

文章编号:1000-1638(2025)03-0276-14

DOI:10.13484/j.nmgdxzbk.20250307

# 偶蹄目动物染色体重排和 相关着丝粒区域重定位的研究\*

赵恒权<sup>1,2</sup>,刘芳远<sup>3</sup>,那顺布和<sup>3</sup>,陆颖<sup>1,2</sup>,李喜和<sup>4,5</sup>

(1.上海海洋大学水产种质资源发掘与利用教育部重点实验室,上海 201306;

2.上海海洋大学农业农村部淡水水产种质资源重点实验室,上海 201306;

3.内蒙古大学生命科学学院,呼和浩特 010070;

4.内蒙古大学蒙古高原动物遗传资源研究中心,呼和浩特 010021;

5.内蒙古赛科星家畜种业与繁育生物技术研究院,呼和浩特 011517)

**摘要:**偶蹄目(Artiodactyla)动物是人类驯化物种的主要来源(如反刍动物),是研究自然选择和人工选择下染色体进化的重要模型生物。在已发表的72个染色体水平的偶蹄目动物基因组中,通过构建全基因组基因共线性的方法,建立了物种之间的直系同源基因集,并由此构建了偶蹄目物种的完整系统发育树,它与已有结果的相似性为86%。同时,通过建立基因共线性的方法,确定了7个反刍动物代表物种的染色体共线性。结合TRF(Tandem repeats finder)软件预测数据和着丝粒蛋白A抗体CUT&Tag(Cleavage under targets and tagmentation)数据定位了反刍动物代表物种——梅花鹿(*Cervus nippon*)、水牛(*Bubalus bubalis*)和黄牛(*Bos taurus*)的染色体着丝粒区域。结果显示,大部分染色体重排-染色体融合或断裂发生在着丝粒区域,它们的着丝粒结构具有共同的起源。同时发现在一些染色体的重排过程中,会在染色体的融合处进化出新的功能性着丝粒区域,而在染色体重排前的对应着丝粒位置仍然存在祖先着丝粒序列的痕迹。从鉴定到的染色体着丝粒区域的特征性串联重复序列结构来看,这些重复序列具有较高的鸟嘌呤和胞嘧啶(GC)含量,暗示着丝粒区域的功能活化和沉默受到了表观遗传学的调控。本研究初步揭示了物种间染色体重排的分子机制,推测由染色体重排带来的染色体多样性是驱动反刍动物物种多样性形成的重要动力。

**关键词:**偶蹄目;染色体重排;基因组;着丝粒;进化

**中图分类号:**Q951 **文献标志码:**A

偶蹄目(Artiodactyla)作为哺乳纲中多样性最丰富的类群之一,广泛分布于全球陆地生态系统,涵盖约240个现存物种,包括牛科(Bovidae)、鹿科(Cervidae)、猪科(Suidae)和骆驼科(Camelidae)等<sup>[1]</sup>。该类群因具有对称的偶数趾结构(通常为两趾或四趾)而得名,其形态与生理特征在长期进化过程中展现出高度适应性<sup>[2]</sup>,尤其在反刍动物中,复杂的消化系统与植食性策略的结合,显著提升了其对纤维素的利用效率<sup>[3]</sup>。

\* 收稿日期:2025-03-10;修回日期:2025-03-25

基金项目:内蒙古自治区“揭榜挂帅”项目(2022JBG0021)

作者简介:赵恒权(2003—),男,安徽宿州人,2024级硕士研究生。E-mail:zhaohengquan@outlook.com

通信作者:李喜和(1963—),男,内蒙古托克托县人,教授,博士。主要从事动物生殖生物学与生物技术研究。

E-mail:lixh@imu.edu.cn

陆颖(1973—),男,上海人,研究员,博士。主要从事动植物基因组学及分子遗传育种研究工作。

E-mail:yinglu@shou.edu.cn

偶蹄目的演化历史与染色体结构的演变密切相关。染色体作为遗传信息的重要载体,其数目和结构的变化不仅反映了物种进化的过程,还可能影响适应性特征的形成<sup>[4]</sup>。近年来研究表明,在不同偶蹄目动物中,染色体数目发生了明显的变化。例如,黄牛(*Bos taurus*)有 30 对染色体<sup>[5]</sup>,绵羊(*Ovis aries*)则有 27 对染色体<sup>[6]</sup>,而作为偶蹄目代表物种的猪(*Sus scrofa*)只有 19 对染色体<sup>[7]</sup>,这种差异反映了不同物种在演化过程中染色体经历了融合和断裂事件<sup>[8]</sup>。除了染色体数目的变化,染色体结构的重排也是偶蹄目动物染色体演化的重要特征。例如,在黄牛和羊中,染色体的倒位、易位和缺失等事件导致了基因组的显著差异<sup>[9]</sup>。这些结构变异不仅影响了染色体的功能,还可能与物种的适应性进化密切相关<sup>[10]</sup>。有研究发现,某些染色体重排事件会赋予物种在不同生态环境中的生存优势。例如,在寒冷地区生活的牦牛(*B. grunniens*)和藏绵羊可能通过染色体重排适应极端气候条件<sup>[11-12]</sup>。

尽管已有大量研究揭示了偶蹄目动物在染色体演化上的多样性,但染色体演化的分子机制仍有许多关键问题亟待解决。在之前的研究中,着丝粒作为染色体上一个重要的功能区域,在物种有丝分裂和减数分裂时期对染色体的正常分离和结构完整起着至关重要的作用,被认为是影响染色体重排的关键结构。尽管着丝粒的功能在物种进化过程中是高度保守的<sup>[13]</sup>,但对着丝粒的研究表明不同物种着丝粒区域的 DNA 序列存在较大差异,这反映了着丝粒序列发生着快速进化<sup>[14-16]</sup>。在大多数动植物中,着丝粒区域通常是由一些着丝粒卫星 DNA 经过高度串联重复形成的兆碱基(Mb)级别序列<sup>[17]</sup>。目前大量的研究表明,不同物种着丝粒中存在高度多样且复杂的重复单元,这些重复单元之间的序列相似性较低<sup>[18-19]</sup>。着丝粒 DNA 序列的快速进化与其功能上的保守性之间明显存在着矛盾,这种进化悖论表明 DNA 序列不是着丝粒位置和功能的主要决定因素<sup>[20]</sup>。实际上,也有大量的研究支持这个观点,如在稳定遗传的黑腹果蝇双着丝粒染色体中,每次分裂时期通常只有一个区域会保留正常的着丝粒功能<sup>[21]</sup>。在人类<sup>[22]</sup>和酿酒酵母<sup>[23]</sup>的双着丝粒染色体中也存在这种现象,表明染色体上某些区域即使具有了着丝粒序列的特征,也不一定会在该区域形成功能性的着丝粒<sup>[24]</sup>。此外,人类染色体上新的功能性着丝粒可以形成于非着丝粒位点上<sup>[25]</sup>,也表明了着丝粒位置具有灵活的可塑性。虽然不同物种间着丝粒的 DNA 序列不同,但它们均依赖于含有着丝粒特异性组蛋白 H3 变体,即着丝粒蛋白 A(Centromere protein A, CENP-A, 又称 CENH3)的核小体<sup>[26-27]</sup>,因此 CENP-A 蛋白被认为是一种特异性定位着丝粒区域的表观遗传标记<sup>[28-29]</sup>。

目前,虽然有大量研究集中于反刍动物的着丝粒卫星 DNA 序列<sup>[30-31]</sup>,例如对黄牛着丝粒卫星 DNA 的早期研究中,通过密度梯度离心方法将黄牛的卫星 DNA 分为了八种<sup>[32]</sup>;以及近期发表的绵羊<sup>[33]</sup>和家山羊<sup>[34]</sup>(*Capra hircus*) T2T(Telomere-to-telomere)基因组文章中对其着丝粒的结构进行了研究,但是对于反刍动物染色体演化过程中的着丝粒重定位的分子机制研究仍是空白。本研究对偶蹄目下已发表的基因组组装达到染色体水平的物种,构建了较为准确的进化树,结合 CENP-A 蛋白的 CUT&Tag(Cleavage under targets and tagmentation)实验数据和 TRF(Tandem repeats finder)软件的预测数据对反刍动物代表物种的染色体重排,以及对重排过程中的着丝粒序列和位置变化进行了分析,解析了着丝粒功能性结构在物种染色体融合和断裂的演化过程中起到的关键作用。

## 1 基因组数据的收集与分析

### 1.1 偶蹄目基因组数据的下载

目前,已经有 177 个偶蹄目物种(不包括杂交种)的基因组被测序,其中 72 个物种的基因组组装达到了染色体水平,它们在系统分类上覆盖了偶蹄目的 4 个亚目中的 16 个科,4 个亚目分别为胼足亚目(Tylopoda)、猪形亚目(Suina)、反刍亚目(Ruminantia)和鲸河马亚目(Whippomorpha)。其中,33 个物种(46%)属于反刍动物下的牛科和鹿科。我们从 NCBI 的 GenBank 数据库中下载了这些物种的基因组文件、基因组注释文件(General feature format, Gff)以及基因编码的氨基酸序列用于研究分析(表 1)。

表 1 72 个偶蹄目物种的基因组序列登录号

Table 1 Accession numbers of genome sequences for 72 Artiodactyla species

物种	GenBank 登录号	Conitg N50 长度/kb	物种	GenBank 登录号	Conitg N50 长度/kb
大羊驼 <i>L. glama</i>	GCA_028534125.1	93.32	沼泽形水牛 <i>B. kerabau</i>	GCF_029407905.1	91172.39
单峰驼 <i>C. dromedarius</i>	GCF_036321535.1	31216.18	山地紫羚 <i>T. e. isaaci</i>	GCA_040438115.1	79.53
野骆驼 <i>C. ferus</i>	GCF_009834535.1	5365.40	藏羚羊 <i>P. hodgsonii</i>	GCA_040182635.1	84561.98
小抹香鲸 <i>K. breviceps</i>	GCF_026419965.1	42780.42	野山羊 <i>C. aegagrus</i>	GCA_000978405.1	19.35
抹香鲸 <i>P. catodon</i>	GCF_002837175.3	42.88	家山羊 <i>C. hircus</i>	GCA_040806595.1	100793.79
亚马逊河豚 <i>I. geoffrensis</i>	GCA_036417435.1	40892.02	维氏盘羊 <i>O. orientalis</i>	GCA_014523465.1	42163.49
白鲸 <i>D. leucas</i>	GCA_029941455.3	608.47	绵羊 <i>O. aries</i>	GCA_040805955.1	103395.56
一角鲸 <i>M. monoceros</i>	GCA_029941395.3	597.13	帕米尔盘羊 <i>O. a. polii</i>	GCA_028583565.1	74870.27
加湾鼠海豚 <i>P. sinus</i>	GCF_008692025.1	20218.76	羚牛 <i>B. taxicolor</i>	GCF_023091745.1	68053.58
鼠海豚 <i>P. phocoena</i>	GCF_963924675.1	3726.69	鬣羚 <i>C. sumatraensis</i>	GCA_032405125.2	100934.92
长江江豚亚种 <i>N. a. sunameri</i>	GCA_026225855.1	84691.50	臆羚 <i>R. rupicapra</i>	GCA_963981305.1	77662.21
虎鲸 <i>O. orca</i>	GCF_937001465.1	45583.38	黑鹿 <i>M. crinifrons</i>	GCA_020276665.1	25225.00
白喙斑纹海豚 <i>L. albirostris</i>	GCF_949774975.1	3398.03	赤鹿 <i>M. muntjak</i>	GCA_008782695.1	215.53
长肢领航鲸 <i>G. melas</i>	GCF_963455315.1	3290.84	小麂 <i>M. reevesi</i>	GCF_963930625.1	5780.20
伪虎鲸 <i>P. crassidens</i>	GCA_039906525.1	60000.72	印度红鹿 <i>M. vaginalis</i>	GCA_039877825.1	29270.23
里氏海豚 <i>G. griseus</i>	GCA_028646425.1	62.04	黇鹿 <i>D. dama</i>	GCF_033118175.1	68791.58
糙齿海豚 <i>S. bredanensis</i>	GCA_028646385.1	59.98	马鹿 <i>C. canadensis</i>	GCF_019320065.1	914.58
真海豚 <i>D. delphis</i>	GCF_949987515.1	3627.55	梅花鹿 <i>C. nippon</i>	GCA_040085125.1	89550.23
条纹原海豚 <i>S. coeruleoalba</i>	GCA_951394435.1	3590.20	欧洲马鹿 <i>C. elaphus</i>	GCF_910594005.1	68741.66
瓶鼻海豚 <i>T. truncatus</i>	GCF_011762595.1	9729.39	东欧马鹿 <i>C. e. hippelaphus</i>	GCA_002197005.1	7.94
北瓶鼻鲸 <i>H. ampullatus</i>	GCA_949752795.1	3126.55	塔里木马鹿 <i>C. h. yarkandensis</i>	GCA_010411085.1	268.66
索氏中喙鲸 <i>M. bidens</i>	GCA_964148845.1	3262.97	獐 <i>H. inermis</i>	GCA_020226075.1	131.41
柏氏中喙鲸 <i>M. densirostris</i>	GCF_025265405.1	48253.04	骡鹿 <i>O. hemionus</i>	GCA_020976825.1	28570.88
黑真鲸 <i>E. glacialis</i>	GCF_028564815.1	37467.37	白尾鹿 <i>O. virginianus</i>	GCA_023699985.2	21769.37
灰鲸 <i>E. robustus</i>	GCA_028021215.1	39209.00	塞尔扣克北美驯鹿 <i>R. t. caribou</i>	GCA_019903745.2	168.82
蓝鲸 <i>B. musculus</i>	GCF_009873245.2	6315.64	斯瓦尔巴德群岛驯鹿	GCA_949782905.1	22482.69
小须鲸 <i>B. acutorostrata</i>	GCF_949987535.1	3015.72	<i>R. t. platyrhynchus</i>		
墨西哥湾鲸 <i>B. ricei</i>	GCF_028023285.1	36541.94	长颈鹿马赛亚种 <i>G. tippelskirchi</i>	GCA_013496395.1	25.03
河马 <i>H. amphibius</i>	GCA_023065835.1	49478.00	长颈鹿乌干达亚种	GCA_017591445.1	11323.48
河马东非亚种 <i>H. a. kiboko</i>	GCF_030028045.1	50348.26	<i>G. c. rothschildi</i>		
倭河马 <i>H. liberiensis</i>	GCA_023065765.1	20244.51	霍加狓 <i>O. johnstoni</i>	GCA_024291935.2	61700.68
野牛 <i>B. gaurus</i>	GCA_014182915.2	13340.58	苏拉威西鹿豚 <i>B. celebensis</i>	GCA_028533215.1	52.66
野牦牛 <i>B. mutus</i>	GCA_027580195.1	38284.42	非洲疣猪 <i>P. africanus</i>	GCF_016906955.1	10602.50
家牦牛 <i>B. grunniens</i>	GCA_005887515.3	44716.74	欧亚野猪 <i>S. scrofa</i>	GCF_000003025.6	48231.28
黄牛 <i>B. taurus</i>	GCF_002263795.3	26402.95	家猪 <i>S. s. domesticus</i>	GCA_017957985.1	51011.53
爪哇野牛 <i>B. javanicus</i>	GCF_032452875.1	47146.24	欧洲野猪 <i>S. s. scrofa</i>	GCA_006511355.2	595.00
瘤牛 <i>B. indicus</i>	GCF_000247795.1	28.38	马 <i>E. caballus</i> *	GCF_002863925.1	1502.75
水牛 <i>B. bubalis</i>	GCF_019923935.1	9586.61			

注: \* 表示偶蹄目物种进化分化时选择马(*E. caballus*, 奇蹄目)作为外类群(outgroup)。

## 1.2 基于基因组共线性方法构建偶蹄目进化树

由于不同亚目的物种之间染色体保守性较低,因此将偶蹄目物种拆分成6个不同类群,即胼足亚目、猪形亚目、鲸河马亚目、反刍亚目的牛科、反刍亚目的鹿科和反刍亚目的长颈鹿科(Giraffidae),分别构建基因共线性。根据 timetree 数据库<sup>[35]</sup>(<https://timetree.org/>)中的偶蹄目进化树选取了每个类群中靠近进化树基部的物种基因组作为该类群的参考物种基因组,分别是大羊驼(*L. glama*, 胼足亚目)、非洲疣猪(*P. africanus*, 猪形亚目)、黑真鲸(*E. glacialis*, 鲸河马亚目)、黄牛(*B. taurus*, 牛科)、獐(*H. inermis*, 鹿科)、霍加狓(*O. johnstoni*, 长颈鹿科)。运用 Diamond<sup>[36]</sup> v2.1.9.163 软件将每个类群下各物种基因编码的氨基酸序列分别与该类群中的参考物种进行比对,利用 MCSanX<sup>[37]</sup> v0.8 软件建立各物种与参考物种间的基因共线关系,并通过各参考物种之间的进化关系,合并获得整个偶蹄目的基因共线性。为了去除多拷贝基因的干扰,最后只保留了在所有物种中都存在的且“一对一共线关系”的基因。这些“一对一共线关系”的基因可以被视为基因序列和位置都高度保守的直系同源基因。在每个类群中,使用 MAFFT<sup>[38]</sup> v7.525 软件进行多序列比对,按照比对结果将所有氨基酸序列串联后,使用 iqtree<sup>[39]</sup> v1.6.12 软件建树。每个类群都选择马(*E. caballus*, 奇蹄目)作为外类群(outgroup)。6个类群的参考物种与马的基因组也按照上述方法,再次鉴定“一对一共线关系”的直系同源基因,进而构建不同亚目和不同亚科基部物种之间的进化树。最后,将6个类群各自的进化树拼接到该基准进化树上,获得偶蹄目物种完整的进化树。

## 1.3 反刍动物代表性物种的全基因组基因共线性的建立

选取在生态分布上较为广泛且染色体数目差异较大的7个反刍动物(梅花鹿、马鹿、獐、黄牛、水牛、家山羊和绵羊)进行了全基因组共线性分析。基于进化树的构建结果,使用 Diamond 软件在两两物种间,进行氨基酸序列比对,获得了不同物种基因间的序列相似性,挑选出相似度大于50%的基因,利用 MCSanX 软件建立两两物种间的基因共线性关系,通过 NGenomeSyn<sup>[40]</sup> v1.41 软件进行可视化,进而建立7个物种之间的染色体共线关系。通过对 MCSanX 软件生成的基因共线性文件进行人工识别,最终确定物种间共线基因区域的起始和终止位置。

## 1.4 染色体着丝粒区域以及特征性重复序列单元的确定

为了鉴定染色体上的着丝粒区域,利用由本实验室测定的獐 CENP-A 抗体的 CUT&Tag 实验数据对选取的7个反刍动物基因组进行了全基因组搜索,鉴定潜在的富集 CENP-A 核小体蛋白的区域,该区域的旁邻区域为染色体的候选着丝粒功能区。具体来说,用 fastp<sup>[41]</sup> v0.23.2 软件对 CUT&Tag 的原始测序数据去除接头序列和低质量的序列得到高质量读段,使用 Bowtie2<sup>[42]</sup> v2.3.2 软件,运行参数为“-end-to-end -very-sensitive -no-mixed -no-discordant -phred33 -I 10 -X 700”,将筛选后的 CUT&Tag 测序数据比对到基因组序列上,得到了每条读段在基因组上的位置。之后使用 MACS2<sup>[43]</sup> v2.2.9.1 软件进行峰响应后,在 IGV<sup>[44]</sup> v2.17.4 软件里进行可视化。

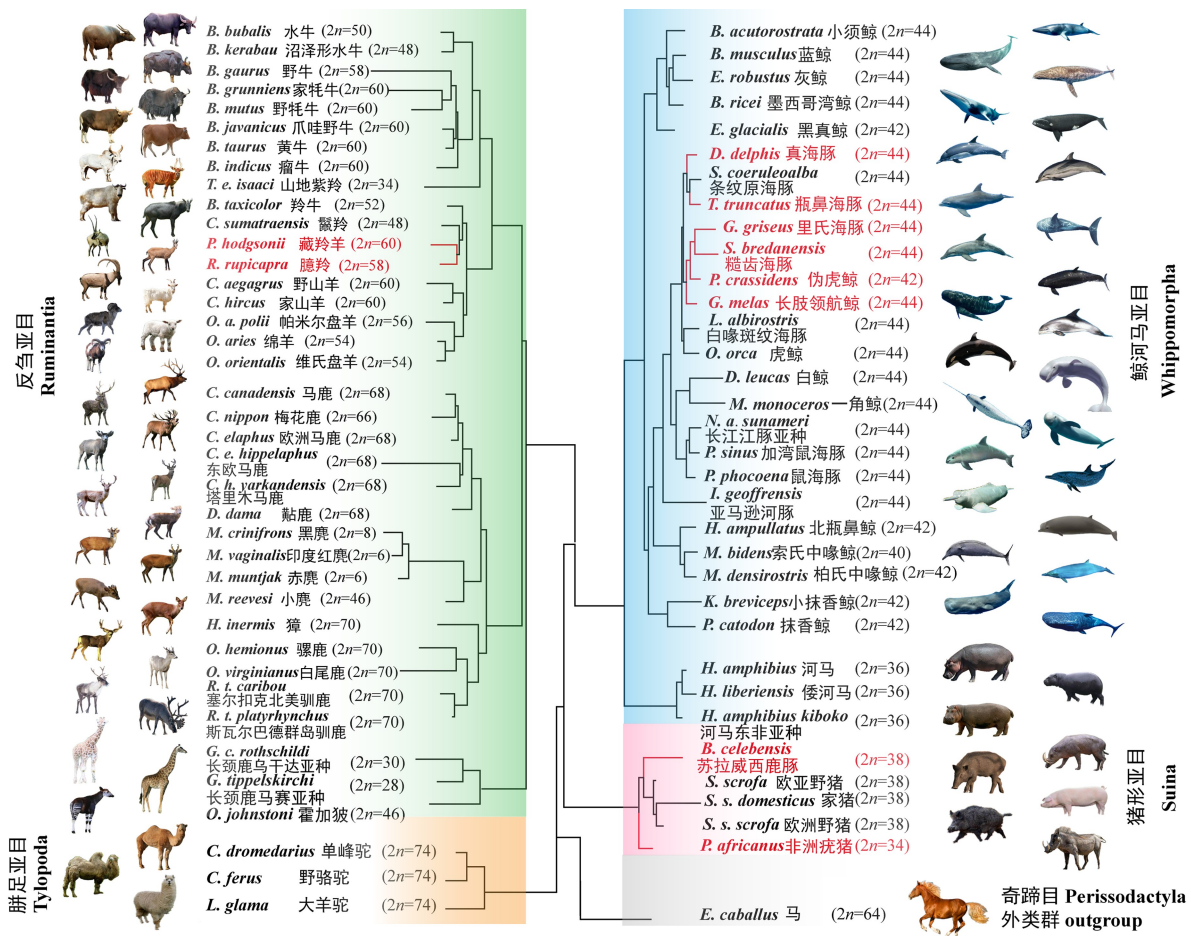
为了预测着丝粒区域的重复单元,使用 TRF<sup>[45]</sup> v4.10.0 软件对发生融合或断裂的染色体区域进行重复单元扫描,运行参数为“2 7 7 80 10 50 500”,搜索具有以下结构的区域:(1)重复单元长度在20个碱基以上;(2)重复次数超过50次。对这些区域进行人工检查后,确定着丝粒的候选区域。

# 2 结果

## 2.1 基于基因组共线性建立偶蹄目进化树

建立基因共线性是了解物种间染色体演化过程的重要手段。但在多个物种之间构建基因共线性时,往往会由于不同物种间基因序列相似度和排列顺序的剧烈变化,只能获得有限数量的共线区块,这影响了染色体水平的整体结构分析。因此,我们按亲缘关系将偶蹄目分成了不同的6个类群,类群内物种的亲缘关系相近,确保能够获得不同类群下所有物种共有的直系同源基因。在18个牛科物种的总计442944个蛋白编码基因中,通过建立全基因组基因共线性的方法,共鉴定到了8412个所有物种共有且均为单拷贝的“一对一直系同源基因”家族,这些基因家族覆盖18个牛科物种的151416个

基因。采用同样的方法,在与牛科存在较近亲缘关系的15个鹿科物种的423435个蛋白编码基因中,鉴定出包含112290个基因的共7486个“一对一直系同源基因”家族。除反刍动物外,在鲸河马亚目下的28个已测序物种的基因组中,亦构建出8242个“一对一直系同源基因”家族,利用这些“一对一直系同源基因”家族分别对相应的类群进行物种进化树的构建,结果如图1所示,进化树中86%(62/72)的物种在进化树上的位置与原先的研究具有一致的拓扑结构,另有10个物种的进化位置发生细微变化,它们中大部分(6个)集中于鲸河马亚目的海豚科中。而海豚科内部分枝上的进化树拓扑结构发生的较大差异,可能是由于海豚科存在不同于其他近缘物种的进化速率和大规模染色体结构重排造成的。在其他物种中,我们发现藏羚羊与髯羚的亲缘关系,比其与羚牛的亲缘关系更近,而猪形亚目下的家猪与苏拉维西鹿豚的亲缘关系,比其与非洲疣猪的亲缘关系更近。



每个物种的染色体数量都已经标明,如马( $2n=64$ ),红色分枝显示该分枝与 timetree 数据库(<https://timetree.org/>)进化树的对应拓扑结构间存在差异。

图1 基于“一对一直系同源基因”构建的72个偶蹄目物种进化树

Fig. 1 Phylogenetic tree of 72 Artiodactyla species constructed based on the one-to-one orthologous genes

在分别完成6个类群进化树的构建后,同样对6个类群中靠近基部的参考物种以及马在内的共7个物种建立了全基因组共线关系。在这7个物种的共187794个编码蛋白的基因中,鉴定出了包含80122个基因的11446个“一对一直系同源基因”家族,构建的7个参考物种系统发育树很好地契合了先前的研究<sup>[46-47]</sup>。在完整的偶蹄目物种进化树中,胼足亚目作为偶蹄目中较为原始的类群被定位到了进化树基部,反刍动物作为偶蹄目中分布广泛、数量众多、具有高度分化趋势的类群被定位到了进化树的分枝末端,而之前从化石、形态和分子等多个方面研究确认的鲸目类群在进化树上的位置,也接近于偶蹄目中的反刍动物<sup>[48]</sup>,即与反刍动物类群位于同一分枝。

### 2.2 反刍动物的染色体共线性

通过建立大片段染色体基因共线性,可以推测各物种染色体演化过程中发生的重排现象。偶蹄目的大多数胼足亚目、猪形亚目和鲸河马亚目物种的染色体数量相近,如鲸河马亚目物种(除河马科外)染色体的数量绝大多数等于或接近其染色体基数  $2n=44$ (图 2)。相比较而言,反刍亚目物种间的染色体数目变化较大,一方面可能是由于反刍亚目物种间存在着较多的远缘杂交(如家山羊和绵羊、马鹿和梅花鹿),另一方面可能是因为大多数反刍亚目物种经历了长期的人工驯化。而鲸河马亚目物种,其水生环境可能很大程度上限制了远缘杂交的发生,使得物种之间的染色体核型基本保持着祖先物种的共同特点。

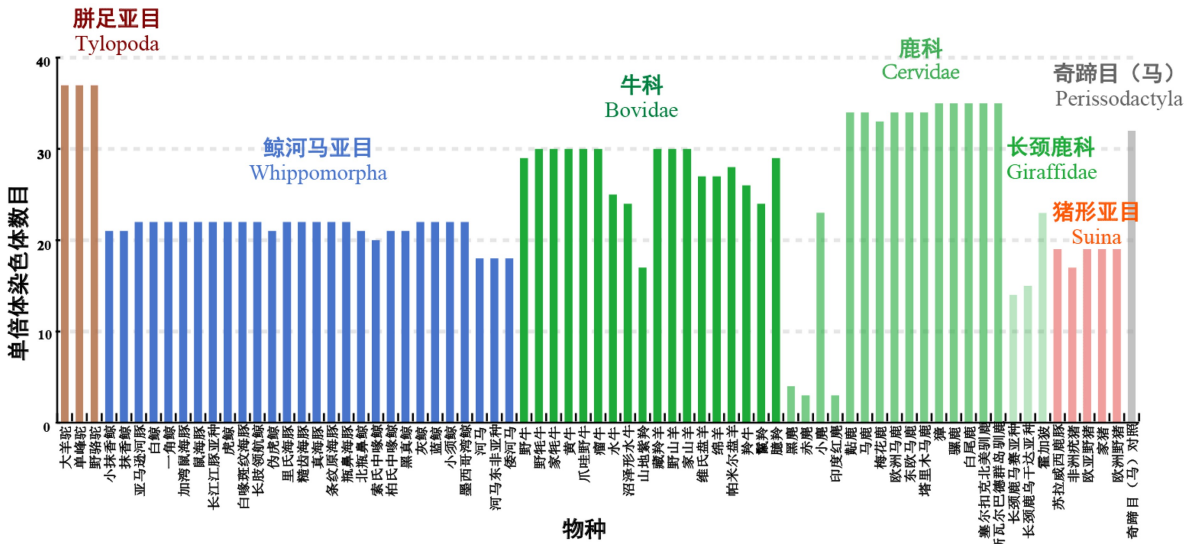


图 2 偶蹄目物种的单倍体染色体数目的比较

Fig. 2 Comparison of the haploid chromosome numbers among Artiodactyla species

为了进一步探究反刍动物染色体结构演化的机制,从进化树的主要分枝上挑选了靠近基部的一些典型物种,构建和比较了这些物种的染色体共线性。这些物种包括:分枝靠近基部的獐——作为较古老的鹿科獐亚科和空齿鹿亚科中的代表物种,生境分布广泛的马鹿和梅花鹿——作为鹿科鹿亚科中的代表物种,生态上分布广且靠近基部的黄牛和水牛——作为牛科牛亚科的代表物种,以及相对于基部较远但已经完成了 T2T 高质量基因组测序的家山羊和绵羊——作为牛科羊亚科的代表物种。在这 7 个物种之间,利用 106378 个共线基因构建了它们之间的全基因组范围的基因共线性。这些基因位于 1334 个共线区段,平均每个区段中约有 80 个共线基因,由此确定了共线染色体的对应关系(表 2),并进一步确定了这 7 个物种的染色体共线性(图 3)。

染色体共线性的结果显示,这 7 个物种间具有高度保守的基因排列顺序。近 40%(84/214)的染色体在所有物种中保持着稳定的同源关系,没有发生过大的融合或者断裂。其中,3 个鹿科物种 91%(93/102)的染色体之间和 4 个牛科物种 64%(72/112)的染色体之间基本上没有发生过大的染色体结构变化,保持了同样的染色体共线性。从具体的单条染色体来看,水牛的 1 号染色体在牛科物种内分别与黄牛的 27 和 1 号、家山羊的 27 和 1 号、绵羊的 26 和 1 号染色体间都存在严格的一对二共线关系,类似情况也发生在水牛的 2、3、4 和 5 号染色体中,这就使得水牛( $2n=50$ )比黄牛( $2n=60$ )和家山羊( $2n=60$ )少了 5 对染色体。同时,在绵羊与家山羊中,存在独特的 3 组二对一的染色体共线,使得绵羊( $2n=54$ )相比家山羊( $2n=60$ )少了 3 对染色体,但是相比水牛( $2n=50$ )多了 2 对染色体。在鹿科物种中,水牛的 1 号染色体分别对应于獐的 32、27 和 6 号、马鹿的 31、27 和 7 号以及梅花鹿的 30、24 和 8 号染色体。这种水牛染色体与其他物种一对三的共线关系,同样也发生在水牛的 2、3、4 号染色体中。除此之外,水牛的 5、7 和 10 号染色体在所有鹿科物种内都存在一对二的共线关系。在

獐中,也发现了1个二对一的共线现象(獐的7号对应着黄牛的28和26号),獐的7号染色体对应着马鹿的8号和梅花鹿9号染色体,因此獐( $2n=70$ )比水牛( $2n=50$ )多了10对染色体。类似地,在马鹿和梅花鹿中也发现了与之相似的二对一的染色体共线关系,这就解释了为什么马鹿( $2n=68$ )相比獐( $2n=70$ )少了1对染色体,而梅花鹿( $2n=66$ )相比獐少了2对染色体(表2)。这些染色体间对应关系的确定,揭示了这些物种的染色体间具有共同的起源,在进化过程中几乎没有产生新的染色体或者发生染色体的丢失,只是染色体在结构上发生了断裂或者融合,说明染色体的重排是生物多样性的重要驱动因素。

表2 7个反刍动物共线染色体的对应关系(染色体编号)

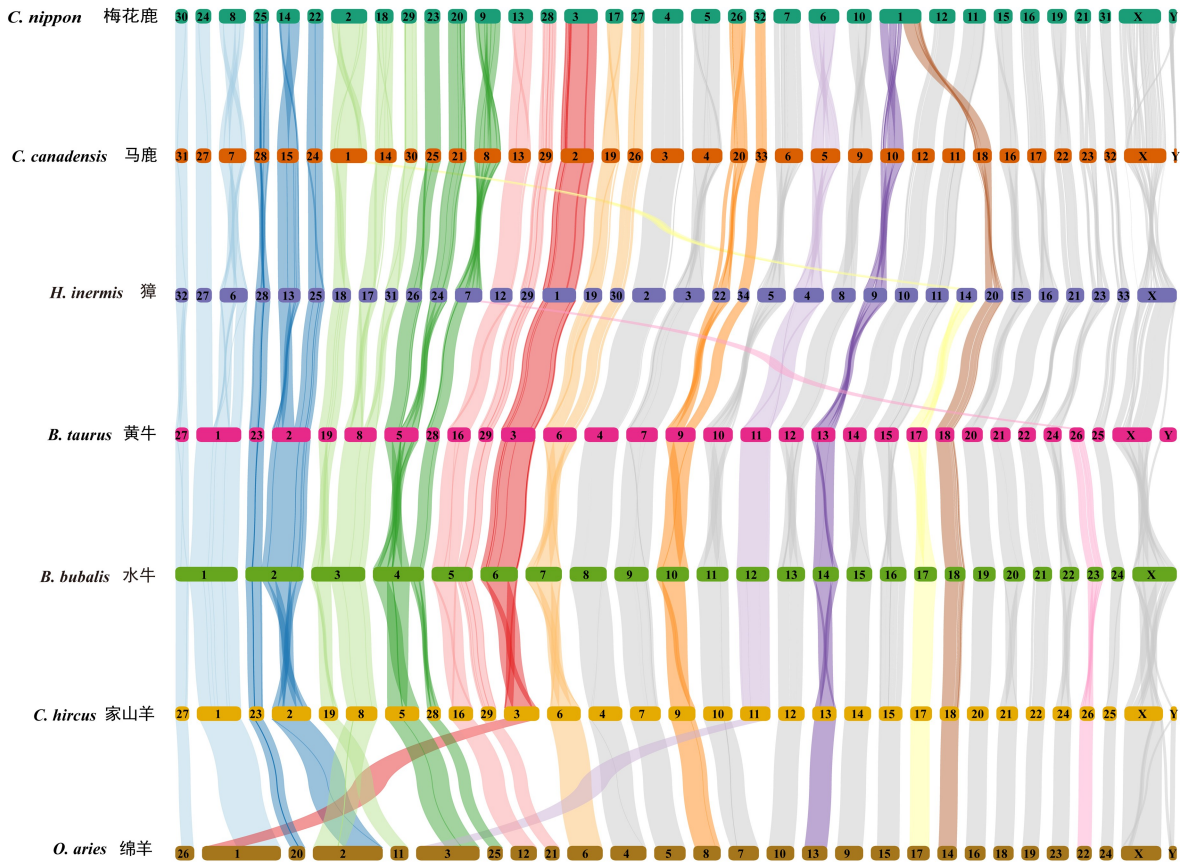
Table 2 Corresponding syntenic chromosomes (Chr ID) among the seven ruminants

水牛	黄牛	家山羊	绵羊	獐	马鹿	梅花鹿
1	27,1	27,1	26,1	32,27,6	31,27,7	30,24,8
2	23,2	23,2	20,2	28,13,25	28,15,24	25,14,22
3	19,8	19,8	11,2	18,17,31	1,14,30	2,18,29
4	5,28	5,28	3,25	26,24,7	25,21,8	23,20,9
5	16,29	16,29	12,21	12,29	13,29	13,28
6	3	3	1	1	2	3
7	6	6	6	19,30	19,26	17,27
8	4	4	4	2	3	4
9	7	7	5	3	4	5
10	9	9	8	22,34	20,33	26,32
11	10	10	7	5	6	7
12	11	11	3	4	5	6
13	12	12	10	8	9	10
14	13	13	13	9	10	1
15	14	14	9	10	12	12
16	15	15	15	11	11	11
17	17	17	17	14	1	2
18	18	18	14	20	18	1
19	20	20	16	15	16	15
20	21	21	18	16	17	16
21	22	22	19	21	22	19
22	24	24	23	23	23	21
23	26	26	22	7	8	9
24	25	25	24	33	32	31
X	X,Y	X,Y	X,Y	X	X,Y	X,Y

### 2.3 反刍动物染色体重排的分子机制

在不同物种的染色体重排过程中,染色体发生了融合或者断裂。我们发现,反刍动物染色体融合过程中,着丝粒区域的序列和结构单元表现出显著的规律性。融合后的着丝粒区域通常保留了原始着丝粒的序列特征,尽管染色体融合导致了着丝粒位置的变化,但着丝粒的结构单元在多个物种中保

持了较高的保守性。为了解这一机制,在已经确定了这7个典型反刍动物染色体共线性的基础上,对着丝粒区域的序列结构和重定位机制进行了分析。



染色体上的数字表示染色体的编号,不同颜色的连线表示物种间共线染色体的对应关系,其中灰色连线表示该染色体在不同物种间没有发生过重排。

图3 7个反刍动物基因组的染色体共线性

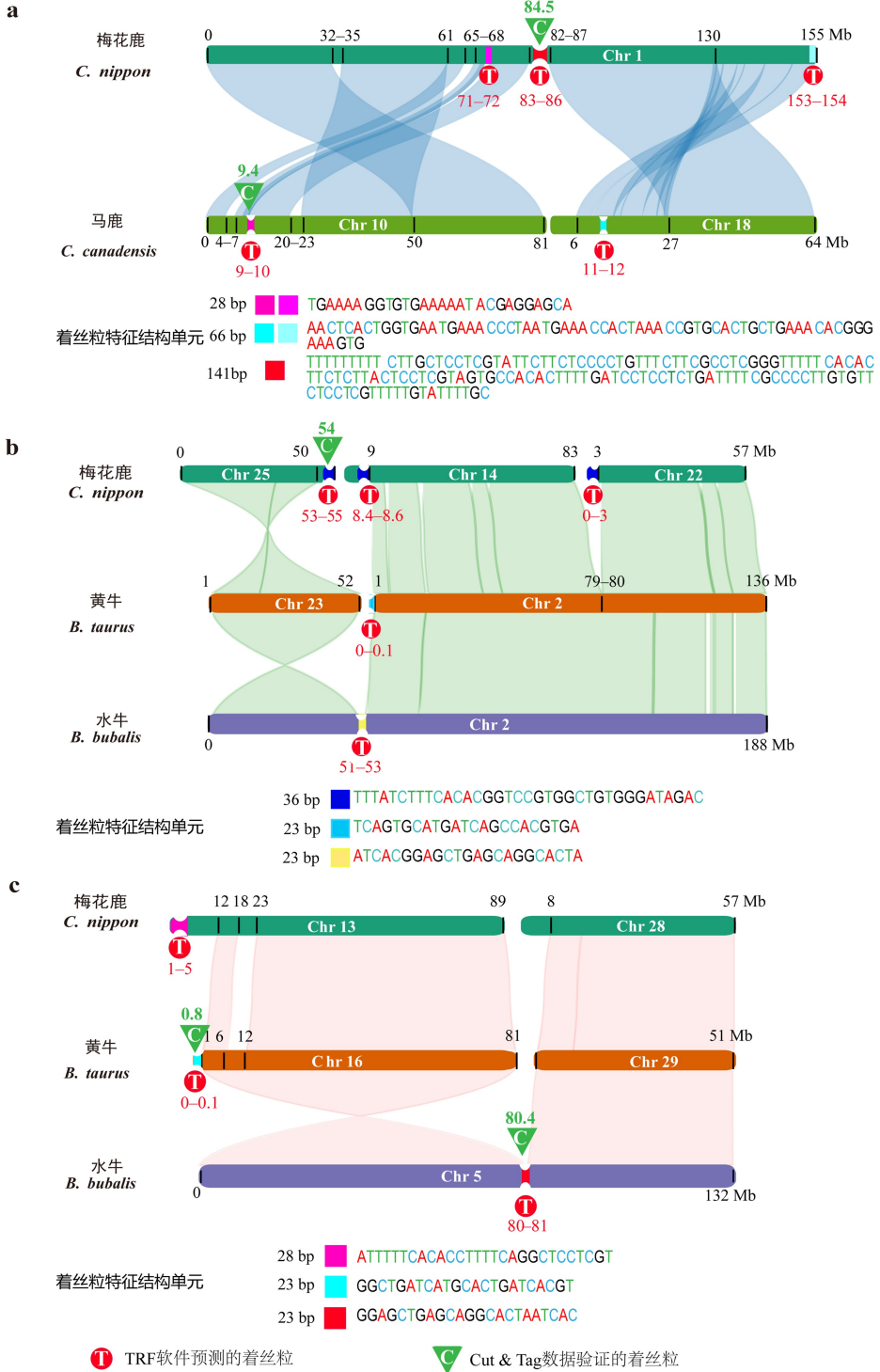
Fig. 3 Genome-wide chromosome synteny among the seven ruminants

梅花鹿的1号染色体(总长约155 Mb)0—82 Mb和87—155 Mb区域分别与马鹿的10号染色体(总长82 Mb)和18号染色体(总长64 Mb)存在共线关系。在其他反刍动物中,也都存在两条分别与马鹿的10号和18号染色体共线的染色体(图4a)。如獐的9号和20号染色体分别与马鹿的10号和18号染色体共线。因此可以推断,梅花鹿1号染色体的形成是由与马鹿10号和18号染色体有共同起源的两条染色体融合产生的,这是梅花鹿中特异发生的染色体融合事件。从染色体的共线区域和融合位置来看,染色体的融合位置发生在梅花鹿1号染色体的82—87 Mb区域。而CUT&Tag的实验数据恰好在84.5 Mb区域附近检测出了CENP-A蛋白的特异结合位点,这两个证据共同表明这个区域应该就是该染色体的功能性着丝粒区域。同时,对梅花鹿1号染色体的串联重复序列(TRF预测)结构进行搜索,结果显示在染色体的3个不同区域存在特殊的串联重复序列:(1)在71—72 Mb区域,检测出一个重复次数为22242、长度为28 bp的单元序列;(2)在83—86 Mb区域,检测出一个重复次数为48095、长度为28 bp的单元序列和一个重复次数为2051、长度为141 bp的单元序列;(3)在153—154 Mb区域,检测出一个重复次数为739、长度为66 bp的单元序列。除了CENP-A抗体的CUT&Tag实验结果和TRF的预测结果共同证明的梅花鹿1号染色体融合后形成的新的着丝粒功能性区域(在82—87 Mb,对应上述第2个串联重复区)之外,在71—72 Mb和153—154 Mb区域发现的高度串联重复序列,在位置上和序列特征上都恰恰对应的是马鹿10号染色体的9—10 Mb和

18号染色体的11—12 Mb这两个区域。在马鹿的这两个区域中,也分别检测到了序列相似、重复次数为8200、长度为28 bp和重复次数为478、长度为66 bp的单元序列。而且,在马鹿10号染色体的9 Mb附近存在着该染色体上唯一的CENP-A特异结合区域,表明9—10 Mb处是马鹿10号染色体的功能性着丝粒区域。由于非T2T组装水平的基因组丢失着丝粒结构是一个非常普遍的现象,我们没能在马鹿的18号染色体上发现CENP-A特异结合位点。概括来说,在梅花鹿1号染色体的融合过程中,没有继承融合前的两条染色体中的着丝粒区域,而是在融合处形成了新的着丝粒,之前旧的着丝粒区域特征序列依然较为完整地保留了下来,但很可能失去了着丝粒的功能。

在这7种反刍动物中,水牛的染色体数目明显少于其他6个物种。染色体的共线性结果显示,水牛的第1、2、3、4和5号染色体,分别与其他物种的2~3条染色体共线。以水牛2号染色体为例,它与黄牛的23和2号染色体共线。这两条染色体的结合点位于水牛2号染色体50—52 Mb区域(图4b)。TRF软件的预测结果显示,水牛2号染色体的51—53 Mb区域和与之对应的黄牛2号染色体末端,分别存在着一个重复次数为4379、长度为23 bp的单元序列和重复次数为3450、长度为23 bp的单元序列,而这两个串联重复相似性达到了75%(图4b中,黄牛23 bp单元序列的互补链与水牛23 bp单元序列相比)。鉴于染色体的其他区域都没有发现与之相似的串联重复单元,因此基本上可以确定,这两个串联重复区域分别是各自染色体的着丝粒区域,而且这两个着丝粒在序列上保守,说明具有共同的起源。当检查与其他反刍动物的染色体共线性以后发现,黄牛的2号染色体与大多数牛亚科物种中的一条染色体存在共线关系(如绵羊的2号染色体),而与大多数鹿亚科动物中的2条染色体存在对应关系(如梅花鹿中的14和22号染色体)。比较黄牛2号染色体与梅花鹿的14和22号染色体的共线关系,显示这两条染色体的结合区域位于黄牛2号染色体的79—80 Mb处。而TRF预测结果显示,在梅花鹿的14号染色体8.4—8.6 Mb区域附近存在一个重复次数为4774、长度为36 bp的单元序列;在22号染色体0—3 Mb区域,也存在一个重复次数为31747、长度为36 bp的单元序列,这些是潜在的着丝粒区域。其中,梅花鹿的14号染色体8.4—8.6 Mb区域的重复单元序列的后半部分(图4b中,36 bp序列的反向互补后的前半部分)与黄牛23 bp序列高度相似,表明梅花鹿的14号染色体、黄牛的2号染色体和水牛的2号染色体具有相同的着丝粒起源。在与梅花鹿22号染色体对应的黄牛2号染色体的79—80 Mb区域,没有发现相似的着丝粒的特征序列,可能是染色体进化过程中,相应的着丝粒区域丢失了。与水牛2号染色体短臂共线的是黄牛23号染色体和梅花鹿25号染色体,与之相似的是在牛亚科和鹿亚科的大多数物种中,这些共线染色体都单独存在,没有发生染色体融合或断裂。在梅花鹿25号染色体的54 Mb处,存在着CENP-A蛋白的结合位点,这也得到了TRF软件预测结果的支持,即在该染色体的53—55 Mb区域位置存在着一个重复次数为19336、长度为36 bp的单元序列。这个36 bp的串联重复单元与梅花鹿的14号染色体8.4—8.6 Mb区域内存在的36 bp单元具有高度的相似性,说明梅花鹿的25号染色体与水牛2号染色体的着丝粒有共同起源。

为了获得更多着丝粒进化的证据,对包括水牛5号染色体在内的其他染色体进行了分析。以水牛5号染色体(总长约132 Mb)为例,它的0—80 Mb和81—132 Mb区域分别与黄牛的16号染色体(总长约81 Mb)和29号染色体(总长约51 Mb)共线(图4c)。黄牛这两条染色体又分别与梅花鹿的13号(总长约89 Mb)和28号染色体(总长约57 Mb)共线。从水牛5号染色体及其对应的共线染色体来看,着丝粒位置和序列也存在着保守性。黄牛的16号和29号染色体在水牛5号染色体上的结合区域位于80—81 Mb。CUT&Tag实验数据显示,染色体的80.4 Mb附近存在唯一的CENP-A特异结合位点。与该区域的重复单元序列相似的结构也存在于黄牛16号染色体的0—0.1 Mb区域和29号染色体的0—1 Mb区域。CUT&Tag数据也表明,在黄牛16号染色体的0.8 Mb区域存在CENP-A的特异结合位点。巧合的是,在与黄牛16号染色体0—1 Mb共线的梅花鹿13号染色体的0—12 Mb区域内也存在着一个相似的、重复次数为66045、长度为28 bp的单元序列,说明它们还保持着相似染色体及着丝粒结构。



a. 梅花鹿 1 号染色体的融合模型；b. 水牛 2 号染色体的融合模型；  
c. 水牛 5 号染色体的融合模型。染色体上的着丝粒位置对应相同颜色的着丝粒特征序列。

图 4 反刍动物染色体的重排机制

Fig. 4 Mechanism of chromosome rearrangement in ruminants

从这些反刍动物的染色体共线性来看,在水牛 1、3、4 号染色体中,都经历了与水牛 2 号染色体相似的重排过程,如水牛的 1 号染色体分别对应黄牛 27 号染色体和 1 号染色体,其中黄牛 1 号染色体又分别对应着梅花鹿 24 号和 8 号染色体(图 3)。结合染色体的共线关系和反刍动物的系统发育,我

们推测,反刍动物祖先的染色体数为35对70条,在牛科中先后发生了两轮染色体的融合,其中第一轮融合发生在6对染色体中,如梅花鹿的8号和24号融合为黄牛的1号,梅花鹿的14号和22号融合为黄牛的2号,梅花鹿的20号和23号融合为黄牛的5号,梅花鹿的18号和29号融合为黄牛的8号,梅花鹿的17号和27号融合为黄牛的6号,以及梅花鹿的26号和32号融合为黄牛的9号(图3)。这一轮融合使得染色体的数量减少为30对60条,这也是目前大多数牛科物种的染色体基数 $2n=60$ 。随后,在水牛等少数物种中发生了特有的第二轮融合,形成了水牛的1、2、3、4和5号染色体,即黄牛的1号和27号融合为水牛的1号,黄牛的2号和23号融合为水牛的2号,黄牛的8号和19号融合为水牛的3号,黄牛的5号和28号融合为水牛的4号,以及黄牛的16号和29号融合为水牛的5号。这一轮融合使得染色体的数量进一步减少为25对50条。而在鹿科中,没有观察到这两轮的染色体融合事件。在染色体的重排过程中,染色体的融合(或断裂)大多发生于着丝粒区域,而且大部分着丝粒的结构和功能在染色体重排过程中被保留下来。从着丝粒的特征序列来看,多富含GC,暗示着着丝粒区域的表观修饰对于着丝粒结构和定位具有非常重要的作用。

### 3 讨论

反刍动物在漫长的演化历程中具有独特的形态特征,如多室胃和骨质角<sup>[49]</sup>,这极大地增加了它们对生境的适应能力,使得反刍动物在不同海拔和纬度的生境中都广泛分布<sup>[50]</sup>。而在牛科和鹿科中大量染色体重排现象的存在,可能就表明了染色体融合和断裂对物种进化和多样性的重要影响。这种染色体重排现象是减数分裂时期染色体间的同源重组导致了染色体的融合和断裂<sup>[51]</sup>,这种融合和断裂如果发生在远缘动物杂交时,就会使得配子染色体中的遗传信息发生嵌合,如果嵌合了亲本遗传信息的染色体可以在杂交后代中得到稳定遗传,并与近缘物种存在生殖隔离,那么可能也就意味着新物种的初步形成。事实上,有大量研究证明,自然生境中存在大量的杂交物种,杂交物种与新物种的形成之间存在一定的关系<sup>[52]</sup>。除此之外,在鳞翅目中染色体的融合和断裂也被认为是与物种的形成相关<sup>[53]</sup>。这些研究结论都与我们在反刍动物中的发现不谋而合。

反刍动物着丝粒区域存在大量的串联重复序列,其中组成着丝粒序列的重复单元只会富集在着丝粒区域,不存在基因组其他区域中。当染色体发生融合时,新融合的染色体上可能同时存在着融合前的多条染色体的着丝粒区域。之前的研究表明,在新融合的染色体中往往只会保留一个着丝粒区域发挥功能,其余的着丝粒区域不再具备着丝粒功能<sup>[21-23]</sup>。发生在水牛2号染色体上的融合现象可能就是这种进化机制,当然这还有待于T2T基因组数据等更多证据来证实。有研究认为,有些染色体区域也可以通过表观遗传修饰形成新的功能性着丝粒,这一过程伴随着染色质重塑,最终使得着丝粒位置发生重定位<sup>[25]</sup>,发生在梅花鹿1号染色体上的着丝粒位置变化可能就是这种类型。这些特殊的融合现象发生在偶蹄目反刍动物这类长期面临远缘杂交的物种中,可能就表明远缘杂交导致了着丝粒区域的改变和染色体结构的重组,进而演化出了新的具有不同核型的物种。

### 参考文献:

- [1] WILSON D E, REEDER D M. Mammal species of the world: A taxonomic and geographic reference[M]. Charles Village: Johns Hopkins University Press, 2005.
- [2] JANIS C M, SCOTT K M. The interrelationship of higher ruminant families; With special emphasis on the members of the Cervoidea[J]. New York: American Museum of Natural History, 1987.
- [3] HOFMANN R R. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: A comparative view of their digestive system[J]. Oecologia, 1989, 78(4): 443-457.
- [4] FERGUSON-SMITH M A, TRIFONOV V. Mammalian karyotype evolution[J]. Nature Reviews Genetics, 2007, 8(12): 950-962.
- [5] POPESCU P C. Chromosomes of the cow and bull[J]. Advances in Veterinary Science and Comparative Medicine, 1990, 34: 41-71.

- [6] LONG S E. Chromosomes of sheep and goats[J]. *Advances in Veterinary Science and Comparative Medicine*, 1990, 34:109-129.
- [7] ZIVKOVIĆ S, JOVANOVIĆ V, ISAKOVIĆ I, et al. Chromosome complement of the European wild pig (*Sus scrofa* L.)[J]. *Experientia*, 1971, 27(2):224-226.
- [8] PROSKURYAKOVA A A, KULEMZINA A I, PERELMAN P L, et al. Comparative chromosome mapping of musk Ox and the X chromosome among some bovidae species[J]. *Genes*, 2019, 10(11):857.
- [9] OLAGUNJU T A, ROSEN B D, NEIBERGS H L, et al. Telomere-to-telomere assemblies of cattle and sheep Y-chromosomes uncover divergent structure and gene content[J]. *Nature Communications*, 2024, 15(1):8277.
- [10] YANG J, LI W R, LÜ F H, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments[J]. *Molecular Biology and Evolution*, 2016, 33(10):2576-2592.
- [11] JI Q M, XIN J W, CHAI Z X, et al. A chromosome-scale reference genome and genome-wide genetic variations elucidate adaptation in yak[J]. *Molecular Ecology Resources*, 2021, 21(1):201-211.
- [12] LIANG X L, DUAN Q J, LI B W, et al. Genomic structural variation contributes to evolved changes in gene expression in high-altitude Tibetan sheep[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2024, 121(27):e2322291121.
- [13] SHANG W H, HORI T, TOYODA A, et al. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences[J]. *Genome Research*, 2010, 20(9):1219-1228.
- [14] GONG Z Y, WU Y F, KOBLÍZKOVÁ A, et al. Repeatless and repeat-based centromeres in potato: Implications for centromere evolution[J]. *The Plant Cell*, 2012, 24(9):3559-3574.
- [15] TALBERT P B, HENIKOFF S. What makes a centromere? [J]. *Experimental Cell Research*, 2020, 389(2):111895.
- [16] CAPPELLETTI E, PIRAS F M, SOLA L, et al. Robertsonian fusion and centromere repositioning contributed to the formation of satellite-free centromeres during the evolution of zebras[J]. *Molecular Biology and Evolution*, 2022, 39(8):msac162.
- [17] MELTERS D P, BRADNAM K R, YOUNG H A, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution[J]. *Genome Biology*, 2013, 14(1):R10.
- [18] SAMONTE R V, RAMESH K H, VERMA R S. Comparative mapping of human alphoid satellite DNA repeat sequences in the great apes[J]. *Genetica*, 1997, 101(2):97-104.
- [19] ALKAN C, CARDONE M F, CATAACCHIO C R, et al. Genome-wide characterization of centromeric satellites from multiple mammalian genomes[J]. *Genome Research*, 2011, 21(1):137-145.
- [20] TORRAS-LLORT M, MORENO-MORENO O, AZORÍN F. Focus on the centre: The role of chromatin on the regulation of centromere identity and function[J]. *The EMBO Journal*, 2009, 28(16):2337-2348.
- [21] AGUDO M, ABAD J P, MOLINA I, et al. A dicentric chromosome of *Drosophila melanogaster* showing alternate centromere inactivation[J]. *Chromosoma*, 2000, 109(3):190-196.
- [22] SULLIVAN B A, SCHWARTZ S. Identification of centromeric antigens in dicentric Robertsonian translocations; CENP-C and CENP-E are necessary components of functional centromeres[J]. *Human Molecular Genetics*, 1995, 4(12):2189-2197.
- [23] THROWER D A, STEMPLE J, YE H E, et al. Nuclear oscillations and nuclear filament formation accompany single-strand annealing repair of a dicentric chromosome in *Saccharomyces cerevisiae*[J]. *Journal of Cell Science*, 2003, 116(3):561-569.
- [24] BODOR D L, MATA J F, SERGEEV M, et al. The quantitative architecture of centromeric chromatin[J]. *ELife*, 2014, 3:e02137.
- [25] WARBURTON P E. Chromosomal dynamics of human neocentromere formation[J]. *Chromosome Research*, 2004, 12(6):617-626.
- [26] HOFFMANN S, DUMONT M, BARRA V, et al. CENP-A is dispensable for mitotic centromere function after initial centromere/kinetochore assembly[J]. *Cell Reports*, 2016, 17(9):2394-2404.
- [27] MCKINLEY K L, CHEESEMAN I M. The molecular basis for centromere identity and function[J]. *Nature Reviews Molecular Cell Biology*, 2016, 17(1):16-29.

- [28] MÜLLER S, ALMOUZNI G. A network of players in H3 histone variant deposition and maintenance at centromeres[J]. *Biochimica et Biophysica Acta*, 2014, 1839(3): 241-250.
- [29] CAPPELLETTI E, PIRAS F M, SOLA L, et al. The localization of centromere protein A is conserved among tissues[J]. *Communications Biology*, 2023, 6(1): 963.
- [30] ESCUDEIRO A, FERREIRA D, MENDES-DA-SILVA A, et al. Bovine satellite DNAs: A history of the evolution of complexity and its impact in the Bovidae family[J]. *The European Zoological Journal*, 2019, 86(1): 20-37.
- [31] VOZDOVA M, KUBICKOVA S, MARTÍNKOVÁ N, et al. Satellite DNA in neotropical deer species[J]. *Genes*, 2021, 12(1): 123.
- [32] FILIPSKI J, THIERY J P, BERNARDI G. An analysis of the bovine genome by Cs<sub>2</sub>SO<sub>4</sub>-Ag<sup>+</sup> density gradient centrifugation[J]. *Journal of Molecular Biology*, 1973, 80(1): 177-197.
- [33] LUO L Y, WU H, ZHAO L M, et al. Telomere-to-telomere sheep genome assembly reveals new variants associated with wool fineness trait[J/OL]. *BioRxiv: The Preprint Server for Biology*, 2024. (2024-07-23) [2024-12-23]. <https://www.biorxiv.org/content/10.1101/2024.07.21.604451v1>.
- [34] WU H, LUO L Y, ZHANG Y H, et al. Telomere-to-telomere genome assembly of a male goat reveals variants associated with cashmere traits[J]. *Nature Communications*, 2024, 15(1): 10041.
- [35] KUMAR S, STECHER G, SULESKI M, et al. TimeTree: A resource for timelines, timetrees, and divergence times[J]. *Molecular Biology and Evolution*, 2017, 34(7): 1812-1819.
- [36] BUCHFINK B, XIE C, HUSON D H. Fast and sensitive protein alignment using DIAMOND[J]. *Nature Methods*, 2015, 12(1): 59-60.
- [37] WANG Y P, TANG H B, DEBARRY J D, et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. *Nucleic Acids Research*, 2012, 40(7): e49.
- [38] KATOH K, MISAWA K Z H, KUMA K I, et al. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform[J]. *Nucleic Acids Research*, 2002, 30(14): 3059-3066.
- [39] NGUYEN L T, SCHMIDT H A, VON HAESLER A, et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular Biology and Evolution*, 2015, 32(1): 268-274.
- [40] HE W M, YANG J, JING Y, et al. NGenomeSyn: An easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes[J]. *Bioinformatics*, 2023, 39(3): btad121.
- [41] CHEN S F, ZHOU Y Q, CHEN Y R, et al. Fastp: An ultra-fast all-in-one FASTQ preprocessor[J]. *Bioinformatics*, 2018, 34(17): i884-i890.
- [42] LANGMEAD B, SALZBERG S L. Fast gapped-read alignment with Bowtie 2[J]. *Nature Methods*, 2012, 9(4): 357-359.
- [43] GASPAR J M. Improved peak-calling with MACS2[J/OL]. *BioRxiv: The Preprint Server for Biology*, 2018. (2018-12-17) [2024-12-17]. <https://www.biorxiv.org/content/10.1101/496521v1>.
- [44] THORVALDSDÓTTIR H, ROBINSON J T, MESIROV J P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration[J]. *Briefings in Bioinformatics*, 2013, 14(2): 178-192.
- [45] BENSON G. Tandem repeats finder: A program to analyze DNA sequences[J]. *Nucleic Acids Research*, 1999, 27(2): 573-580.
- [46] CALAMARI Z T. Total evidence phylogenetic analysis supports new morphological synapomorphies for Bovidae (Mammalia, Artiodactyla)[J]. *American Museum Novitates*, 2021, 2021(3970): 1-38.
- [47] YIN Y, FAN H Z, ZHOU B T, et al. Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer[J]. *Nature Communications*, 2021, 12(1): 6858.
- [48] GEISLER J H. New morphological evidence for the phylogeny of Artiodactyla, Cetacea, and Mesonychidae[J]. *American Museum Novitates*, 2001, 2001(3344): 1-53.
- [49] DAVIS E B, BRAKORA K A, LEE A H. Evolution of ruminant headgear: A review[J]. *Proceedings of the Royal Society B: Biological Sciences*, 2011, 278(1720): 2857-2865.
- [50] WANG B, CHEN L, WANG W. Genomic insights into ruminant evolution: From past to future prospects[J]. *Zoological Research*, 2019, 40(6): 476-487.
- [51] RIESEBERG L H. Chromosomal rearrangements and speciation[J]. *Trends in Ecology & Evolution*, 2001, 16

(7):351-358.

- [52] BULLINI L. Speciation by hybridization in animals[J]. Italian Journal of Zoology, 1985, 52(1/2):121-137.
- [53] DE VOS J M, AUGUSTIJNEN H, BÄTSCHER L, et al. Speciation through chromosomal fusion and fission in Lepidoptera[J]. Philosophical Transactions of the Royal Society B, 2020, 375(1806):20190539.

## A Study of Chromosome Rearrangement and Relevant Centromere Relocation in Artiodactyla

ZHAO Hengquan<sup>1,2</sup>, LIU Fangyuan<sup>3</sup>, NASHUN Buhe<sup>3</sup>, LU Ying<sup>1,2</sup>, LI Xihe<sup>4,5</sup>

- (1. Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China;
2. Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, China;
3. School of Life Sciences, Inner Mongolia University, Hohhot 010070, China;
4. Research Center for Animal Genetic Resources of Mongolian Plateau, Inner Mongolia University, Hohhot 010021, China;
5. Inner Mongolia Saikexing Institute of Breeding and Reproductive Biotechnology in Domestic Animal, Hohhot 011517, China)

**Abstract:** As the main source of animals domesticated by humans (such as ruminants), Artiodactyla serves as crucial materials for studying chromosomal evolution under natural and artificial selection. We identified the orthologous gene clusters by construction of genome-wide gene colinearity among the 72 published chromosomal-level Artiodactyla genomes, which enabled us to build an integrated phylogenetic tree, which was in high agreement with known topology, in which 86% (62/72) of the species were localized at the same clade. The corresponding chromosomal synteny among seven representative species of ruminants was determined by the constructed gene colinearity. By combining the predictions by TRF (Tandem repeats finder) software and CUT&Tag (Cleavage under Targets and Tagmentation) data with centromere protein A (CENP-A) antibodies, we localized the centromeric regions of chromosomes in the representative ruminant species of sika deer (*Cervus nippon*), buffalo (*Bubalus bubalis*), and cattle (*Bos taurus*). The result reveals occurrence of the chromosome rearrangements (fusions or breaks) in the centromere region, which suggests their shared common origins. In addition, new functional centromere regions evolve during the fusion of a few chromosomes, while traces of their ancestral centromere sequences could still be observed at the corresponding regions in their syntenic chromosomes of other species. The characteristic tandem repeats identified in the centromere region are indicative of a high GC content, suggesting that the functional activation and silencing of the centromeres are probably regulated by epigenetics. This study preliminarily reveals the molecular mechanism of chromosome arrangement among different species, which speculates that chromosome diversity caused by chromosome arrangement is an important driving force for the formation of species diversity in ruminants.

**Key words:** Artiodactyla; chromosome rearrangement; genome; centromere; evolution