



畜牧科学研究

新版本鸡基因组 GGswu1 的评价与应用

寇文艳, 王菲, 姜雨*

(西北农林科技大学 动物科技学院, 陕西杨凌 712100)

摘要: 鸡基因组的完整性和准确性对于脊椎动物进化研究以及鸟类生物学特性的理解具有重要意义。同时, 它在家禽的遗传机制解析和遗传育种改良方面也发挥着至关重要的作用。近期公布了接近完整的鸡基因组 GGswu1, 但以此基因组为参考的遗传变异研究尚未广泛开展。本研究旨在对 GGswu1, 以及现行的鸡参考基因组 GRCg7b 的差异和优势进行评估。基于所收集的 60 份公开可用的鸡全基因组重测序数据和 103 份鸡转录组数据, 对两个基因组在组装质量、比对率、SNPs 检出能力以及基因组组装的完整性等方面进行了全面的评估与比较。结果表明, 与 GRCg7b 相比, GGswu1 在组装质量上实现了显著提升, 基因组组装的完整性也得到了改善, 显示出在遗传变异解析方面的较大潜力。这些发现为更精准地理解这两个基因组组装版本之间的差异和优势提供了帮助, 可为未来鸡基因组的功能性研究和家禽遗传资源的深入挖掘提供一定参考。

关键词: 鸡; 参考基因组; 基因; 遗传变异

[中图分类号] S831.2 [文献标志码] A [文章编号] 1004-6704(2024)-04-0029-08

Evaluation and Application of the New Chicken Genome Version GGswu1

KOU Wenyan, WANG Fei, JIANG Yu*

(College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: The integrity and accuracy of the chicken genome hold significant implications for the study of vertebrate evolution and the understanding of avian biological characteristics. Additionally, it plays an essential role in deciphering the genetic mechanisms of poultry and in the improvement of genetic breeding. The nearly complete chicken genome map was recently published, but genetic variation studies using this genome as a reference have not been widely conducted. This study aims to evaluate the differences and advantages of GGswu1 and the current chicken reference genome GRCg7b. Based on 60 publicly available chicken whole-genome resequencing data and 103 chicken transcriptome data collected, a comprehensive assessment and comparison of the two reference genomes were conducted in terms of assembly quality, alignment rate, SNPs detection capability, and completeness of gene assembly. The results indicate that, compared to GRCg7b, GGswu1 has achieved a significant improvement in assembly quality and has also improved the completeness of gene assembly to some extent, demonstrating considerable potential in genetic variation analysis. These findings aid with a more precise understanding of the differences and advantages between these two genome assembly versions and offer a reference for future functional research of the chicken genome and in-depth exploration of poultry genetic resources.

Key words: chicken; reference genome; genes; genetic variation

[收稿日期] 2024-06-25

[基金项目] 国家重点研发计划(2022YFD1301600); 陕西省自然科学基金基础研究计划项目(2022JQ-171); 陕西省博士后科研基金项目(2023BSHYDZZ86)

[第一作者] 寇文艳(1999-), 女, 主要从事动物遗传育种研究。
E-mail: kouwenyan99@163.com

*[通信作者] 姜雨, E-mail: yu.jiang@nwfau.edu.cn

随着高通量测序技术的飞速发展, 各类生物物种的基因组测序工作呈现出爆发式增长, 尤其是在关键性状基因定位及其遗传机制解析等前沿研究领域。近年来, 众多学者利用表型测定和变异挖掘等手段, 对畜禽品种进行了深入研究, 不仅丰富了对生

物多样性的理解,也为畜禽育种和遗传改良提供了宝贵的科学依据。解析关键性状的遗传机制,首先需要在全基因组范围挖掘与该性状紧密相关的关键变异,在尚未实现群体基因组组装的物种中,变异的检测都是基于该物种的参考基因组,即将测序得到的序列与参考基因组——“标准地图”进行比对。参考基因组质量的高低直接决定了变异鉴定的全面性和准确性,从而影响对功能基因和性状遗传机制的解析^[1]。因此,一个物种是否拥有高质量的参考基因组是深入解析其遗传与表型关系的重要前提^[2]。

鸡(*Gallus gallus*)是最重要的农业动物和模式生物之一,也是人类蛋白质消费的主要来源,全球鸡肉消费量在肉类中占比最高。同时,鸡也是最早进行基因组测序的农业动物,这可能与其基因组较小的尺寸有关,其基因组大小仅为家畜的 1/2 至 1/3^[3]。随着测序技术和组装水平的发展,多个版本的红原鸡(galgal2-galgal5 和 GRCg6a)和商品鸡(GRCg7b 和 GRCg7w)基因组相继更新和发布^[4]。这些基因组被广泛应用于鸡的遗传变异图谱构建、遗传多样性分析、关键性状候选因果变异挖掘,以及驯化和进化历史研究等领域^[5-8]。

然而,尽管二十年来不断努力构建和完善鸡的基因组序列,由于鸡基因组中小微染色体较高的 GC 含量、重复序列含量以及 DNA 甲基化水平等因素^[9],鸡的参考基因组仍未完全完善,存在一些限制全面分析的障碍。例如,当前的参考基因组 GRCg7b 包含 39 条常染色体,与鸡的核型分析结果(2n=78)不符,序列中仍存在较多的组装间隙,着丝粒和端粒序列有待完善。此外,GRCg7b 基因组中有 12.2 Mb 的未定位序列,这些序列与主染色体分开表示,许多基因组分析会排除它们,以避免识别错误的变异和错误的调控关系。直到 2023 年,才公布了首个基本完整的鸡基因组序列^[10]。然而,当前鸡的遗传变异研究多基于参考基因组 GRCg7b,以 GGswu1 基因组为参考的遗传变异研究尚未广泛开展,鸡的完整基因组序列对于遗传变异解析能力的改进还有待探索。因此,本研究收集了两种参考基因组,并对其组装质量、重测序和转录组数据比对率、遗传变异检出能力以及基因组组装的完整性等方面进行了评估与比较,以期全面、准确解析鸡的遗传变异提供参考。

1 材料和方法

1.1 试验样品收集

本研究从 NCBI 数据库([https://www.ncbi.](https://www.ncbi.nlm.nih.gov)

[nlm.nih.gov](https://www.ncbi.nlm.nih.gov))下载了 GRCg7b 和 GGswu1 的基因组序列。同时,本研究共收集到 60 个来自世界 48 个国家或地区的 56 个代表性鸡品种的全基因组重测序(Whole genome resequencing, WGS)数据,平均测序深度为 $10 \times$ ^[11-15]。此外,还收集了来自心脏、肾脏、皮肤、小脑、胸腺、脾脏、肝脏、结肠、肌肉和皮肤等多种组织的 103 个转录组(RNA-seq)数据,以代表丰富的组织多样性^[16-17]。

1.2 基因组组装质量评估

本研究使用 GAEP (v1.2.3)^[18]评估基因组的基础信息,并使用 python 脚本 getgaps.py 统计基因组的间隙(gap)的位置和长度。使用 BUSCO (Benchmarking Universal Single-Copy Orthologs)软件(v 5.4.7)^[19]对基因组的完整性进行评估,其中基因预测采用 augustus 程序,参考库为包含 8 338 个基因的鸟类 aves_odb10 基因数据库。

1.3 重复序列注释

本研究使用 RepeatMasker (v4.1.1)^[20]进行基因组的重复序列注释,指定 rmbblast(v2.13.0+)为搜索引擎,利用 RepBase(版本:20181026)和 Dfam 数据库(版本:3.6)。随后,通过 Tandem Repeats Finder(TRF)(v4.09)^[21]进一步对基因组的串联重复(Tandem Repeats, TRs)序列进行注释。根据 TRs 的长度将其分为两类:重复单元长度大于 6 bp 的为可变数目串联重复(variable number tandem repeat, VNTR),重复单元长度不超过 6 bp 的为短串联重复(short tandem repeat, STR)。

1.4 基因组非同源区域鉴定

本研究参考 Nurk 等^[22]的方法,使用 Winnow-map (v2.03)^[23]软件将 GGswu1 比对至 GRCg7b,生成 sam 文件。接着,使用 paftools^[24]工具进行格式转换,最后使用 bedtools(v2.26.0)^[25]工具提取大于 50 bp 的未比对区域。

1.5 基因组的比对率

本研究共收集了 60 个 WGS 数据和 103 个 RNA-seq 数据,将这些测序 reads 分别比对到 GRCg7b 和 GGswu1 基因组上,以评估不同基因组的比对率。具体步骤如下:首先使用 fastp (v 0.20.0)^[26]对所有 WGS 数据进行质控过滤。接着,使用 BWA-MEM (v 0.7.15)^[27]以默认参数将 reads 分别比对至上述基因组。接着使用 picard-2.18.7.jar (<https://broadinstitute.github.io/picard/>)程序的 MarkDuplicates 工具去除 PCR 重

复。之后,使用 samtools(v1.14)^[28]工具对 reads 比对情况进行统计,提取各样本基于不同基因组的 properly paired reads 比对率进行后续比较分析。对于转录组数据,首先使用 trimmomatic-0.38.jar^[29]对数据进行质控过滤,而后使用 Hisat2(v2.0.3)^[30]将 reads 以默认参数比对至上述基因组,根据 Hisat2 输出的日志文件提取各样本基于不同基因组的 reads overall alignment rate 用于统计分析。

1.6 单核苷酸多态性(SNPs)检测

本研究随机选取了 10 个样本,这些样本的覆盖深度为 10~25×。基于经过 picard 工具去除重复的 bam 文件,使用 GATK(v4.1.8.1)工具^[31]进行 SNPs 检测。根据以下标准过滤低质量的 SNPs:(1)QD(Quality By Depth)<2.0;(2)MQ(RMS Mapping Quality)<30.0。此外,使用 python 包 pysam (<https://github.com/pysam-developers/pysam>)统计每条 reads 的平均比对质量(mapping quality, MAPQ)、比对得分(alignment score, AS)和错配数量(Number of mismatches, NM)。

1.7 转录本组装和比对

本研究收集了 103 个转录组数据,其中 46 个样本来自科尼什鸡、乌当鸡、洛岛红鸡、丝羽乌骨鸡、藏鸡和白来航鸡这 6 个个体的多个组织。使用 SPAdes (v3.14.1)^[32]软件分别对这 6 个个体的转录组数据进行了从头组装。随后,使用 blat (The BLAST-Like Alignment Tool)(v.36)^[33]将组装得到的转录本序列比对至参考基因组,覆盖度大于 95%即认为比对成功。

2 结果与分析

2.1 GRCg7b 与 GGswu1 基因组组装质量的比较

根据表 1 的数据,结果显示 GGswu1 基因组相较于 GRCg7b 基因组增加了约 50 Mb 的序列。同时,GGswu1 基因组的组装间隙数量和长度显著减少,仅剩下约 5 572 bp 的空白序列,主要由 W 染色体上的 14 个间隙组成。GGswu1 基因组共组装出 38 条常染色体、一条 W 染色体和一条 Z 染色体,未定位序列大幅减少,同时 Scaffold N50 和最长 scaffold 长度也有所提升。根据 BUSCO 评估结果,GGswu1 基因组的 BUSCO 完整性得分达到 97.1%,并且片段化和缺失的单拷贝直系同源基因也有所减少,展现出更优异的组装完整性。进一步

的基因组序列注释显示,GGswu1 基因组的 GC 含量和重复序列含量明显增加。特别是在重复序列方面,GGswu1 基因组的重复序列含量比 GRCg7b 增加了 3.75%。详细的重复序列分类显示,GGswu1 基因组中增加的重复序列主要集中在卫星序列(satellites)、可变数目串联重复(VNTR)和短串联重复(STR)等串联重复序列上(图 1A)。

2.2 GRCg7b 与 GGswu1 基因组的非同源区域分析

通过 GGswu1 与 GRCg7b 的比对,分别在 GRCg7b 和 GGswu1 中鉴定到 10.2 Mb 和 70.0 Mb 的非同源序列,并且发现这些非同源序列与基因组重复序列区域的重叠率较高(GGswu1:73.3%,GRCg7b:47.2%)。除了两个基因组测序个体间的遗传差异,这些非同源序列可能与 GRCg7b 和 GGswu1 基因组的组装质量差异有关。值得注意的是,GGswu1 的非同源序列中也包含较多的卫星序列、VNTR 和 STR 等 TRs 序列(图 1B),表明 GGswu1 基因组可能对于这些区域的组装质量有所改善。在 GRCg7b 的非同源区域共注释到 718 个基因。对这些基因进行 Reactome 富集分析显示,它们显著富集在基因表达调控、细胞代谢、免疫、疾病和生命周期等多个通路中(图 2),参与一些重要的基础生物学过程。

表 1 两个参考基因组的组装质量评估
Table 1 Assessment of assembly quality
of two reference genomes

特征	参考基因组	
	GRCg7b	GGswu1
测序技术	CLR/NGS/Hi-C	HiFi/ONT/Hi-C
总长度/Gb	1.05	1.10
总组装间隙数量	463	14
总组装间隙长度/bp	3 383 918	5 572
脚手架数量	213	40
脚手架 N50/Mb	90.86	91.36
最长脚手架长度/Mb	196.45	200.04
完整 BUSCO/%	95.50	97.10
完整且单拷贝 BUSCO/%	95.20	96.80
完整且重复 BUSCO/%	0.30	0.30
片段化 BUSCO/%	1.40	0.70
缺失 BUSCO/%	3.10	2.20
GC 含量/%	42.20	42.64
串联重复序列含量/%	4.00	7.75

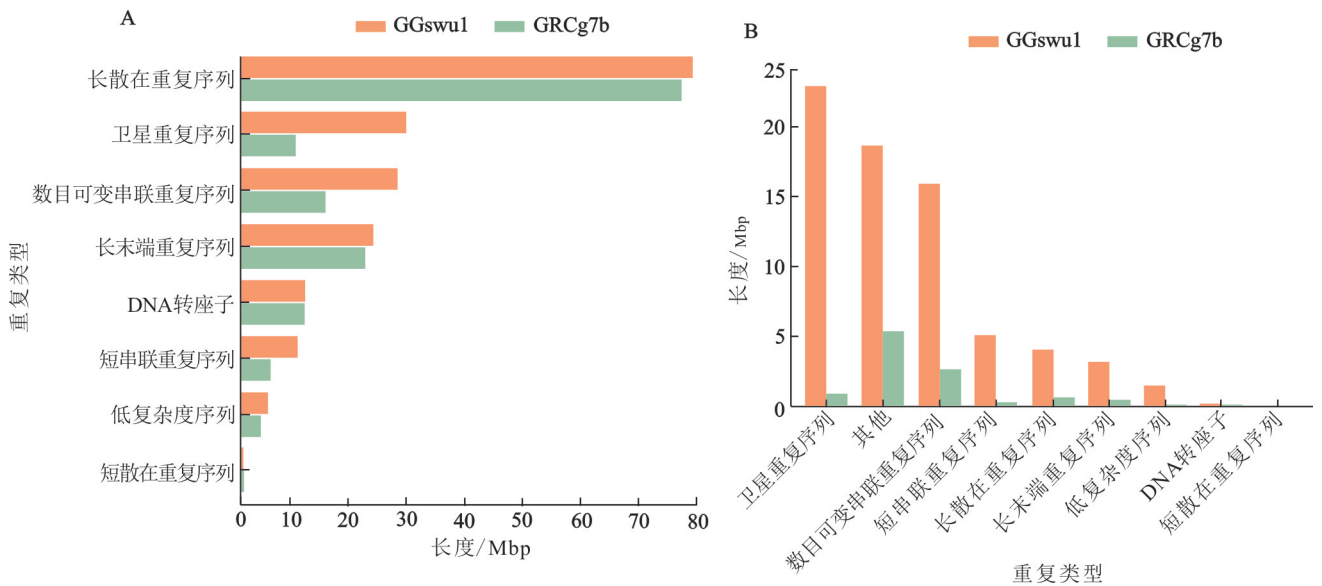


图 1 GRCg7b 和 GGswu1 基因组(A)和基因组间非同源区域(B)的重复序列类型和长度分布
 Fig. 1 The types and length distributions of repetitive sequences in the GRCg7b and GGswu1 genomes (A) and non-homologous regions between the genomes (B)

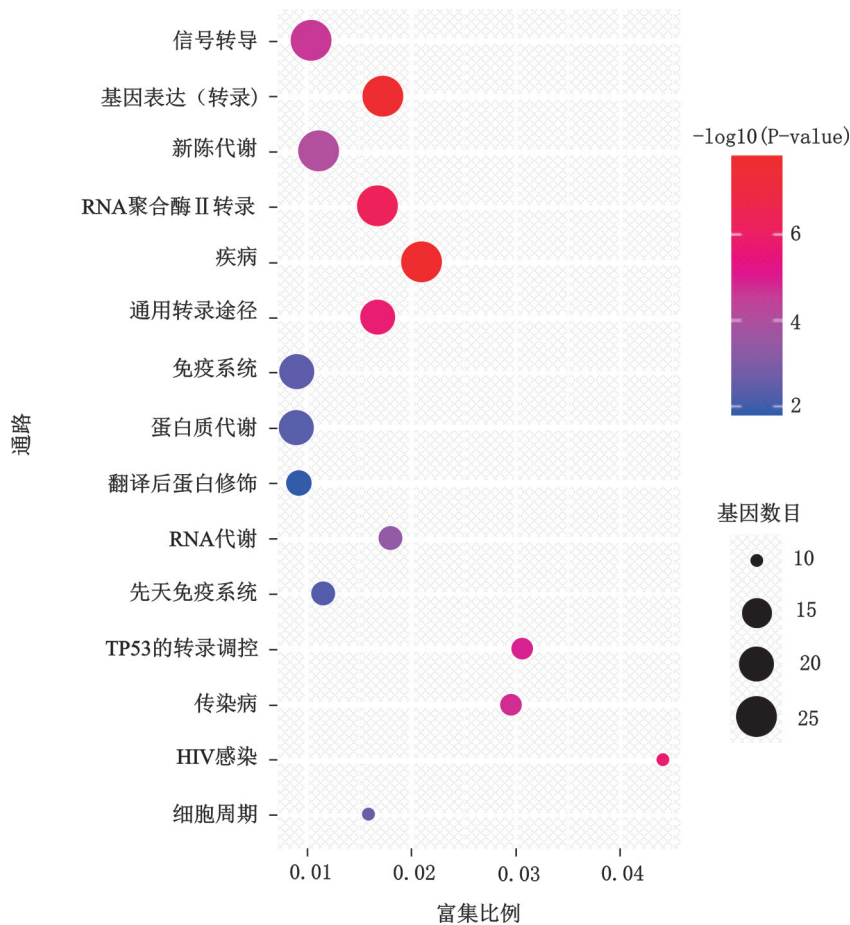


图 2 GRCg7b 非同源区域的基因富集数量最多的前 15 个显著 Reactome 通路

Fig. 2 The top 15 significantly enriched Reactome pathways with the largest number of genes in non-homologous region of GRCg7b

2.3 GRCg7b 与 GGswu1 基因组的比对率分析

本研究分别基于 60 个 WGS 样本和 103 个 RNA-seq 样本对两个参考基因组的比对率进行统计。结果显示,无论是基因组还是转录组数据,

GGswu1 基因组的比对率均显著提升。特别是在转录组数据中,基于 GGswu1 基因组的比对率的中位数相较于 GRCg7b 增加了约 1.7%(图 3)。

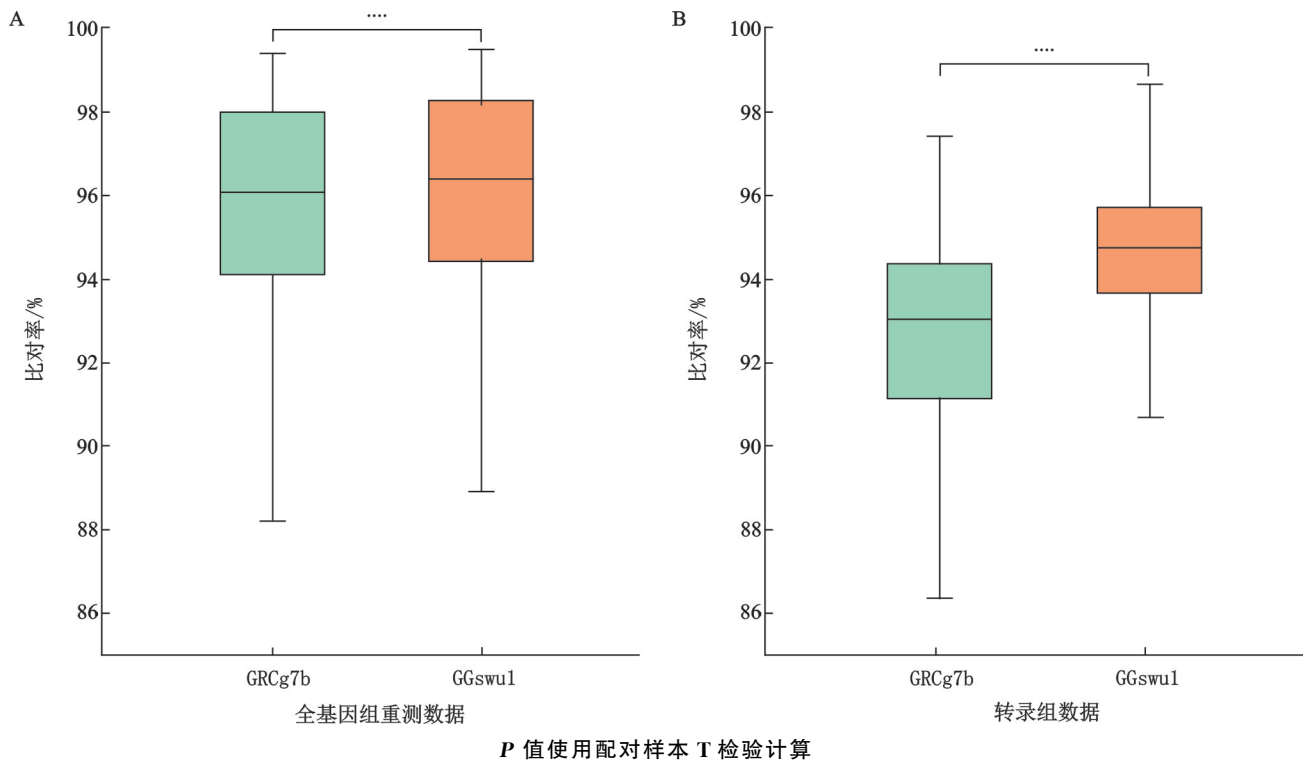


图 3 两个参考基因组的全基因组重测序数据(A)和 RNA 测序数据 (B)的比对率

Fig. 3 Mapping rates of WGS data (A) and RNA-seq data (B) based on two reference genomes

表 2 基于不同参考基因组的 SNPs 检测结果

Table 2 SNPs detection results based on different reference genomes

样本	参考基因组	SNPs 数量	比对质量	比对得分	错配数量
CNGGD_BY01	GRCg7b	5 599 013	56.095 1	93.609 3	1.101 8
CNGGD_BY01	GGswu1	5 608 702	56.212 5	94.287 7	1.067 1
CNGGD_WC01	GRCg7b	6 521 150	56.853 0	94.665 8	0.927 7
CNGGD_WC01	GGswu1	6 456 043	56.964 9	95.191 6	0.897 4
USGGD_WPR01	GRCg7b	5 695 750	55.275 0	95.004 7	0.840 8
USGGD_WPR01	GGswu1	5 927 475	55.191 7	95.696 1	0.809 0
CNGGD_DX01	GRCg7b	6 044 786	56.658 8	94.376 6	0.938 4
CNGGD_DX01	GGswu1	6 011 734	56.766 6	94.919 6	0.909 3
CNGGD_JYSK01	GRCg7b	6 845 568	56.536 5	139.569 7	1.518 8
CNGGD_JYSK01	GGswu1	6 819 383	56.688 7	140.213 5	1.483 1
CNGGD_PXY01	GRCg7b	6 446 316	55.420 8	137.861 5	1.631 2
CNGGD_PXY01	GGswu1	6 555 223	55.642 3	138.662 5	1.619 4
CNGGD_TFBB01	GRCg7b	5 856 953	52.754 3	111.969 6	1.597 0
CNGGD_TFBB01	GGswu1	6 004 676	53.138 4	113.187 3	1.601 6
CAGGD_LOC01	GRCg7b	6 102 972	53.390 0	135.062 9	1.987 0
CAGGD_LOC01	GGswu1	6 413 684	53.720 2	136.331 4	1.984 6
CNGGD_YNLOC01	GRCg7b	7 225 346	54.163 2	93.432 5	0.995 3
CNGGD_YNLOC01	GGswu1	7 264 987	54.415 1	94.420 4	0.959 4
CNGGD_MY01	GRCg7b	6 870 132	54.214 2	116.331 4	1.235 0
CNGGD_MY01	GGswu1	6 944 541	54.599 9	117.554 4	1.219 1

2.4 GRCg7b 与 GGswu1 基因组的 SNPs 检测能力分析

上述结果显示 GGswu1 基因组显著提高了重测序数据的比对率,表明 GGswu1 基因组在变异检测方面具有潜在的改进效果。本研究进一步分析了 GGswu1 基因组在遗传变异检测方面的表现。为减少测序随机性带来的影响,本研究随机选择了 10 个测序深度在 $10\times$ 以上的样本进行 SNPs 检测。结果显示,这 10 个样本在 GGswu1 全基因组范围内共检测到 16 479 839 个 SNPs 位点,比 GRCg7b (16 170 405 个 SNPs)增加了 309 434 个。同时,平均每个样本检出的 SNPs 数量增加了约 30 727 个。此外,除 SNPs 变异位点的增加外,比对至 GGswu1 基因组的每条 reads 的平均比对质量(MAPQ)、比对得分(AS)均有所提升,同时错配数量(NM)有所减少(表 2)。

2.5 GRCg7b 与 GGswu1 基因组的转录本序列比对结果

本研究分别对 6 个个体的 RNA-seq 数据进行了从头组装,每个样本分别得到了 277 833 ~ 409 907 个转录本序列。从表 3 中可以看出,对于不同个体的转录本数据,相较于 GRCg7b 基因组,每个个体比对到 GGswu1 基因组后均有 2 000 ~ 3 000 个更多的转录本可以成功比对。

表 3 基于不同参考基因组的转录本比对结果

Table 3 Transcript alignment results based on different reference genomes

样本	总数	比对到参考基因组的转录本数量	
		GRCg7b	GGswu1
科尼什鸡	357 859	318 215	321 232
乌当鸡	277 833	240 664	243 115
洛岛红鸡	375 892	322 640	325 294
丝羽乌骨鸡	409 907	364 000	367 920
藏鸡	301 858	256 002	258 436
白来航鸡	299 964	250 648	253 106

3 讨论

首个接近完整的鸡基因组序列的公布为我们更好地理解脊椎动物的基因组进化历程和染色体生物学提供了帮助^[10]。作为一个重要的基因组学研究资源,它对全基因组范围内的遗传变异解析和功能基因挖掘也具有重要意义。本研究基于当前鸡的参

考基因组 GRCg7b,对 GGswu1 基因组在组装质量、基因组比对率、变异检测及基因组完整性等方面进行了综合评估和比较。结果显示,GGswu1 基因组在组装连续性和完整性上均有显著提升。得益于高准确性的 PacBio HiFi 和高连续性 ONT Ultra-long 测序技术的应用,为 GGswu1 基因组的完整组装提供了有力的技术支持。先前基于 GRCg6a 的泛基因组研究指出,非参考基因组序列具有高含量 TRs 的特征,而这些序列有可能因测序技术的限制而在组装中容易缺失。在本研究中,GGswu1 基因组不仅显示出更高的 TRs 含量,而且在与 GRCg7b 非同源的区域也观察到了较高比例的 TRs,这表明 GGswu1 在处理基因组的复杂重复区域组装方面取得了实质性的改进。

人类完整基因组序列(T2T-CHM13)公布于 2022 年,Aganezov 等^[34]对其进行了深入评估。他们的研究显示,完整的参考基因组对于改进遗传变异分析至关重要。类似地,本研究以全基因组中分布最广、密度最大的 SNPs 为代表,评估了 GGswu1 的遗传变异解析能力。结果显示,GGswu1 不仅提高了基因组重测序数据的比对率和比对质量,而且能够揭示出更多的变异位点。除 SNPs 检测质量的改进外,T2T-CHM13 的应用还有效改善了以往结构变异(SVs)研究中插入和缺失事件的不平衡问题,使得对全基因组的 SV 有了更全面的了解。同样地,GGswu1 基因组的应用,也可能为鸡的 SV 及其他类型的变异提供更全面、精确的解析。这种全面的遗传信息有助于识别与重要经济性状和生产性能相关的遗传标记,推动家禽遗传改良发展。

GGswu1 基因组在提升转录组数据比对率方面表现出显著的优势,这为更精确地解析可变剪接事件和识别差异表达基因提供了可能。此外,本研究中利用来自 6 个个体的转录本序列,对 GGswu1 基因组组装的完整性进行了初步评估。比对分析结果表明,GGswu1 可能在一定程度上改进了 GRCg7b 中尚未完全组装或有待进一步优化的基因区域。GRCg7b 中的 GGswu1 非同源区域包含较多参与基因表达调控、代谢和免疫等关键生物学过程的基因,这些基因在 GGswu1 的完整组装中可能得到了更好的表征。这有助于更深入地理解基因的调控网络和生物学路径,为揭示鸡的生物学特性提供新的视角。

总之,GGswu1 基因组的完整组装为鸡的基因组学研究提供了一个高分辨率的框架。在 GGswu1 基因组的基础上,结合泛基因组构建、单细胞测序技术和 CRISPR-Cas9 基因编辑等前沿技术,我们可以更深入地解析鸡种质资源的遗传特征^[35]。这种多技术融合的方法将极大地推动基因组学研究和分子育种的发展,加速鸡的遗传改良和新品种的培育。随着这些技术的不断发展和应用,我们期待在鸡的遗传学和育种领域取得更多突破性的进展。

4 结 论

本研究初步阐明了两个关键的鸡基因组之间的差异,表明 GGswu1 基因组相较于当前参考基因组 GRCg7b 在精确性和完整性方面显著提升,并在改进基因组遗传变异解析方面具有较大潜力,可为未来鸡基因组学的深入研究和精准分子育种发展提供一定参考。

参考文献:

- [1] 霍冬敖,王跃斌,陈庆富. 动植物高质量基因组的获取及其主要应用[J]. 生命科学,2019,31(7):739-747.
HUO D A, WANG Y B, CHEN Q F. Acquisition of high quality genomes of animals and plants and their main applications[J]. Chinese Bulletin of Life Sciences, 2019,31(7):739-747.
- [2] 边培培,张禹,姜雨. 泛基因组:高质量参考基因组的新标准[J]. 遗传,2021,43(11):1 023-1 037.
BIAN P P, ZHANG Y, JIANG Y. Pan-genome: Setting a new standard for high-quality reference genomes[J]. Hereditas(Beijing), 2021,43(11):1 023-1 037.
- [3] 梁素芸,周正奎,侯水生. 基于测序技术的畜禽基因组学研究进展[J]. 遗传,2017,39(4):276-292.
LIANG S Y, ZHOU ZH K, HOU SH SH. The research progress of farm animal genomics based on sequencing technologies [J]. Hereditas, 2017, 39 (4): 276-292.
- [4] WU S W, WANG K, DOU T F, et al. High quality assemblies of four indigenous chicken genomes and related functional data resources[J]. Scientific Data, 2024, 11(1):300.
- [5] WONG G K S, LIU B, WANG J, et al. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms [J]. Nature, 2004, 432 (7 018): 717-722.
- [6] 刘振宇. 略阳乌鸡和镇坪乌鸡的群体遗传分析及乌色性状关键候选基因挖掘[D]. 陕西杨凌:西北农林科技大学,2023.
- [7] WANG M S, LI Y, PENG M S, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens[J]. Molecular Biology and Evolution, 2015,32(7):1 880-1 889.
- [8] WANG M S, THAKUR M, PENG M S, et al. 863 genomes reveal the origin and domestication of chicken [J]. Cell Research, 2020,30(8):693-701.
- [9] CONSORTIUM I C G S. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution[J]. Nature, 2004, 432 (7 018):695-716.
- [10] HUANG Z, XU Z X, BAI H, et al. Evolutionary analysis of a complete chicken genome[J]. Proceedings of the National Academy of Sciences of the United States of America, 2023,120(8):e2216641120.
- [11] YI G Q, QU L J, LIU J F, et al. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing[J]. BMC Genomics, 2014,15(1):962.
- [12] LI D Y, CHE T D, CHEN B L, et al. Genomic data for 78 chickens from 14 populations [J]. GigaScience, 2017,6(6):1-5.
- [13] XU N Y, SI W, LI M, et al. Genome-wide scan for selective footprints and genes related to cold tolerance in Chantecler chickens[J]. Zoological Research, 2021, 42 (6):710-720.
- [14] WANG M S, LI Y, PENG M S, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens[J]. Molecular Biology and Evolution, 2015,32(7):1 880-1 889.
- [15] FU W W, WANG R, XU N Y, et al. Galbase: A comprehensive repository for integrating chicken multiomics data[J]. BMC Genomics, 2022,23(1):364.
- [16] LI M, SUN C J, XU N Y, et al. De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions[J]. Molecular Biology and Evolution, 2022,39(4):msac066.
- [17] SHI S R, SHAO D, YANG L Y, et al. Whole genome analyses reveal novel genes associated with chicken ad-

- aptation to tropical and frigid environments[J]. *Journal of Advanced Research*, 2023, 47:13-25.
- [18] ZHANG Y, LU H W, RUAN J. GAEP: A comprehensive genome assembly evaluating pipeline[J]. *Journal of Genetics and Genomics*, 2023, 50(10):747-754.
- [19] SIMÃO F A, WATERHOUSE R M, IOANNIDIS P, et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs [J]. *Bioinformatics*, 2015, 31(19):3 210-3 212.
- [20] TARAİLO-GRAOVAC M, CHEN N S. Using RepeatMasker to identify repetitive elements in genomic sequences[J]. *Current Protocols in Bioinformatics*, 2009, Chapter 4:4. 10. 1-4. 10. 4. 10. 14.
- [21] BENSON G. Tandem repeats finder: A program to analyze DNA sequences [J]. *Nucleic Acids Research*, 1999, 27(2):573-580.
- [22] NURK S, KOREN S, RHIE A, et al. The complete sequence of a human genome[J]. *Science*, 2022:376.
- [23] JAIN C, RHIE A, HANSEN N F, et al. Long-read mapping to repetitive reference sequences using Winnowmap2[J]. *Nature Methods*, 2022, 19:705-710.
- [24] LI H. New strategies to improve minimap2 alignment accuracy [J]. *Bioinformatics*, 2021, 37 (23): 4 572-4 574.
- [25] QUINLAN A R, HALL I M. BEDTools: A flexible suite of utilities for comparing genomic features [J]. *Bioinformatics*, 2010, 26(6):841-842.
- [26] 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.
- [27] LI H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [J]. *arXiv E-Prints*, 2013:1303. 3997v2[q-bio. GN].
- [28] DANECEK P, BONFIELD J K, LIDDLE J, et al. Twelve years of SAMtools and BCFtools [J]. *GigaScience*, 2021, 10(2):giab008.
- [29] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: A flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30(15):2 114-2 120.
- [30] KIM D, PAGGI J M, PARK C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype [J]. *Nature Biotechnology*, 2019, 37:907-915.
- [31] VAN DER AUWERA G A, CARNEIRO M O, HARTL C, et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline [J]. *Current Protocols in Bioinformatics*, 2013, 43(1 110):1-33.
- [32] BUSHMANOVA E, ANTIPOV D, LAPIDUS A, et al. rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data [J]. *GigaScience*, 2019, 8(9):giz100.
- [33] KENT W J. BLAT—the BLAST-like alignment tool [J]. *Genome Research*, 2002, 12(4):656-664.
- [34] AGANEZOV S, YAN S M, SOTO D C, et al. A complete reference genome improves analysis of human genetic variation [J]. *Science*, 2022, 376 (6 588): eabl3533.
- [35] 王宇龙, 李珂, 李栋, 等. 基因组组装技术及其在畜禽遗传变异中的研究进展 [J]. *中国畜牧兽医*, 2023, 50(10):4 085-4 095.
- WANG Y L, LI K, LI D, et al. Genome assembly technology and its research progress on genetic variation of livestock and poultry [J]. *China Animal Husbandry & Veterinary Medicine*, 2023, 50(10):4 085-4 095.