BMC Genomics

DATABASE

Open Access

Yak genome database: a multi-omics analysis platform

Hui Jiang[1,2†], Zhi-Xin Chai[3†], Xiao-Ying Chen[1,2†], Cheng-Fu Zhang[1,2], Yong Zhu[1,2], Qiu-Mei Ji[1,2*] and Jin-Wei Xin[1,2*]

Abstract

Background The yak (*Bos grunniens*) is a large ruminant species that lives in high-altitude regions and exhibits excellent adaptation to the plateau environments. To further understand the genetic characteristics and adaptive mechanisms of yak, we have developed a multi-omics database of yak including genome, transcriptome, proteome, and DNA methylation data.

Description The Yak Genome Database (http://yakgenomics.com/) integrates the research results of genome, transcriptome, proteome, and DNA methylation, and provides an integrated platform for researchers to share and exchange omics data. The database contains 26,518 genes, 62 transcriptomes, 144,309 proteome spectra, and 22,478 methylation sites of yak. The genome module provides access to yak genome sequences, gene annotations and variant information. The transcriptome module offers transcriptome data from various tissues of yak and cattle strains at different developmental stages. The proteome module presents protein profiles from diverse yak organs. Additionally, the DNA methylation module shows the DNA methylation information at each base of the whole genome. Functions of data downloading and browsing, functional gene exploration, and experimental practice were available for the database.

Conclusion This comprehensive database provides a valuable resource for further investigations on development, molecular mechanisms underlying high-altitude adaptation, and molecular breeding of yak.

Keywords Yak, Genome, Database, Multi-omics, Plateau environment

Background

Although single omics study provides information and insights into specific biological or molecular processes, it is hard to confirm the real molecular mechanisms underlying the functionality of an organism and the relationships between biological processes and environmental factors. Integrating and analyzing multiple omics data provide an effective and systematic approach to life science researchers. In general, genomics provides DNA sequence information, transcriptomics examines gene transcription patterns under specific conditions, proteomics explores the composition and expression levels of proteins in cells, and DNA methylation involves chemical modifications on DNA molecules [1]. Multi-omics

[†] Hui Jiang, Zhi-Xin Chai and Xiao-Ying Chen contributed equally to this work.

*Correspondence:
Qiu-Mei Ji
jiqiumei07@163.com
Jin-Wei Xin
xinjinwei18@163.com
[1] State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement, 850000 Lhasa, Tibet, China
[2] Institute of Animal Science and Veterinary, Tibet Academy of Agricultural and Animal Husbandry Sciences, 850000 Lhasa, Tibet, China
[3] Key Laboratory of Qinghai-Tibetan Plateau Animal Genetic Resource Reservation and Utilization, Sichuan Province and Ministry of Education, Southwest Minzu University, 610041 Chengdu, Sichuan, China

analysis combines data at different levels to comprehensively explore biological processes. Multi-omics analysis reveals connections between genomics, transcriptomics, proteomics, and DNA methylation data, facilitating to understand how genomic variations impact gene transcription and protein expression, as well as the associations between DNA methylation and gene activities [2]. These pieces of information contribute novel information to the gene regulatory networks, which are important to molecular mechanisms underlying biological functions, development, metabolism, etiopathology, and environmental adaptation.

The yak (*Bos grunniens*) is a unique species in the Qinghai-Tibet Plateau, and widely distributes in high-altitude areas of Western China and neighboring regions. As a large mammal at the highest-altitude area, yak has survived and adapted to the harsh and cold environment after thousands of years of evolution [3]. Their unique biological features make them an ideal model for studying adaptive evolution and high-altitude ecosystems. Yak also plays important roles in agriculture and economic development. As a significant livestock species, yak provides meat, fur, and other economic resources. Their dung is also an important source of agricultural fertilizer and energy production. Moreover, yak positively impacts the ecological balance and vegetation restoration in the plateau grasslands through their grazing behaviors [4]. In recent years, we have analyzed yak using different omics approaches. These data preliminarily explored the yak genetic characteristics, gene transcription, protein expression, and DNA methylation patterns, as well as molecular regulatory mechanisms in response to different conditions [5–11], providing novel insights into the mechanisms underlying evolution, and high-altitude adaptation in yak.

Currently, the data resources of yak omics researches are generally stored in public databases in their raw data format, such as NCBI. These databases primarily provide storage and retrieval functions, but lack an integrated platform for data integration and in-depth analysis. Hu et al. [12] developed a yak genome database (http://me.lzu.edu.cn/yak), which incorporated genome sequences, predicted genes and associated annotations, non-coding RNA sequences, transposable elements, and single nucleotide variants of yak, as well as three-way whole-genome alignments between human, cattle and yak. However, this database did not include other omics datasets, such as transcriptome, proteome, and DNA methylation. Given the vast and diverse nature of omics data, the traditional database retrieval methods could not fully explore the relationship between different types of datasets [13]. Thus, an integrated platform of different omics data is crucial to facilitate data integration, interaction, and analysis. An integrated platform can also offer advanced data mining and machine learning algorithms to help researchers discover the complex relationships among yak genomics, transcriptomics, proteomics, and other omics levels, further deepening our understanding of biological processes and diseases in yak.

In this study, the Yak Genome Database (http://yak-genomics.com/) was constructed, which successfully assembled a comprehensive yak fine-scale genome map at the chromosome level, using PacBio sequencing, Illumina sequencing, Bionano assembly, and Hi-C three-dimensional genome scaffolding. Moreover, this platform also integrated transcriptome, proteome, and DNA methylation data of yak, which were not available in Yak Genome Database developed by Hu et al. [12]. This database provides basic information for yak researches in future, such as molecular breeding, molecular evolution, disease prevention and control.

Construction and content

The Yak Genome Database was deployed in the Ubuntu 20.04 operation system using the AKKA 2.13 (web server), MySQL 8.0.30 (database server), Scala 2.13.2, and SBT 1.3.9. All data were managed and stored using the MySQL Database Management System. The query function was enforced based on Slick 3.3.2 middleware tier. The Jbrowse 1.16.11 was used to visualize the genome. The website interfaces were designed and implemented using the Bootstrap 4.6.0 and the Play Framework 2.8.7. The software versions and statistical tools used for data analyses and plot preparation have been presented in Xin et al. [6–11]. The boxplots, and heatmaps were prepared using R 4.2.1. The website has been tested in several popular web browsers, including Firefox, Google Chrome, and Internet Explorer.

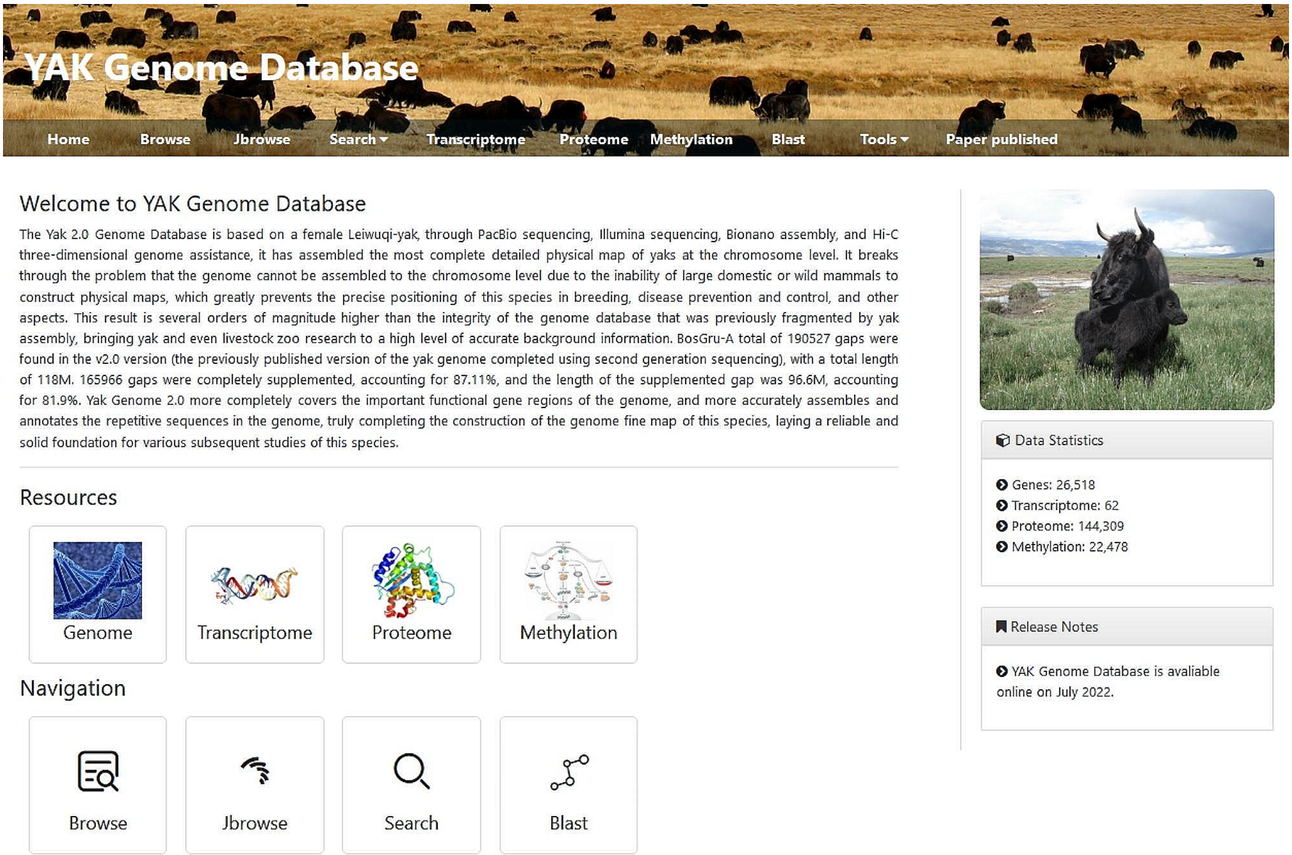
Utility and discussion

The yak genome database content

The multi-omics data in the Yak Genome Database are categorized into two central functional domains: data resources and navigation (Fig. 1). The data resources contain four main modules, including genome, transcriptome, proteome, and methylation information. The database contains 26,518 genes, 62 transcriptomes, 144,309 proteome spectra, and 22,478 methylation sites of yak. The navigation page consists of Browser, Jbrowse, Search and Blast functions. Currently, the database supports individual download of images and gene data. In the future, we will add functions such as one-click download of whole genome information.

Genome module

The Genome module incorporates the complete genomic DNA sequence of yaks obtained by the third-generation high-throughput sequencing platform (PacBio RSII) [14].



Welcome to YAK Genome Database

The Yak 2.0 Genome Database is based on a female Leiwuqi-yak, through PacBio sequencing, Illumina sequencing, Bionano assembly, and Hi-C three-dimensional genome assistance, it has assembled the most complete detailed physical map of yaks at the chromosome level. It breaks through the problem that the genome cannot be assembled to the chromosome level due to the inability of large domestic or wild mammals to construct physical maps, which greatly prevents the precise positioning of this species in breeding, disease prevention and control, and other aspects. This result is several orders of magnitude higher than the integrity of the genome database that was previously fragmented by yak assembly, bringing yak and even livestock zoo research to a high level of accurate background information. BosGru-A total of 190527 gaps were found in the v2.0 version (the previously published version of the yak genome completed using second generation sequencing), with a total length of 118M. 165966 gaps were completely supplemented, accounting for 87.11%, and the length of the supplemented gap was 96.6M, accounting for 81.9%. Yak Genome 2.0 more completely covers the important functional gene regions of the genome, and more accurately assembles and annotates the repetitive sequences in the genome, truly completing the construction of the genome fine map of this species, laying a reliable and solid foundation for various subsequent studies of this species.

Resources

Genome
Transcriptome
Proteome
Methylation

Navigation

Browse
Jbrowse
Search
Blast

Data Statistics

- Genes: 26,518
- Transcriptome: 62
- Proteome: 144,309
- Methylation: 22,478

Release Notes

- YAK Genome Database is avaliable online on July 2022.

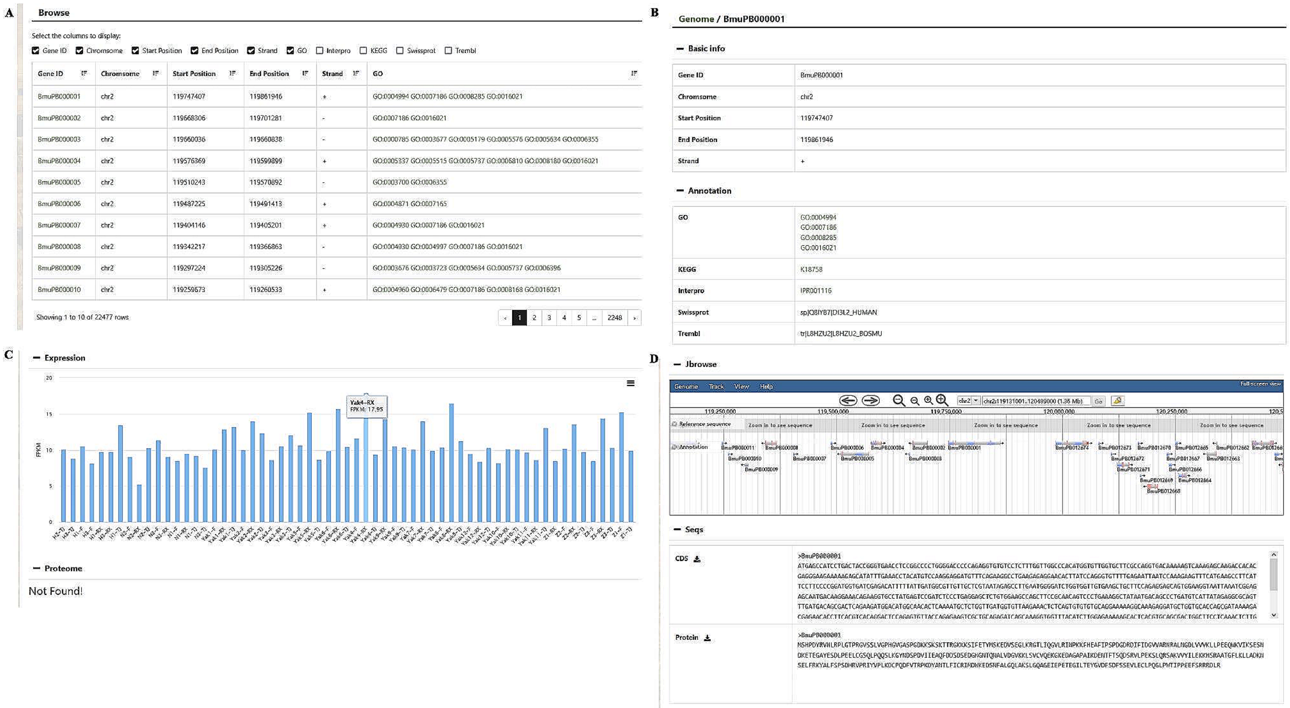
Fig. 1 The homepage of yak genome database

The yak genome was sequenced at a coverage of 70X, with the second-generation sequencing data used to correct errors. The Bionano assisted assembly technology was used for high-quality assembly, and analysis. Next, a refined physical map of the yak chromosome was generated, providing a more readable and complete genome database than the fragmented information in another Yak Genome Database (BosGru_v2.0) [15], and contributing a novel genome tool to yak researchers.

When accessing the 'Genome' section on the homepage, a new page will display information of genes at all locations, such as Gene ID, Chromosome, Start Position, End Position, Strand, GO (Gene Ontology) terms, Interpro, KEGG (Kyoto Encyclopedia of genes and Genomes), Swissprot, and Trembl in a user-friendly table format (Fig. 2A). When clicking each gene, users can access detailed information of this gene, including annotations, transcriptional levels, proteome data, Jbrowse page, and nucleotide sequences associated with the gene (Fig. 2B-2D). The 'Annotation' tav provides comprehensive gene annotation information, including GO terms, KEGG pathways, and Interpro annotations, which can be further explored by clicking them. The 'Expression' tab displays gene expression levels across different cattle breeds and tissues, and users can download the images in various formats by selecting the menu in the upper right corner of the image. 'Jbrowse' is used to display integrated information from annotated genomic datasets, while 'Seqs' provides the coding sequence (CDS) and protein sequence on the selected gene.

Transcriptome module

Previously, comparative transcriptome sequencing was performed on lung, gluteal muscle, and mammary gland tissues of low-altitude cattle (Sanjiang and Holstein cattle), Tibetan cattle (living at a moderate altitude), and yaks (living at a high altitude). In addition, these tissues of yaks at different ages (6, 30, 60, and 90 months) were also subjected to transcriptome sequencing. These analyses identified the functional genes involved in the major biochemical, metabolic, and signal transduction pathways involved in yak development and high-altitude adaptation [10, 11]. These data are included in the transcriptome module on the website, providing a valuable transcriptome database for specific tissue biomarkers, molecular research, and breeding of yaks. After clicking the "Transcriptome" button, users can select the strain in the 'Sample' dialog box, enter the gene ID in the 'Gene ID' dialog box, and then click 'Search' (Fig. 3A), and then the website will return the transcriptional levels of the



Browse

Select the columns to display:

☑ Gene ID ☑ Chromsome ☑ Start Position ☑ End Position ☑ Strand ☑ GO ☐ Interpro ☐ KEGG ☐ Swissprot ☐ Trembl

Gene ID	Chromsome	Start Position	End Position	Strand	GO
BmuPB000001	chr2	119747407	119861946	+	GO:0004994 GO:0007186 GO:0008285 GO:0016021
BmuPB000002	chr2	119668306	119701281	-	GO:0007186 GO:0016021
BmuPB000003	chr2	119660036	119660838	-	GO:0000785 GO:0003677 GO:0005179 GO:0005576 GO:0005634 GO:0006355
BmuPB000004	chr2	119576369	119599899	+	GO:0005337 GO:0005515 GO:0005737 GO:0006810 GO:0008180 GO:0016021
BmuPB000005	chr2	119510243	119570892	-	GO:0003700 GO:0006355
BmuPB000006	chr2	119487225	119491413	+	GO:0004871 GO:0007165
BmuPB000007	chr2	119404146	119405201	+	GO:0004930 GO:0007186 GO:0016021
BmuPB000008	chr2	119342217	119366863	-	GO:0004930 GO:0004997 GO:0007186 GO:0016021
BmuPB000009	chr2	119297224	119305226	-	GO:0003676 GO:0003723 GO:0005634 GO:0005737 GO:0006396
BmuPB000010	chr2	119259673	119260533	+	GO:0004960 GO:0006479 GO:0007186 GO:0008168 GO:0016021

Showing 1 to 10 of 22477 rows ‹ 1 2 3 4 5 ... 2248 ›

Genome / BmuPB000001

Basic info

Gene ID	BmuPB000001
Chromsome	chr2
Start Position	119747407
End Position	119861946
Strand	+

Annotation

GO	GO:0004994 GO:0007186 GO:0008285 GO:0016021
KEGG	K18758
Interpro	IPR001116
Swissprot	sp|Q8IYB7|DI3L2_HUMAN
Trembl	tr|L8HZU2|L8HZU2_BOSMU

Proteome

Not Found!

Seqs

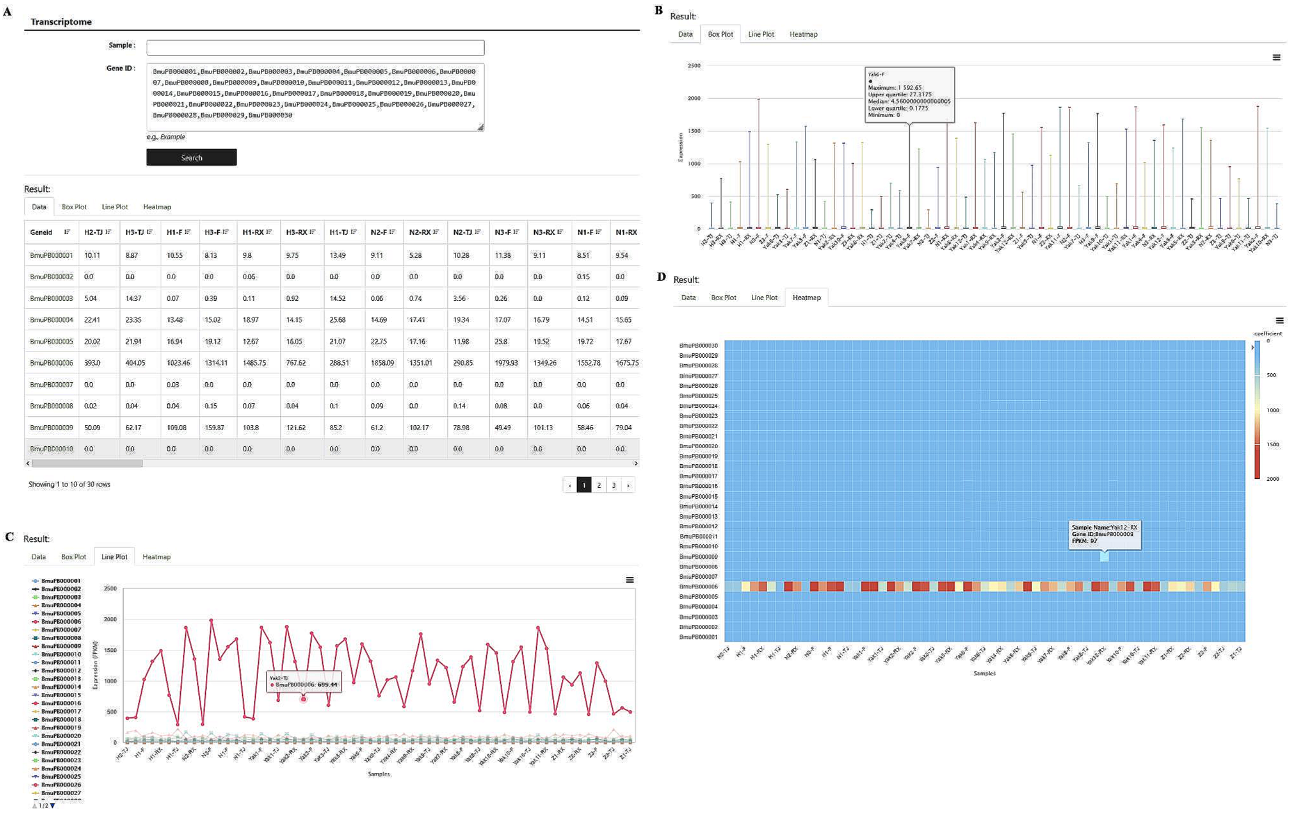
CDS

```
>BmuPB000001
ATGAGCCATCCTGACTACCGGGTGAACCTCCGGCCCCTGGGGACCCCCAGAGGTGTGTCCTCTTTGGTTGGCCCACATGGTGTTGGTGCTTCGCCAGGTGACAAAAAGTCAAAGAGCAAGACCACAC
GAGGGAAGAAAAAGAGCATATTTGAAACCTACATGTCCAAGGAGGATGTTTCAGAAGGCCTGAAGAGAGGAACACTTATCCAGGGTGTTTTGAGAATTAATCCAAAGAAGTTTCATGAAGCCTTCAT
TCCTTCCCCGGATGGTGATCGAGACATTTTTATTGATGGCGTTGTTGCTCGTAATAGAGCCTTGAATGGGGATCTGGTGGTTGTGAAGCTGCTTCCAGAGGAGCAGTGGAAGGTAATTAAATCGGAG
AGCAATGACAAGGAAACAGAAGGTGCCTATGAGTCCGATCTCCCTGAGGAGCTCTGTGGAAGCCAGCTTCCGCAACAGTCCCTGAAAGGCTATAATGACAGCCCTGATGTCATTATAGAGGCGCAGT
TTGATGACAGCGACTCAGAAGATGGACATGGCAACACTCAAAATGCTCTGGTTGATGGTGTTAAGAAACTCTCAGTGTGTGTGCAGGAAAAAGGCAAAGAGGATGCTGGTGCACCAGCGATAAAAGA
CGAGAACACCTTCACGTCACAGGACTCCAGAGTGTTACCAGAGAAGTCGCTGCAGAGATCAGCAAAGGTGGTTTACATCTTGGAGAAAAAGCACTCACGTGCAGCGACTGGCTTCCTCAAACTCTTG
```

Protein

```
>BmuPB000001
MSHPDYRVNLRPLGTPRGVSSLVGPHGVGASPGDKKSKSKTTRGKKKSIFETYMSKEDVSEGLKRGTLIQGVLRINPKKFHEAFIPSPDGDRDIFIDGVVARNRALNGDLVVVKLLPEEQWKVIKSESN
DKETEGAYESDLPEELCGSQLPQQSLKGYNDSPDVIIEAQFDDSDSEDGHGNTQNALVDGVKKLSVCVQEKGKEDAGAPAIKDENTFTSQDSRVLPEKSLQRSAKVVYILEKKHSRAATGFLKLLADKN
SELFRKYALFSPSDHRVPRIYVPLKDCPQDFVTRPKDYANTLFICRIANDWKEDSNFALGQLAKSLGQAGEIEPETEGILTEYGVDFSDFSSEVLECLPQGLPWTIPPEEFSRRRDLR
```

Fig. 2 Features of the genome module. (**A**) Genome browse. (**B**) Basic information and annotation of a gene. (**C**) Gene expression. (**D**) Gene Jbrowse and sequences



Transcriptome

Sample :

Gene ID : BmuPB000001,BmuPB000002,BmuPB000003,BmuPB000004,BmuPB000005,BmuPB000006,BmuPB000007,BmuPB000008,BmuPB000009,BmuPB000010,BmuPB000011,BmuPB000012,BmuPB000013,BmuPB000014,BmuPB000015,BmuPB000016,BmuPB000017,BmuPB000018,BmuPB000019,BmuPB000020,BmuPB000021,BmuPB000022,BmuPB000023,BmuPB000024,BmuPB000025,BmuPB000026,BmuPB000027,BmuPB000028,BmuPB000029,BmuPB000030

e.g., *Example*

Search

Result:

Data Box Plot Line Plot Heatmap

Geneid	H2-TJ	H3-TJ	H1-F	H3-F	H1-RX	H3-RX	H1-TJ	N2-F	N2-RX	N2-TJ	N3-F	N3-RX	N1-F	N1-RX
BmuPB000001	10.11	8.87	10.55	8.13	9.8	9.75	13.49	9.11	5.28	10.28	11.38	9.11	8.51	9.54
BmuPB000002	0.0	0.0	0.0	0.0	0.06	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.15	0.0
BmuPB000003	5.04	14.37	0.07	0.39	0.11	0.92	14.52	0.06	0.74	3.56	0.26	0.0	0.12	0.09
BmuPB000004	22.41	23.35	13.48	15.02	18.97	14.15	25.68	14.69	17.41	19.34	17.07	16.79	14.51	15.65
BmuPB000005	20.02	21.94	16.94	19.12	12.67	16.05	21.07	22.75	17.16	11.98	25.8	19.52	19.72	17.67
BmuPB000006	393.0	404.05	1023.46	1314.11	1485.75	767.62	288.51	1858.09	1351.01	290.85	1979.93	1349.26	1552.78	1675.75
BmuPB000007	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BmuPB000008	0.02	0.04	0.04	0.15	0.07	0.04	0.1	0.09	0.0	0.14	0.08	0.0	0.06	0.04
BmuPB000009	50.09	62.17	109.08	159.87	103.8	121.62	85.2	61.2	102.17	78.98	49.49	101.13	58.46	79.04
BmuPB000010	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Showing 1 to 10 of 30 rows ‹ 1 2 3 ›

Fig. 3 Features of the transcriptome module. (**A**) Transcriptome browse. (**B**) Box plot, (**C**) Line plot and (**D**) Heatmap of gene expression

selected genes in selected samples in the forms of data table, Boxplot, Lineplot, and Heatmap (Fig. 3B and D).

Proteome module

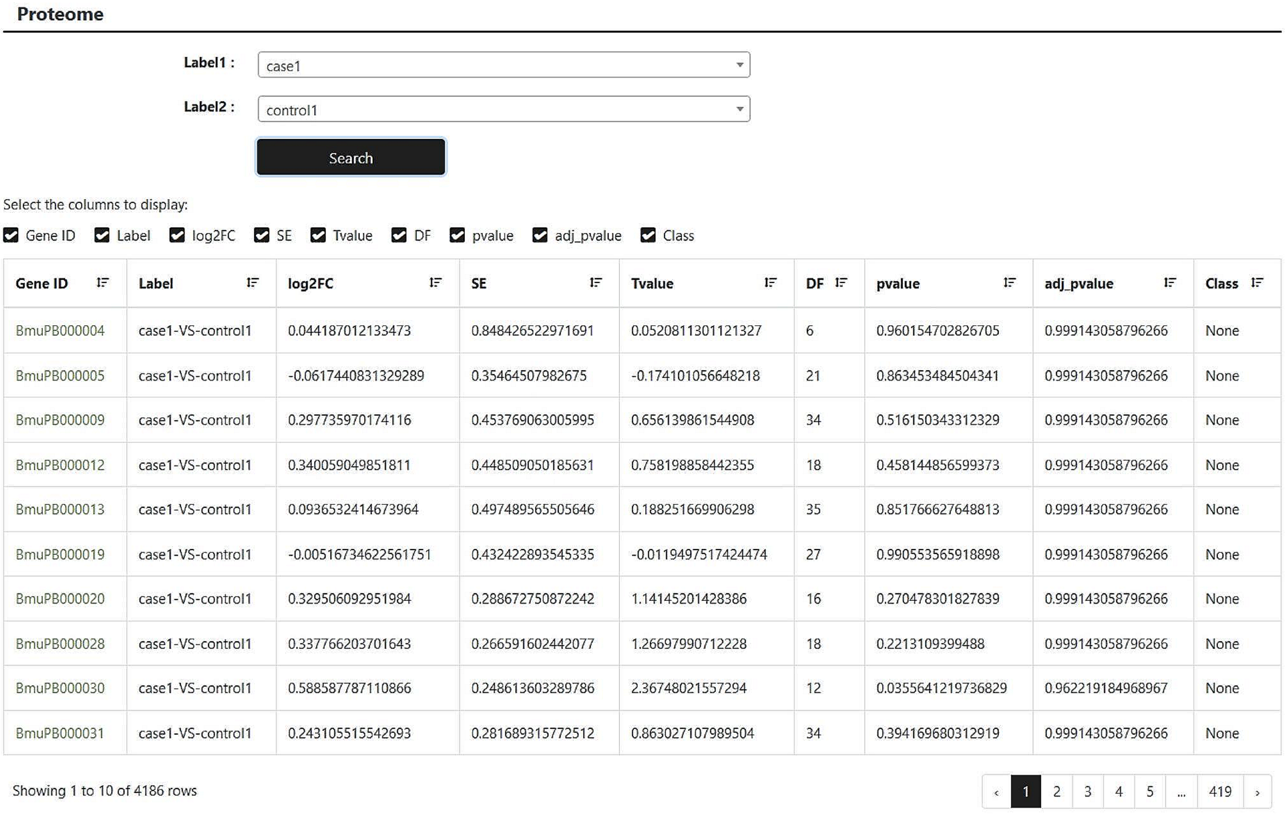
Using the liquid chromatography-mass spectrometry (LC-MS) method, proteomic analyses were conducted for four specific tissues from four different species (yak, Tibetan cattle, Sanjiang cattle, and Holstein cattle) [7–9]. All the animals were female and 60 months of age. The proteome module provides two input dialog boxes. Users can select two samples and then click the "search" button. Next, the website will return the comparison results of the expression levels of all genes in the two selected samples, including log2(fold change) and statistical parameters (Fig. 4).

Methylation module

DNA methylation is a critical epigenetic modification that occurs in both animals and plants, playing pivotal roles in chromosome structure, gene expression and regulation [16]. The establishment of a comprehensive DNA methylation database for yak can significantly advance the comprehension of cellular gene expression and regulation, and provide deeper insights into the spatiotemporal specificity of DNA methylation across various developmental stages and organs [17]. The DNA methylation database of yak presents single-base methylation maps and tissue-specific methylation maps. The single-base methylation maps include: 1) DNA methylation levels at the single-base resolution, 2) DNA methylation levels specific to different base types, 3) DNA methylation levels specific to different gene structures, 4) DNA methylation levels in repetitive sequences, and 5) DNA methylation levels in non-coding sequences and regulatory regions. The tissue-specific methylation maps involve three tissues: mammary gland, lung, and muscle [6]. On the website, users can select 'Sample' and 'Chromosome' in the Methylation module, set the 'Start Position' and 'End Position,' and finally click 'Search' to obtain the corresponding DNA methylation results on the selected sequences (Fig. 5).

Navigation

'Browse' allows users to read the yak genome directly. 'JBrowse' is a next-generation genome browser built with JavaScript and HTML5. The Jbrowse of Yak Genome Database includes tracks describing gene, gene sequence, mRNAs, structure, and other gene-related features, and provides a graphical display of annotations on the yak genome (Fig. 6). Users can browse gene models on chromosomes and unanchored contigs. For example, if user set the genomic region from 4,454,001 bp to 5,878,000 bp



Proteome

Label1 : case1

Label2 : control1

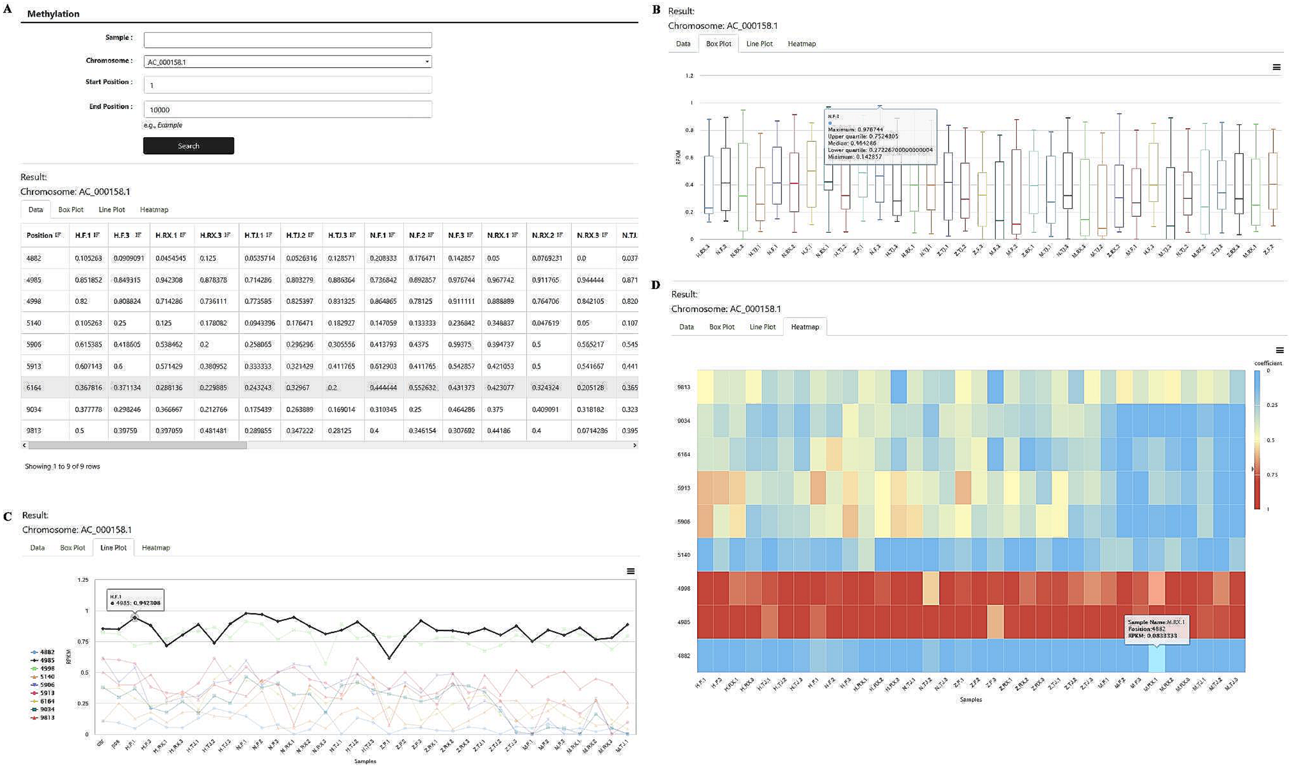
Select the columns to display:

☑ Gene ID ☑ Label ☑ log2FC ☑ SE ☑ Tvalue ☑ DF ☑ pvalue ☑ adj_pvalue ☑ Class

Gene ID	Label	log2FC	SE	Tvalue	DF	pvalue	adj_pvalue	Class
BmuPB000004	case1-VS-control1	0.044187012133473	0.848426522971691	0.0520811301121327	6	0.960154702826705	0.999143058796266	None
BmuPB000005	case1-VS-control1	-0.0617440831329289	0.35464507982675	-0.174101056648218	21	0.863453484504341	0.999143058796266	None
BmuPB000009	case1-VS-control1	0.297735970174116	0.453769063005995	0.656139861544908	34	0.516150343312329	0.999143058796266	None
BmuPB000012	case1-VS-control1	0.340059049851811	0.448509050185631	0.758198858442355	18	0.458144856599373	0.999143058796266	None
BmuPB000013	case1-VS-control1	0.0936532414673964	0.497489565505646	0.188251669906298	35	0.851766627648813	0.999143058796266	None
BmuPB000019	case1-VS-control1	-0.00516734622561751	0.432422893545335	-0.0119497517424474	27	0.990553565918898	0.999143058796266	None
BmuPB000020	case1-VS-control1	0.329506092951984	0.288672750872242	1.14145201428386	16	0.270478301827839	0.999143058796266	None
BmuPB000028	case1-VS-control1	0.337766203701643	0.266591602442077	1.26697990712228	18	0.2213109399488	0.999143058796266	None
BmuPB000030	case1-VS-control1	0.588587787110866	0.248613603289786	2.36748021557294	12	0.0355641219736829	0.962219184968967	None
BmuPB000031	case1-VS-control1	0.243105515542693	0.281689315772512	0.863027107989504	34	0.394169680312919	0.999143058796266	None

Showing 1 to 10 of 4186 rows

Fig. 4 Browse of the proteome module



Methylation

Sample :

Chromosome : AC_000158.1

Start Position : 1

End Position : 10000

e.g., *Example*

Result:

Chromosome: AC_000158.1

Position	H.F.1	H.F.3	H.RX.1	H.RX.3	H.TJ.1	H.TJ.2	H.TJ.3	N.F.1	N.F.2	N.F.3	N.RX.1	N.RX.2	N.RX.3	N.TJ.
4882	0.105263	0.0909091	0.0454545	0.125	0.0535714	0.0526316	0.128571	0.208333	0.176471	0.142857	0.05	0.0769231	0.0	0.037
4985	0.851852	0.849315	0.942308	0.878378	0.714286	0.803279	0.886364	0.736842	0.892857	0.976744	0.967742	0.911765	0.944444	0.871
4998	0.82	0.808824	0.714286	0.736111	0.773585	0.825397	0.831325	0.864865	0.78125	0.911111	0.888889	0.764706	0.842105	0.820
5140	0.105263	0.25	0.125	0.178082	0.0943396	0.176471	0.182927	0.147059	0.133333	0.236842	0.348837	0.047619	0.05	0.107
5906	0.615385	0.418605	0.538462	0.2	0.258065	0.296296	0.305556	0.413793	0.4375	0.59375	0.394737	0.5	0.565217	0.545
5913	0.607143	0.6	0.571429	0.380952	0.333333	0.321429	0.411765	0.612903	0.411765	0.542857	0.421053	0.5	0.541667	0.441
6164	0.367816	0.371134	0.288136	0.229885	0.243243	0.32967	0.2	0.444444	0.552632	0.431373	0.423077	0.324324	0.205128	0.365
9034	0.377778	0.298246	0.366667	0.212766	0.175439	0.263889	0.169014	0.310345	0.25	0.464286	0.375	0.409091	0.318182	0.323
9813	0.5	0.39759	0.397059	0.481481	0.289855	0.347222	0.28125	0.4	0.346154	0.307692	0.44186	0.4	0.0714286	0.395

Showing 1 to 9 of 9 rows

N.F.3
Maximum: 0.976744
Upper quartile: 0.7524305
Median: 0.464286
Lower quartile: 0.27226700000000004
Minimum: 0.142857

H.F.1
● 4985: 0.942308

4882
4985
4998
5140
5906
5913
6164
9034
9813

Sample Name:M.RX.1
Position:4882
RPKM: 0.0833333

Fig. 5 Features of the methylation module. (**B**) Box plot, (**C**) Line plot and (**D**) Heatmap of methylated gene expression

on Chr1 for browsing, all genes in this region will appear in order (Fig. 6A). When clicking on 'BmuPB021145', an extra layer will appear with the detailed information, such as mRNAs, CDS and other features (Fig. 6B). For more operational details, users can click the 'Help' button, which provides comprehensive instructions and guidance.

The 'Search' tab supplies users with two methods (search by gene ID or range) for genome searching. When users click on 'Blast', three options 'Blastn Gene', 'Blastn Genome' and 'Blastp' will display. Users can select the Blast type and enter a DNA or protein sequence, and set the parameters of 'Evalue', 'Word size' and 'Max target seqs'. After clicking the 'Search' button, the nucleotide or protein sequence complying the search conditions will display and could be downloaded by the users.

Additional tools

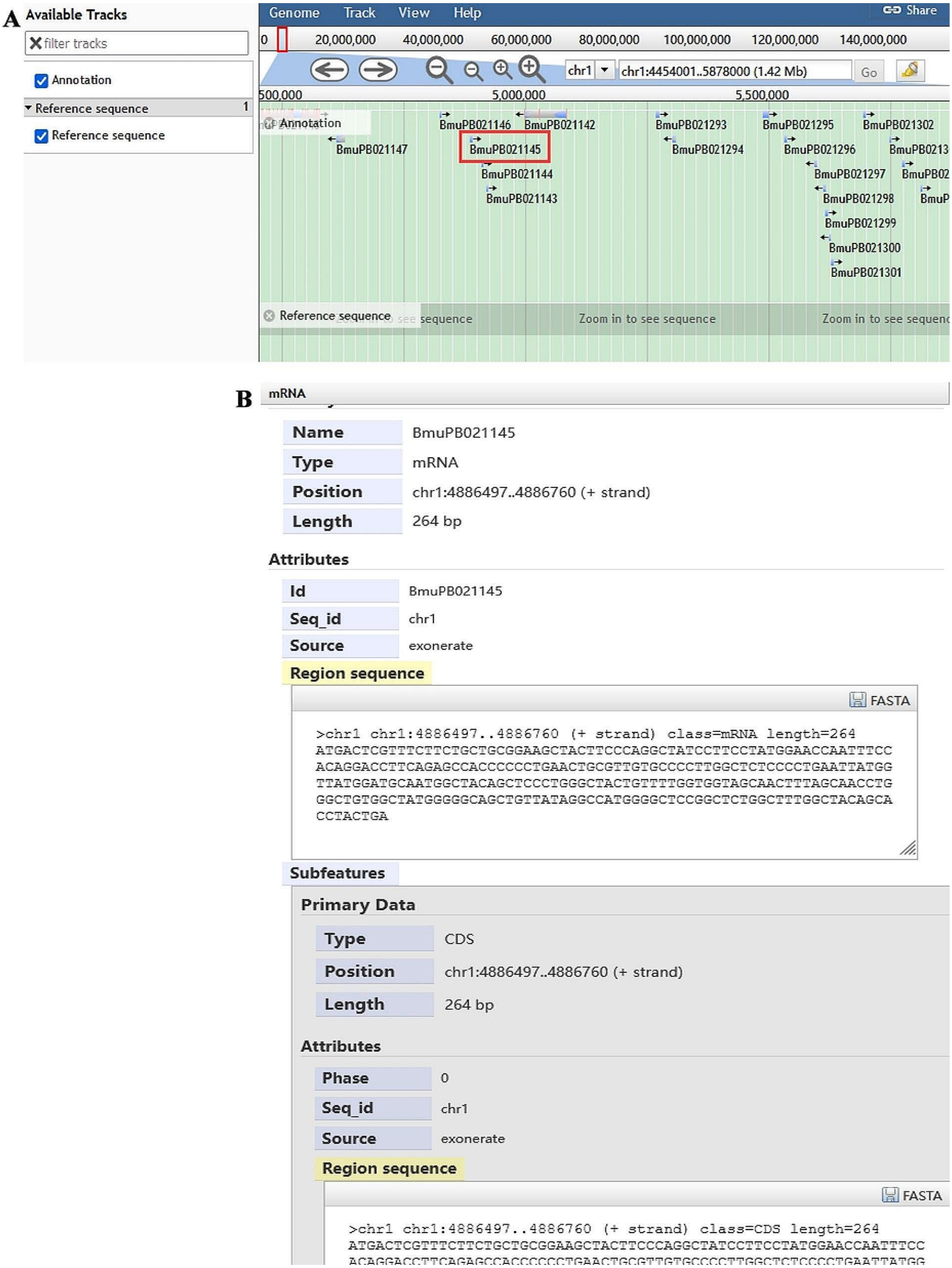
The Yak Genome Database also provides users with several convenient online tools, including Primer designer, GO and KEGG enrichment. The 'Primer designer' tool offers primer design function to amplify a selected sequence. The 'GO enrichment' and 'KEGG enrichment' tools facilitate the users to obtain the GO and KEGG enrichment results of a set of genes.

Maintenance of the yak genome database in future

To ensure continuous operation of the Yak Genome Database, we would assign an administrator to manage the website regularly. We would keep omics studies on yak in future, and all the omics data we obtained would be uploaded to this database. In addition, we would keep cooperations with other investigators and find more cooperators who work on yak. Next, all the progresses on yak omics would also be encouraged to supplement in this database.

Conclusions

The Yak Genome Database is a comprehensive platform of genomic physical map, which integrates genome, transcriptome, proteome, and DNA methylation data. Information in the database can be downloaded, and shared through the Internet. Users who want to upload their own data can contact the administrator of the website. By providing timely updates on yak research progress, the Yak Genome Database enables efficient and interactive sharing of existing scientific data among researchers worldwide who are interested in yak, cattle, livestock, ruminant animals, and even medical research. Comparative analysis of multidimensional data from key yak tissues aims to uncover the mechanisms underlying high-altitude adaptation, disease resistance, cold tolerance, and starvation resistance of large animals in the plateau. These findings contribute to molecular breeding



Name	BmuPB021145
Type	mRNA
Position	chr1:4886497..4886760 (+ strand)
Length	264 bp

Attributes

Id	BmuPB021145
Seq_id	chr1
Source	exonerate

Region sequence

FASTA

```
>chr1 chr1:4886497..4886760 (+ strand) class=mRNA length=264
ATGACTCGTTTCTTCTGCTGCGGAAGCTACTTCCCAGGCTATCCTTCCTATGGAACCAATTTCC
ACAGGACCTTCAGAGCCACCCCCCTGAACTGCGTTGTGCCCCTTGGCTCTCCCCTGAATTATGG
TTATGGATGCAATGGCTACAGCTCCCTGGGCTACTGTTTTGGTGGTAGCAACTTTAGCAACCTG
GGCTGTGGCTATGGGGGCAGCTGTTATAGGCCATGGGGCTCCGGCTCTGGCTTTGGCTACAGCA
CCTACTGA
```

Subfeatures

Primary Data

Type	CDS
Position	chr1:4886497..4886760 (+ strand)
Length	264 bp

Attributes

Phase	0
Seq_id	chr1
Source	exonerate

Region sequence

FASTA

```
>chr1 chr1:4886497..4886760 (+ strand) class=CDS length=264
ATGACTCGTTTCTTCTGCTGCGGAAGCTACTTCCCAGGCTATCCTTCCTATGGAACCAATTTCC
ACAGGACCTTCAGAGCCACCCCCCTGAACTGCGTTGTGCCCCTTGGCTCTCCCCTGAATTATGG
```

Fig. 6 Regional view of the genome using Jbrowse. (**A**) A graphic view of the region 4,454,001 bp to 5,878,000 bp on Chr1. (**B**) The interface after clicking on 'BmuPB021145'.

of livestock animals and the understanding of human responses to harsh environments.

Abbreviations

CDS	Coding Sequence
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center for Biotechnology Information

Author contributions

HJ and ZXC performed the analysis. XYC conducted the database. CFZ and YZ wrote the paper. QMJ supervised the database and JWX revised the manuscript. All authors have approved the final article.

Funding

This work was supported by the Program of Provincial Department of Finance of the Tibet Autonomous Region, the Major Special Projects of Tibet Autonomous Region (XZ202101ZD0002N-01), the Second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK0501), and the program National Beef Cattle and Yak Industrial Technology System (CARS-37).

Data availability

The datasets generated and analyzed in the current study are freely available on the Download page of Yak database with the web link: http://yakgenomics.com/.

Declarations

Ethics approval and consent to participate

All procedures and experiments involving animals followed the guidelines for the Care and Use of Laboratory Animals. The Ethics Committee at Institute of Animal Science and Veterinary, Tibet Academy of Agricultural and Animal Husbandry Sciences (Permit Number: 2015 – 216) approved this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 30 October 2023 / Accepted: 31 March 2024
Published online: 05 April 2024

References

1. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform. 2018;19(2):286–302.
2. Liao Y, Wang J, Zou J, Liu Y, Liu Z, Huang Z. Multi-omics analysis reveals genomic, clinical and immunological features of SARS-CoV-2 virus target genes in pan-cancer. Front Immunol. 2023;14:1112704.
3. Ge Q, Guo Y, Zheng W, Zhao S, Cai Y, Qi X. Molecular mechanisms detected in yak lung tissue via transcriptome-wide analysis provide insights into adaptation to high altitudes. Sci Rep. 2021;11(1):7786.
4. Ayalew W, Chu M, Liang C, Wu X, Yan P. Adaptation mechanisms of Yak (*Bos grunniens*) to high-Altitude Environmental stress. Animals. 2021;11(8):2344.
5. Gao X, Wang S, Wang YF, Li S, Wu SX, Yan RG, et al. Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak. Nat Commun. 2022;13(1):4887.
6. Xin J, Chai Z, Zhang C, Zhang Q, Zhu Y, Cao H, et al. Methylome and transcriptome profiles in three yak tissues revealed that DNA methylation and the transcription factor ZGPAT co-regulate milk production. BMC Genom. 2020;21(1):731.
7. Xin JW, Chai ZX, Zhang CF, Zhang Q, Zhu Y, Cao HW, et al. Signature of high altitude adaptation in the gluteus proteome of the yak. J Exp Zool B Mol Dev Evol. 2020;334(6):362–72.
8. Xin JW, Chai ZX, Zhang CF, Zhang Q, Zhu Y, Cao HW, et al. Differences in proteomic profiles between yak and three cattle strains provide insights into molecular mechanisms underlying high-altitude adaptation. J Anim Phys Anim Nutr. 2022;106(3):485–93.
9. Xin JW, Chai ZX, Zhang CF, Yang YM, Zhang Q, Zhu Y, et al. Comparative analysis of Skeleton muscle Proteome Profile between Yak and cattle provides insight into high-altitude adaptation. Curr Proteom. 2021;18(1):62–70.
10. Xin JW, Chai ZX, Zhang CF, Zhang Q, Zhu Y, Cao HW, et al. Transcriptome profiles revealed the mechanisms underlying the adaptation of yak to high-altitude environments. Sci Rep. 2019;9(1):7558.
11. Xin JW, Chai ZX, Zhang CF, Zhang Q, Zhu Y, Cao HW, et al. Comparisons of lung and gluteus transcriptome profiles between yaks at different ages. Sci Rep. 2019;9(1):14213.
12. Hu Q, Ma T, Wang K, Xu T, Liu J, Qiu Q. The yak genome database: an integrative database for studying yak biology and high-altitude adaption. BMC Genomics. 2012;13:600.
13. Tarazona S, Arzalluz-Luque A, Conesa A. Undisclosed, unmet and neglected challenges in multi-omics studies. Nat Comp Sci. 2021;1(6):395–402.
14. Ji QM, Xin JW, Chai ZX, Zhang CF, Dawa Y, Luo S, et al. A chromosome-scale reference genome and genome-wide genetic variations elucidate adaptation in yak. Mol Ecol Res. 2021;21(1):201–11.
15. Jiangfeng F, Yuzhu L, Sijiu Y, Yan C, Gengquan X, Libin W, et al. Transcriptional profiling of two different physiological states of the yak mammary gland using RNA sequencing. PLoS ONE. 2018;13(7):e0201628.
16. Lucibelli F, Valoroso MC, Aceto S, Plant DNA, Methylation. An epigenetic Mark in Development, Environmental interactions, and evolution. Int j mol sci. 2022;23(15):8299.
17. Chai Z, Wu Z, Ji Q, Wang J, Wang J, Wang H, et al. Genome-wide DNA methylation and hydroxymethylation changes revealed epigenetic regulation of Neuromodulation and Myelination in Yak Hypothalamus. Front Genet. 2021;12:592135.

Publisher's Note