics: a new assembly strategy and validating the power of read-based approach

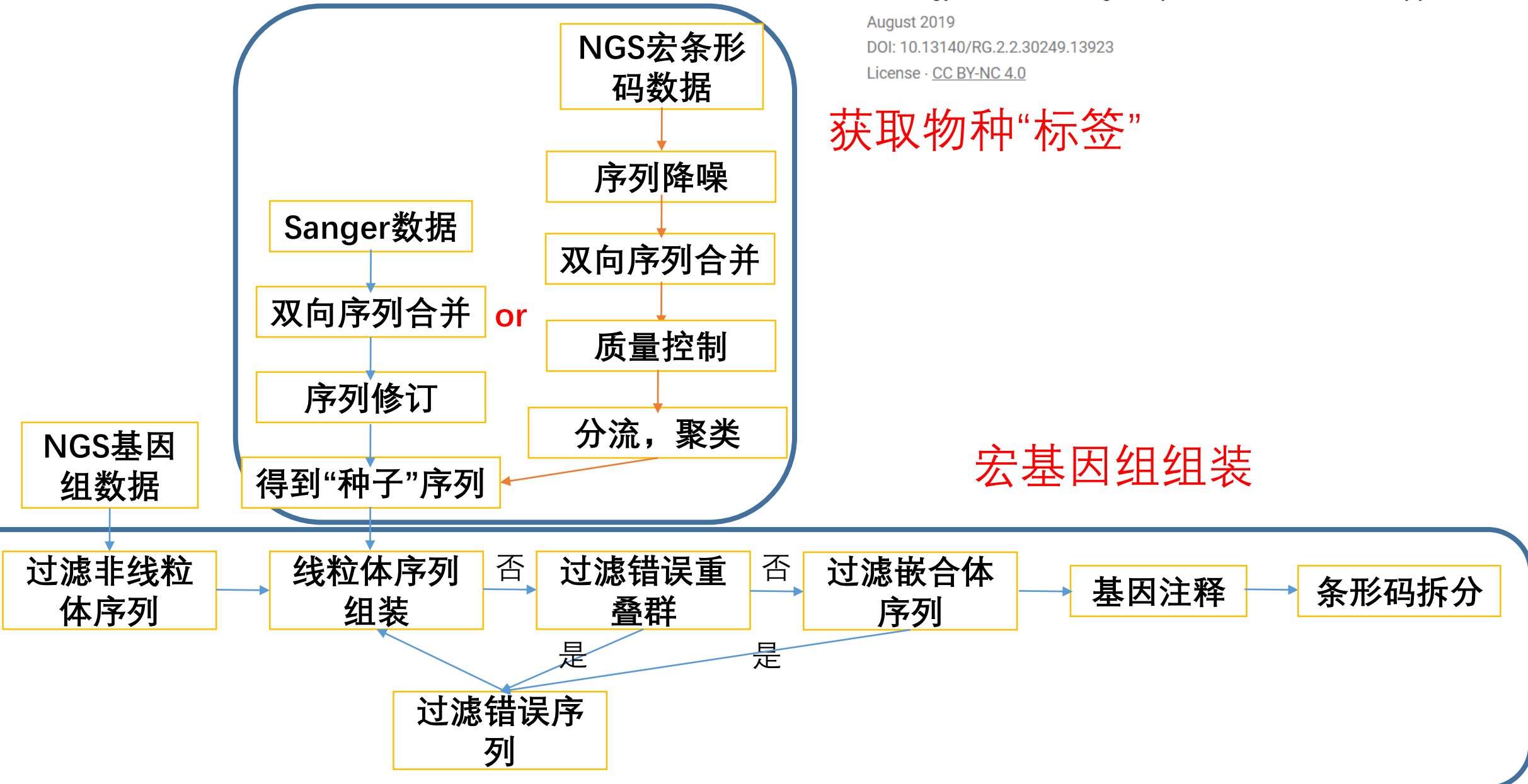
August 2019

DOI: 10.13140/RG.2.2.30249.13923

License · [CC BY-NC 4.0](#)

获取物种“标签”

宏基因组组装



三、已完成工作

测试共计87个物种，完成相关实验、组装及注释。

- ◆ 蜘蛛：20
- ◆ 蚂蚁：8
- ◆ 大蚊：3
- ◆ 金龟：8
- ◆ 蓟马：10
- ◆ 跳虫：38

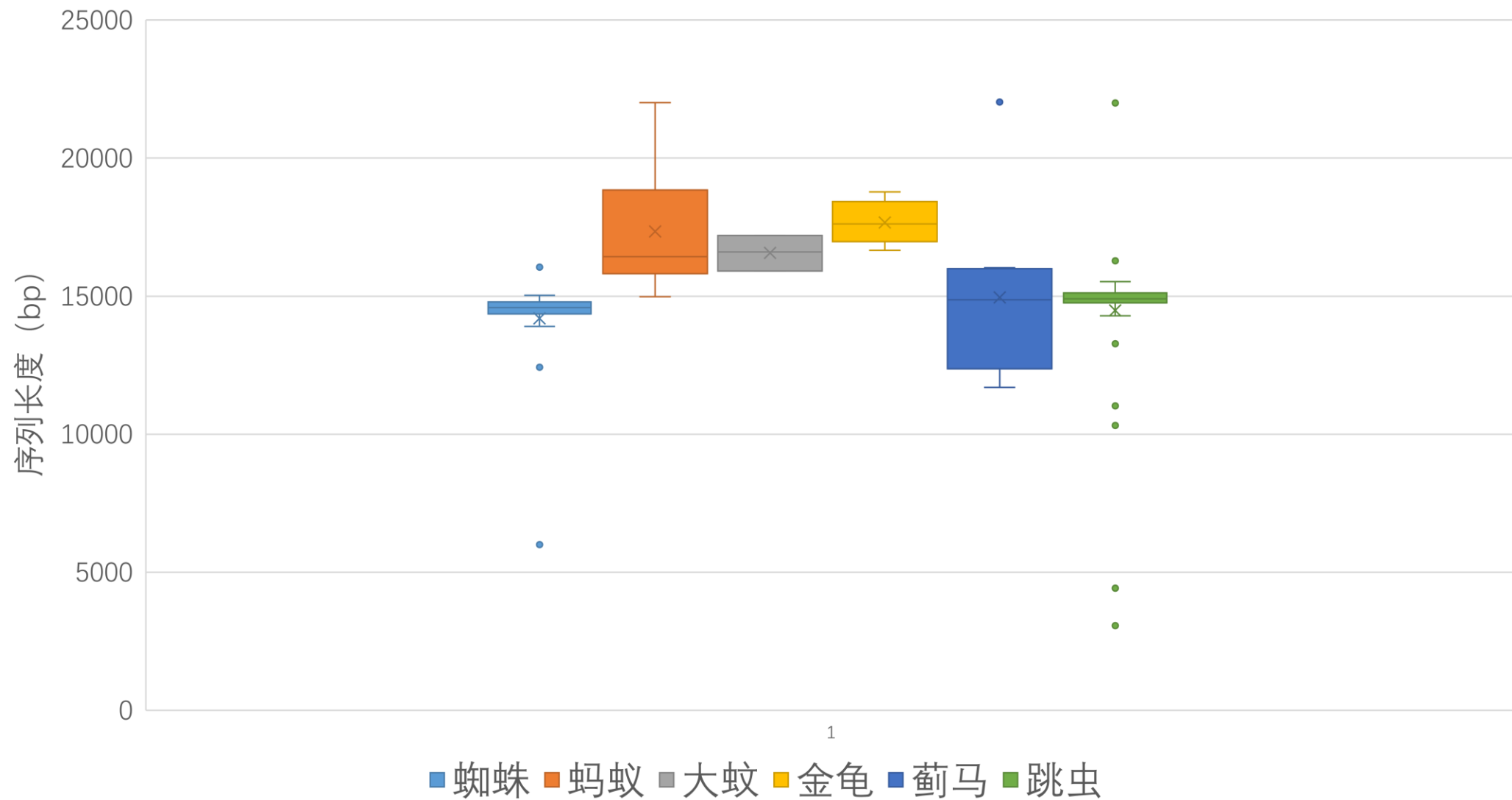


四、结果与分析： 组装结果

- 共成功组装87条线粒体基因组序列，其中长度大于13,500bp及以上的序列共77条，占88.5%。

序号	拉丁名	碱基数目	是否环化	测序深度
1	<i>Poeciloneta variegata</i>	14332	v	3096
2	<i>Xysticus sikkimus</i>	12427	v	115
3	<i>Leucauge celebesiana</i>	13906	v	304
4	<i>Dictyna namulinensis</i>	16051	v	261
5	<i>Hamataliwa</i> sp6	14482	v	961
6	<i>Hahnia corticicola</i>	14728	v	165
7	<i>Thanatus</i> sp.	14802	v	496
8	Uloboridae	14978	v	192
9	<i>Trachelas alticolus</i>	14632	v	1850

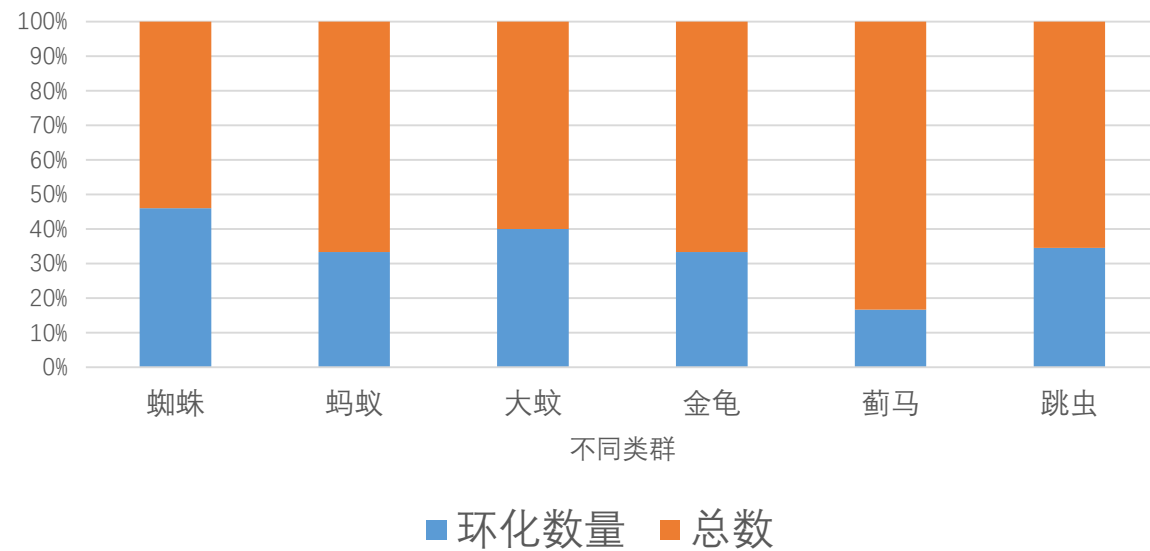
六种不同类群线粒体序列长度分布



不同类群的序列平均长度



不同类群序列环化比例

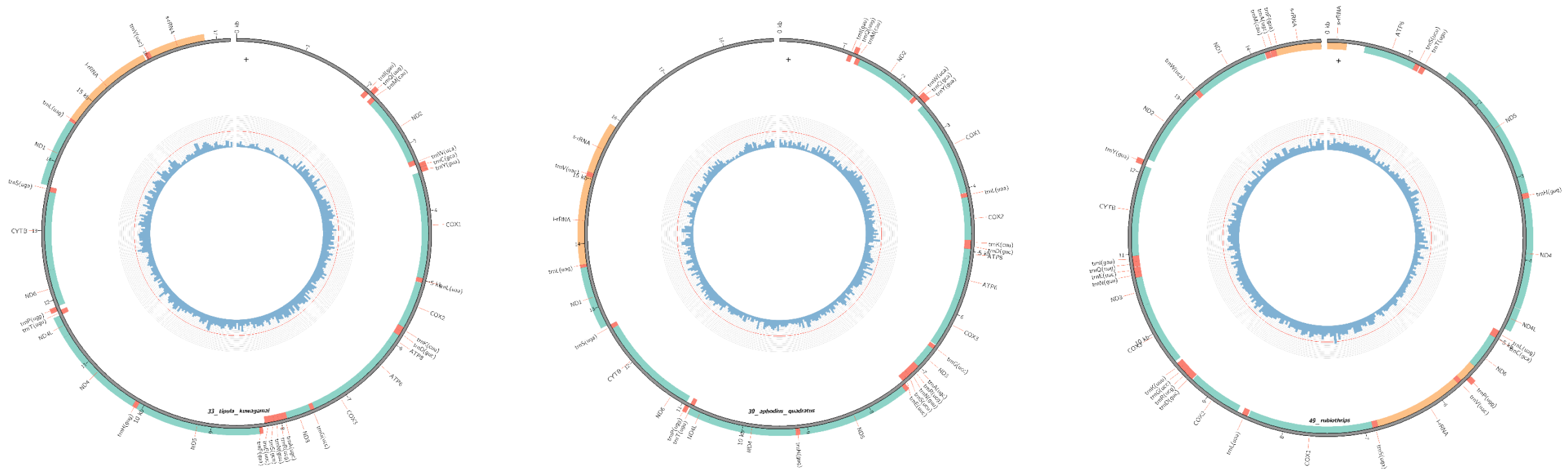


四、结果与分析：注释结果

类群	PCGs(13)	tRNAs(22)	rRNAs(2)	总计
蜘蛛	251(12.55)	172(8.6)	29(1.45)	452(22.6)
蚂蚁	102(12.75)	140(17.5)	13(1.62)	255(31.875)
大蚊	39(13)	66(22)	6(2)	111(37)
金龟	104(13)	176(22)	16(2)	296(37)
蓟马	107(10.7)	190(19)	17(1.7)	314(31.4)
弹尾纲	466(12.26)	791(20.82)	70(1.84)	1327(34.92)
总计	1069(12.38)	1535(18.31)	151(1.80)	2755(32.46)

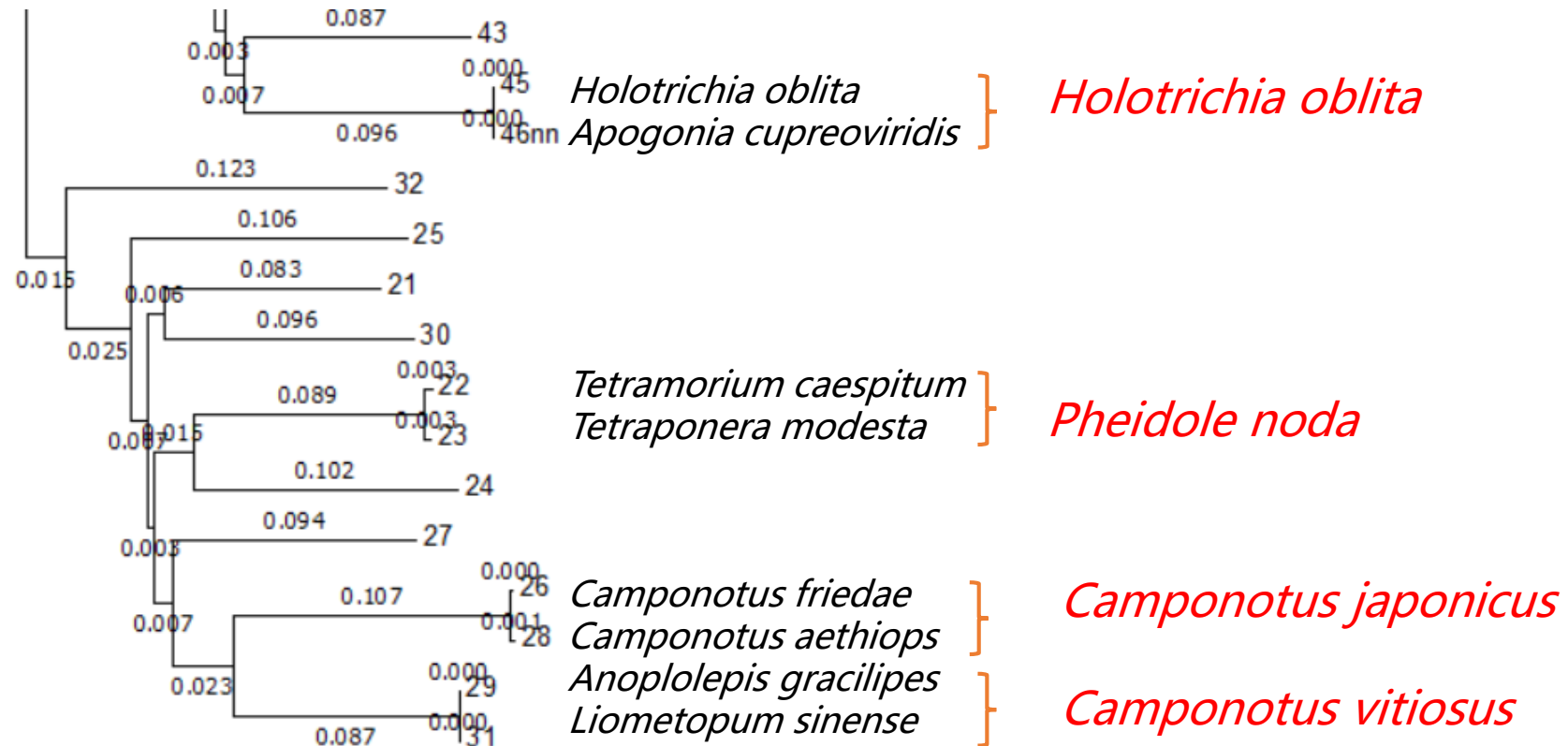
平均每个物种能生成**32**条序列

四、结果与分析：线粒体结构展示



五、讨论：形态学与分子定种冲突

- 约有1/5的样本形态学信息与分子证据不符。



五、讨论：组装准确性

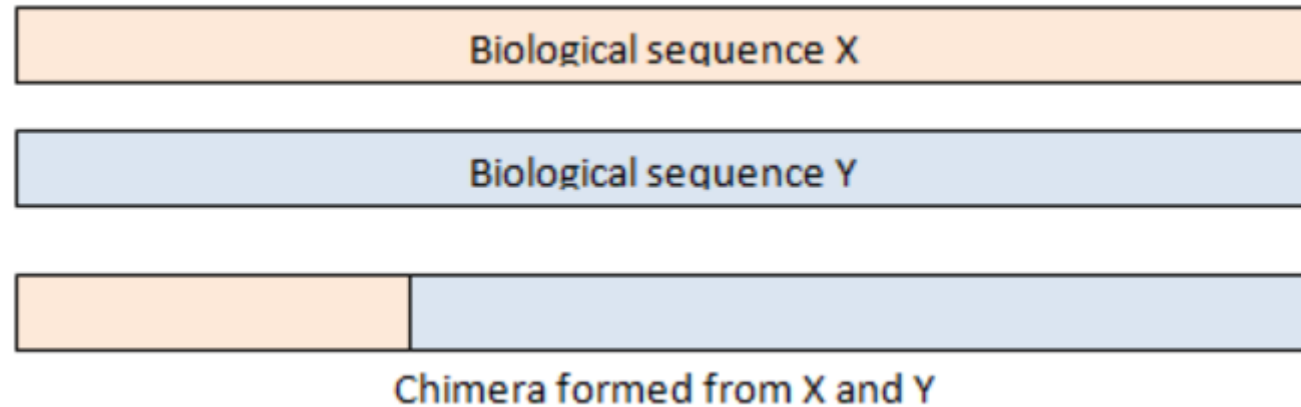
Table 1. Species information, mitogenome reference sequences, and assembly results for datasets extracted from Tang et al. (2014)

Species	Class	Order	Mitogenome reference		Seed	Mitogenome assembly		Accuracy (%)	
			NCBI accession	length (bp)		Length (bp)	Genome coverage (%)	Aligned length to the reference	Percentage (%)
<i>Araneae</i>	Arachnida	Araneae	KM244680	13689	mini-COI	14037	102.5	12225	100.0
<i>Opiliones</i>	Arachnida	Opiliones	KM244669	10922	mini-COI	6241	57.1	6192	100.0
<i>Daphnia magna</i>	Branchiopoda	Cladocera	KM244710	14377	mini-COI	16726	116.3	14377	100.0
<i>Blattodea</i>	Insecta	Blattodea	KM244690	16139	mini-COI	15311	94.9	15311	100.0
<i>Cheilomenes sexmaculata</i>	Insecta	Coleoptera	KM244706	17192	mini-COI	15993	93.0	15993	100.0
<i>Propylea japonica</i>	Insecta	Coleoptera	KM244660	15027	mini-COI	6265	41.7	6265	100.0
<i>Coleoptera</i>	Insecta	Coleoptera	KM244695	15050	mini-COI	15184	100.9	15034	100.0
<i>Tribolium castaneum</i>	Insecta	Coleoptera	KM244661	15877	mini-COI	15883	100.0	15877	100.0
<i>Drosophila erecta</i>	Insecta	Diptera	KM244700	14853	mini-COI	15232	102.6	12973	99.9
<i>Drosophila melanogaster</i>	Insecta	Diptera	KM244693	12956	mini-COI	14972	115.6	12956	100.0

五、讨论：存在的问题

1. 混合物种组装可能会导致**嵌合体序列**的产生。

- 解决方案：
- a. 对嵌合体序列进行分段过滤；
 - b. 尽量避免**近缘种** (<0.10)的混合测序。

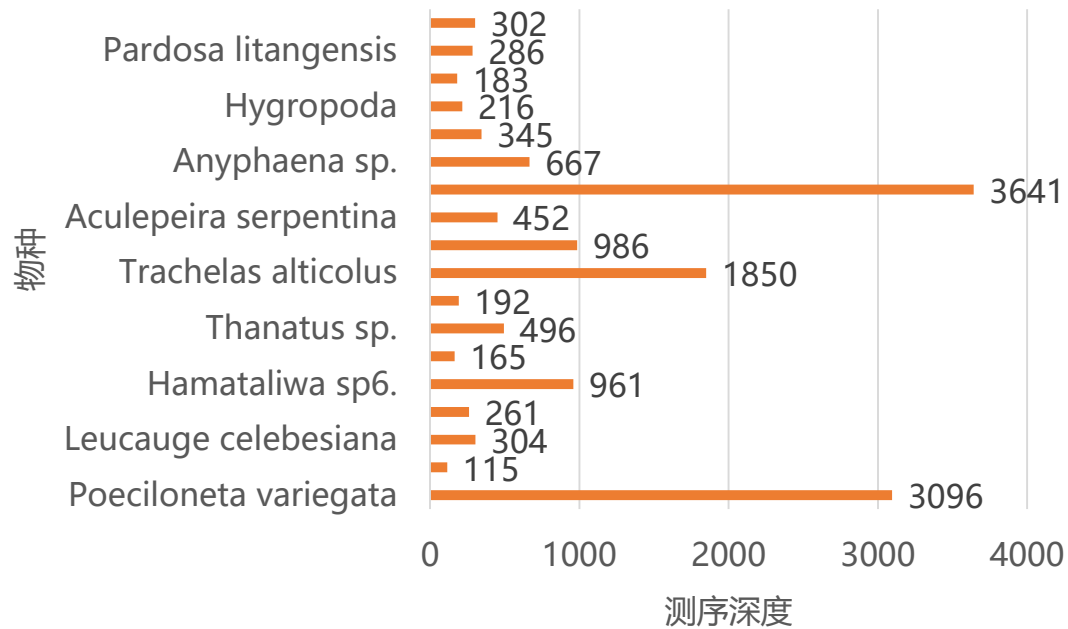


五、讨论：存在的问题

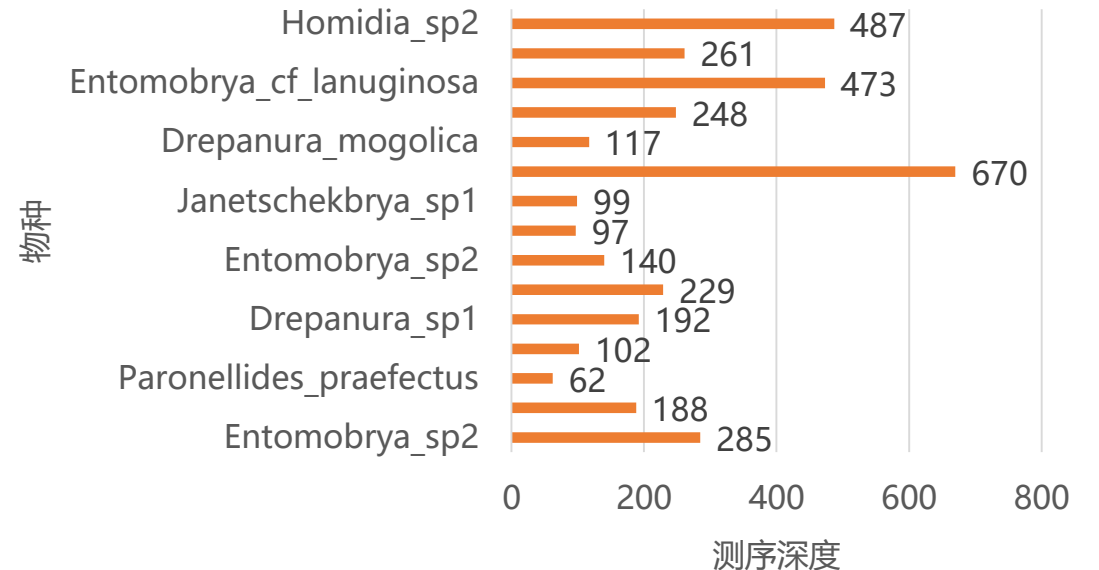
2. 相同物质的量的DNA混合后，不同物种的**测序深度**并不一致 (mtDNA/nuDNA种间差异...)。

建议1.5G/种 (>1G/种)。

蛛形纲样本测序深度



跳虫样本测序深度



五、讨论：存在的问题

- 后期数据库的在线搜索比对
NCBI-blast+?

BLAST® >> blastn suite

Standard Nucleotide BLAST

blastn [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

BLASTN programs search nucleotide databases using a nucleotide query sequence

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#) Query subrange [?](#)

From

To

Or, upload file 未选择任何文件 [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

敬请批评指正！
欢迎合作！