

树鼩参考基因组数据集

（国家非人灵长类实验动物资源库编制，云南省昆明市盘龙区茨坝街道龙欣路 17 号， 650201，
0871-68424851， nhp@mail.kiz.ac.cn， 2025-12-12）

资源名称	中文	树鼩参考基因组数据集
	英文	Tree shrew reference genome dataset
资源标识		CSTR:13153.11.20220430.OD.TSRGD.01.KIZ
数据集内容		a. 描述摘要 本数据集基于高精度二代与三代（PacBio）的树鼩基因组测序数据进行重新组装与整合，显著提高了序列连续性与基因结构解析的准确性。除常规的基因与转录本注释外，还系统性地纳入了非编码元素的注释信息，特别是 环状 RNA（circRNA）和 核线粒体 DNA 片段（NUMTs）。circRNA 作为一种具有组织特异性和进化保守性的非编码 RNA，在基因表达调控、疾病发生以及病毒感染反应中发挥重要作用；而 NUMTs 的识别与标注则有助于厘清核基因组与线粒体基因组之间的遗传关系，为进化分析与基因组结构研究提供可靠参考。这些新增的注释内容不仅拓展了树鼩基因组资源的深度与广度，还为研究者开展 非编码 RNA 功能解析、基因组结构变异研究、以及树鼩在疾病模型中的应用探索 提供了更为全面的数据支持。 b. 关键词 树鼩，基因组，三代测序 c. 数据的时间范围 2022 年 4 月 d. 数据的空间范围 中国科学院昆明动物研究所 e. 学科范围 18031（遗传学） f. 行业范围 M7310（自然科学研究和试验发展） g. 数据格式 .txt h. 数据量 146KB i. 名词解释与量纲 j. 数据精度 数据精度主要通过错误率、比对率阈值与碱基质量分布三方面进行量化与控制。原始测序数据在导入前均经严格质控与过滤，仅保留 Q≥30 的高质量碱基比例不低于 80%–85% 的样本，质量分布曲线在 read 全长范围内平滑，无明显 3’ 端质量衰减现象，保证整体错误率控制在 10⁻³ 量级。在比对层面，对基因组重测序及转录组数据统一设置 比对率纳入阈值为 ≥80%，低于该阈值的样本不被收录；最终进入数据库的样本基因组 reads 平均比对率约在 90%–95%，RNA-seq 样本比对率稳定在 90% 左右，显著提高了变异检测和表达定量结果的可信度。

	k. 数据更新频度 无																																																																																																																																																									
缩略图	 Tree Shrew Database V3.0 HomeSearchExpressionGenomeToolsDownloadsMore NUMTs Search by gene name... <table><tr><th>ID</th><th>Chromosome</th><th>Chr_start</th><th>Chr_end</th><th>MT_start</th><th>MT_end</th><th>Length</th><th>Gene locus</th><th>Get sequence</th></tr><tr><td>TS_2</td><td>chr1</td><td>58895384</td><td>58895638</td><td>15000</td><td>15254</td><td>254</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_3</td><td>chr1</td><td>58896235</td><td>58896349</td><td>15281</td><td>15401</td><td>114</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_4</td><td>chr1</td><td>58896499</td><td>58896963</td><td>15541</td><td>16010</td><td>464</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_5</td><td>chr1</td><td>67684258</td><td>67684384</td><td>852</td><td>977</td><td>126</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_6</td><td>chr1</td><td>67684387</td><td>67684848</td><td>16678</td><td>17140</td><td>461</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_7</td><td>chr1</td><td>67684896</td><td>67684997</td><td>16528</td><td>16628</td><td>101</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_8</td><td>chr1</td><td>67685411</td><td>67686260</td><td>15320</td><td>16171</td><td>849</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_9</td><td>chr1</td><td>82051778</td><td>82052698</td><td>612</td><td>1546</td><td>920</td><td>PPP4R2</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_10</td><td>chr1</td><td>82052963</td><td>82053708</td><td>16676</td><td>17423</td><td>745</td><td>PPP4R2(exon)</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_11</td><td>chr1</td><td>82053745</td><td>82054081</td><td>16399</td><td>16072</td><td>336</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_12</td><td>chr1</td><td>82054083</td><td>82054603</td><td>15541</td><td>16046</td><td>520</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_13</td><td>chr1</td><td>82054772</td><td>82055382</td><td>14788</td><td>15401</td><td>610</td><td>-</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_14</td><td>chr1</td><td>90732383</td><td>90732494</td><td>15078</td><td>15189</td><td>111</td><td>MX2/MX1</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_15</td><td>chr1</td><td>93698778</td><td>93698922</td><td>14647</td><td>14790</td><td>144</td><td>TSPEAR</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_16</td><td>chr1</td><td>106398728</td><td>106398797</td><td>19</td><td>87</td><td>69</td><td>EPB41L5</td><td> Get FASTA msbpg: = 0 bp flank </td></tr><tr><td>TS_17</td><td>chr1</td><td>113958411</td><td>113958682</td><td>8067</td><td>8337</td><td>271</td><td>NCKAP5</td><td> Get FASTA msbpg: = 0 bp flank </td></tr></table>	ID	Chromosome	Chr_start	Chr_end	MT_start	MT_end	Length	Gene locus	Get sequence	TS_2	chr1	58895384	58895638	15000	15254	254	-	Get FASTA msbpg: = 0 bp flank	TS_3	chr1	58896235	58896349	15281	15401	114	-	Get FASTA msbpg: = 0 bp flank	TS_4	chr1	58896499	58896963	15541	16010	464	-	Get FASTA msbpg: = 0 bp flank	TS_5	chr1	67684258	67684384	852	977	126	-	Get FASTA msbpg: = 0 bp flank	TS_6	chr1	67684387	67684848	16678	17140	461	-	Get FASTA msbpg: = 0 bp flank	TS_7	chr1	67684896	67684997	16528	16628	101	-	Get FASTA msbpg: = 0 bp flank	TS_8	chr1	67685411	67686260	15320	16171	849	-	Get FASTA msbpg: = 0 bp flank	TS_9	chr1	82051778	82052698	612	1546	920	PPP4R2	Get FASTA msbpg: = 0 bp flank	TS_10	chr1	82052963	82053708	16676	17423	745	PPP4R2(exon)	Get FASTA msbpg: = 0 bp flank	TS_11	chr1	82053745	82054081	16399	16072	336	-	Get FASTA msbpg: = 0 bp flank	TS_12	chr1	82054083	82054603	15541	16046	520	-	Get FASTA msbpg: = 0 bp flank	TS_13	chr1	82054772	82055382	14788	15401	610	-	Get FASTA msbpg: = 0 bp flank	TS_14	chr1	90732383	90732494	15078	15189	111	MX2/MX1	Get FASTA msbpg: = 0 bp flank	TS_15	chr1	93698778	93698922	14647	14790	144	TSPEAR	Get FASTA msbpg: = 0 bp flank	TS_16	chr1	106398728	106398797	19	87	69	EPB41L5	Get FASTA msbpg: = 0 bp flank	TS_17	chr1	113958411	113958682	8067	8337	271	NCKAP5	Get FASTA msbpg: = 0 bp flank
ID	Chromosome	Chr_start	Chr_end	MT_start	MT_end	Length	Gene locus	Get sequence																																																																																																																																																		
TS_2	chr1	58895384	58895638	15000	15254	254	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_3	chr1	58896235	58896349	15281	15401	114	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_4	chr1	58896499	58896963	15541	16010	464	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_5	chr1	67684258	67684384	852	977	126	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_6	chr1	67684387	67684848	16678	17140	461	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_7	chr1	67684896	67684997	16528	16628	101	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_8	chr1	67685411	67686260	15320	16171	849	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_9	chr1	82051778	82052698	612	1546	920	PPP4R2	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_10	chr1	82052963	82053708	16676	17423	745	PPP4R2(exon)	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_11	chr1	82053745	82054081	16399	16072	336	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_12	chr1	82054083	82054603	15541	16046	520	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_13	chr1	82054772	82055382	14788	15401	610	-	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_14	chr1	90732383	90732494	15078	15189	111	MX2/MX1	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_15	chr1	93698778	93698922	14647	14790	144	TSPEAR	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_16	chr1	106398728	106398797	19	87	69	EPB41L5	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
TS_17	chr1	113958411	113958682	8067	8337	271	NCKAP5	Get FASTA msbpg: = 0 bp flank																																																																																																																																																		
数据质量描述	实验用成年中华树鼯由中国科学院昆明动物研究所实验动物中心饲育。所有动物实验均经中国科学院昆明动物研究所动物伦理委员会批准。动物经戊巴比妥钠麻醉后，采用磷酸盐缓冲盐水（PBS）进行心脏灌注，采集多组织进行提取，包括大脑、肌肉、肝脏、肠道、睾丸等。RNA 提取、文库构建和测序由安诺优达基因科技（中国）完成。																																																																																																																																																									
数据产生方式	RNA-seq 步骤如下：每个样本取约 1 μg 总 RNA，使用 NEBNext® Ultra™ RNA Library Prep Kit for Illumina® 构建 500–1 000 bp 的 RNA-seq 文库。每个文库的质量通过 Agilent Bioanalyzer 2100 系统检测。文库在 Illumina NovaSeq 平台上进行测序，生成 150 bp 的双端读段。数据处理遵循与公开 RNA-seq 数据相同的 QC 流程。 SMRT 测序步骤如下：每个样本的总 RNA 使用 NEBNext® Ultra™ RNA Library Prep Kit for Illumina®分离，并按照制造商的说明书进行处理。通过 1% 琼脂糖凝胶监测 RNA 的降解和污染情况，RNA 纯度使用 NanoPhotometer® 分光光度计检测。RNA 完整性使用 Qubit® RNA Assay Kit 结合 Qubit® 2.0 荧光计和 RNA Nano 6000 Assay Kit 结合 Bioanalyzer 2100 系统进行评估。来自两只树鼯各组织的等量 RNA 被混合为一个混合 RNA 样品，用于 ISO-seq。按照 PacBio 描述的 Isoform Sequencing (ISO-seq™) 协议，使用 Clontech SMARTer PCR cDNA 合成试剂盒和 BluePippin™ 尺寸选择系统分别制备了两个 ISO-seq 文库（<4 kb 和 >4 kb）。随后在 Pacific Bioscience Sequel 系统上，使用两个 SMRT cells 进行了 SMRT 测序。																																																																																																																																																									
数据采集、加工处理方法	a. 环状 RNA 注释 我们使用了来自四个树鼯肝脏的剔除核糖体的 RNA 测序文库来鉴定树鼯的 circRNA。首先，我们使用 STAR 软件将转录组数据与树鼯参考基因组（TS_3.0）进行比对，并获得了关于嵌合剪接位点的信息。然后，使用 DCC 软件根据嵌合剪接位点来识别 circRNA。在数据库中，用户可根据 circRNA 注释来提取 circRNA 序列。 b. 核基因组线粒体假基因（NUMTs）鉴定 首先,使用用 Tantan 软件对线粒体序列和树鼯参考基因组的重复性和低复杂性区域进行屏蔽。然后，我们使用 LAST（https://github.com/michaeldubuyu/LAST/）将树鼯线粒体基因组与染色体水平的基因组进行比对，参数为： "+1 for matches, -1 for mismatches,																																																																																																																																																									

	+7 for gap-open penalty, +1 for gap-extension penalty, 10 ⁻³ for E-value threshold". 为了获得无偏见的比对结果，无论选择何种切割点（圆形线粒体基因组被切割和线性化），我们在比对前在切割点连接两个线性线粒体基因组。
数据使用条件、方法	文件均可用常用办公软件打开
知识产权	a. 标注知识产权说明 使用本数据集时，请在文章中引用以下文献：无 b. 数据标注参考以下规范： 数据来源引用参考以下规范： 中文表达方式：数据来源于国家科技基础条件平台—国家非人灵长类实验动物资源库（https://nhp.kiz.ac.cn）； 英文表达方式：National Resource Center for Non-Human Primates, National Science & Technology Infrastructure of China（https://nhp.kiz.ac.cn）。 致谢方式参考以下规范： 中文致谢方式：感谢国家科技基础条件平台-国家非人灵长类实验动物资源库（https://nhp.kiz.ac.cn）提供数据支撑。 英文致谢方式：Acknowledgement for the data support from National Resource Center for Non-Human Primates, National Science & Technology Infrastructure of China.（https://nhp.kiz.ac.cn）。 c. 数据贡献者信息 姓名：叶茂森 单位：中国科学院昆明动物研究所（姚永刚课题组） 电话：15682010335 邮箱：yemaosen@mail.kiz.ac.cn
其它说明内容	若使用方希望利用该资源的任何材料开展宣传等活动，须事先得到资源管理方的书面授权。