DATABASE   **Open Access**

iMarmot: an integrative platform for comparative and functional genomics of marmots

Baoning Liu[1,2], Liang Bai[1,2], Qingqing Yu[1,2], Fang Hu[1,2], Jing Wu[1,3], Sihai Zhao[1,2], Rong Wang[1,2], Weirong Wang[1,2], Yuanqing Tao[4], Jianglin Fan[5] and Enqi Liu[1,2*]

Abstract

Background: Marmots are large Holarctic rodents with unique biological features, making them potential animal models in various research fields. Due to the rapid accumulation of the genetic data in marmots, a highly integrative database is urgent needed.

Description: iMarmot is freely available on the web at http://www.marmotdb.org/ and currently contains the biological information of 14 marmots, genomic sequence of 6 marmots, syntenic relationship and orthologs among 3 marmots, and expression profiles of several hibernators and plague hosts. To assist with the genomic and transcriptomic analysis, we also integrated a set of analysis and visualization tools, such as KEGG or GO enrichment analysis, PCA, Blast, Muscle, GeneWise, Lastz, and JBrowse. Particularly, one DEGs (differentially expressed genes) module has been implemented in this database to visualize the gene expression changes in hibernators and plague hosts.

Conclusion: This database will provide comprehensive information and analysis platform for researchers interested in understanding the biological features of marmots.

Background

Marmots have been known since antiquity and firstly recognized by Thorington and Hoffman [1]. They are placed in the genus *Marmota*, family *Sciurid* with 14 extant species [1]. Most of them are living in social colonies and use loud whistles to communicate with one another. They prefer to habitat in the rock mountains or alpine meadows in the Holarctic, where there is a long and cold winter time without food [2–4]. To cope with the harsh environment, marmots have developed the capability of hibernation in burrows during winter.

Recently, global climate changes are affecting fauna and flora in various ways. Marmots living in alpine might encounter problems caused by rising of the timberline or from the abandonment of pastures. For instance, Vancouver Island marmot (*Marmota vancouverensis*), a typical alpine-dwelling marmot, has become one of the most endangered mammals in the world [5]. An international marmot network of scientists and managers has been established to work on the conservation issues of these species since 2008 (http://www.cons-dev.org/marm/MARM/EMARM/framarm/framarm.html). Meanwhile, plagues prevalence in some marmots were frequently reported to co-occur with sporadic epidemic in people all over the world, which also give rise to special supervision to the population fluctuations and epizootic activity of marmots [6–8]. Therefore, not only sociobiologists, conservationists and epidemiologist pay high

* Correspondence: liuenqi@xjtu.edu.cn
[1]Laboratory Animal Center, Xi'an Jiaotong University Health Science Center, No.76, Yanta West Road, Xi'an 710061, Shaanxi, China
[2]Research Institute of Atherosclerotic Disease, Xi'an Jiaotong University Cardiovascular Research Center, Xi'an 710061, Shaanxi, China
Full list of author information is available at the end of the article

Liu *et al. BMC Genomics*   (2020) 21:266

Page 2 of 8

attention to marmots, but also physiologists, who consider marmots as excellent animal models for the research of hibernation mechanism with possible applications to human medicine [4, 9].

Thanks to the development of sequencing technologies, large volumes of genomic and transcriptomic data have been generated. In 2015, the Alpine marmot (*Marmota marmota*) draft genome sequence was firstly released in the GenBank database of National Center for Biotechnology Information (NCBI) (accession number: GCA_001458135.2). In 2018, the Yellow-bellied marmot (*Marmota flaviventris*) draft genome was also released (accession number: GCA_003676075.2). Meanwhile, our group has assembled and published the genome sequence of the Himalayan marmot (*Marmota himalayana*) and re-sequenced the whole genomes of another four marmots, including Mongolia marmot (*Marmota sibirica*), Gray marmot (*Marmota baibacina*), Long-tailed marmot (*Marmota caudata*) and Yellow-bellied marmot in 2019 [10]. Particularly, RNA-seq or microarray data from a broad range of mammal hibernators and plague hosts provide novel insight into understanding the special biological features of marmots, especially the molecular mechanism of hibernation [10–16]. However, there is no available database to save, exploit, analyze and distribute these large-scale datasets. To facilitate the usage and application of these genomic, transcriptomic data and comparative results, we have developed the marmot database (iMarmot: www.marmotdb.org) as a highly integrative information platform with online analysis tools. We believe that the web-based platform with user-friendly interface and useful functions of will undoubtedly help the researches from a broad range of ecological and biomedical fields.

Construction and content
Database implementation
The iMarmot was implemented in Linux operation system and built using Akka HTTPServer (web server) and MySQL (database server). MySQL Database Management System (version 5.7.23) was used to process and organize all the data. The Twirl template engine and Bootstrap were applied to design and execute the interface components of the website. The query function was enforced based on Slick middleware tier. Boxplot, Barplot, Scatterplot and Heatmap plot were produced by Plotly 2.6.0 and Highcharts 5.0.12.

Biological features of marmots
We have collected and integrated the species information of 14 extant marmots from the international marmot network, ADW (animal diversity web), IUCN Redlist (International Union for Conservation of Nature Redlist) and published literatures. For enhancing the readability of the integrated infor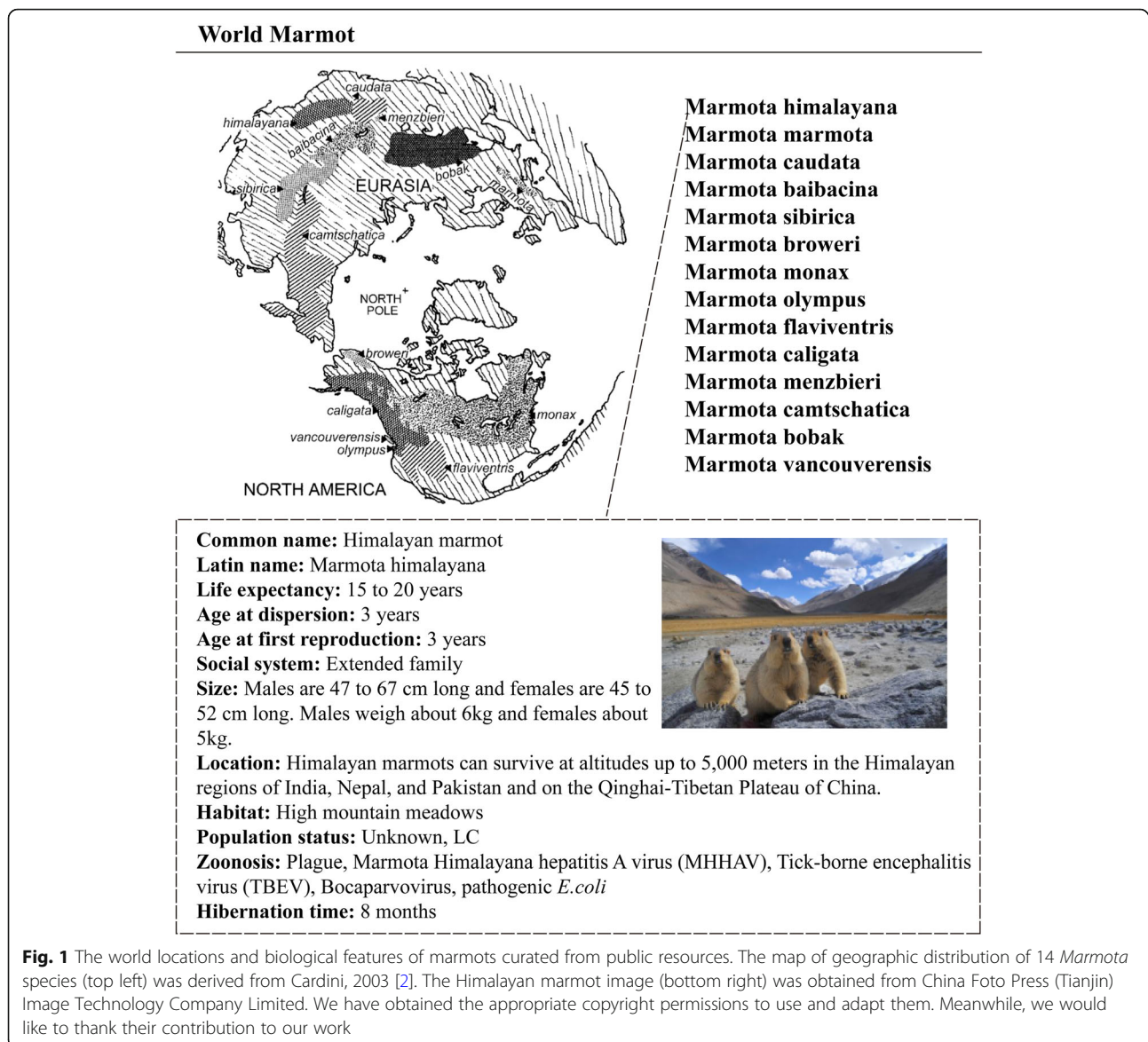mation, detailed data about marmots were shown in the categories of size, habitat, location, population status, zoonosis, hibernation time, social system, dispersion and reproductive age (Fig. 1).

Genome sequence and annotation
The iMarmot serves as a public warehouse to annotate, release and update the whole genome sequence of marmots. At present the iMarmot possesses three de novo genome assemblies, including Himalayan marmot, Alpine marmot and Yellow-bellied marmot. The assembly sequence of Alpine marmot and Yellow-bellied marmot were downloaded from the GenBank assembly database of NCBI, while the assembly of Himalayan marmot was derived from our previous work [10]. Totally, 21,609 protein-coding genes have been predicted for Himalayan marmot, 23,923 for Alpine marmot and 24,575 for Yellow-bellied marmot. A uniform protocol was applied to comprehensively annotate the function of these putative protein-coding genes. Functions of these protein-coding genes were assigned according to the best match of the alignments against various protein databases using BLASTP (*e*-value = $1e^{-5}$) [17], including the nonredundant protein (Nr) database, Swiss-Prot database and TrEMBL. Furthermore, unigenes were searched against the NCBI nonredundant nucleotide sequence (Nt) database using BLASTN [17] with a threshold of *e*-value = $1e^{-5}$. InterProScan (v4.3) [18] was used to collect the protein domain information and Gene Ontology (GO) terms [19]. Meanwhile, all protein-coding genes were aligned against KEGG (Kyoto Encyclopedia of Genes and Genomes), then KAAS (KEGG Automatic Annotation Server) was used for the KEGG pathway annotation [20]. As a result, 19,470 protein-coding genes were annotated for Himalayan marmot, 20,832 for Alpine marmot genome and 20,217 for Yellow-bellied marmot genome. Furthermore, 290 KEGG pathways and 13,334 GO terms were involved in the Himalayan marmot genome, 293 KEGG pathways and 11,389 GO terms in the Alpine marmot genome, 290 KEGG pathways and 11,400 GO terms in the Yellow-bellied marmot genome.

Synteny blocks and orthologous genes
The synteny blocks and homologous genes were detected for the genome comparisons of Himalayan marmot versus Alpine marmot and Himalayan marmot versus Yellow-bellied marmot. All of the protein sequence were firstly BLASTP against each other with an *e*-value cut-off of $1e^{-5}$ [17]. Based on these results, MCScanX was applied to determine the synteny blocks by default parameters [21]. Totally, 949 synteny blocks and 14,397 orthologous genes have been identified for the comparison of Himalayan marmot versus Alpine marmot, 951 synteny blocks and 13,935 orthologous genes for comparison of Himalayan marmot versus Yellow-bellied marmot. Each synteny block was embedded as a 'Synteny' subsection in its

Liu *et al. BMC Genomics* (2020) 21:266

Page 3 of 8

World Marmot

Marmota himalayana
Marmota marmota
Marmota caudata
Marmota baibacina
Marmota sibirica
Marmota broweri
Marmota monax
Marmota olympus
Marmota flaviventris
Marmota caligata
Marmota menzbieri
Marmota camtschatica
Marmota bobak
Marmota vancouverensis

Common name: Himalayan marmot
Latin name: Marmota himalayana
Life expectancy: 15 to 20 years
Age at dispersion: 3 years
Age at first reproduction: 3 years
Social system: Extended family
Size: Males are 47 to 67 cm long and females are 45 to 52 cm long. Males weigh about 6kg and females about 5kg.
Location: Himalayan marmots can survive at altitudes up to 5,000 meters in the Himalayan regions of India, Nepal, and Pakistan and on the Qinghai-Tibetan Plateau of China.
Habitat: High mountain meadows
Population status: Unknown, LC
Zoonosis: Plague, Marmota Himalayana hepatitis A virus (MHHAV), Tick-borne encephalitis virus (TBEV), Bocaparvovirus, pathogenic *E.coli*
Hibernation time: 8 months

Fig. 1 The world locations and biological features of marmots curated from public resources. The map of geographic distribution of 14 *Marmota* species (top left) was derived from Cardini, 2003 [2]. The Himalayan marmot image (bottom right) was obtained from China Foto Press (Tianjin) Image Technology Company Limited. We have obtained the appropriate copyright permissions to use and adapt them. Meanwhile, we would like to thank their contribution to our work

related gene feature (Fig. 2). The database users could view the detailed information of all the orthologous gene pairs located in the synteny region in a further linked web page (Fig. 3).
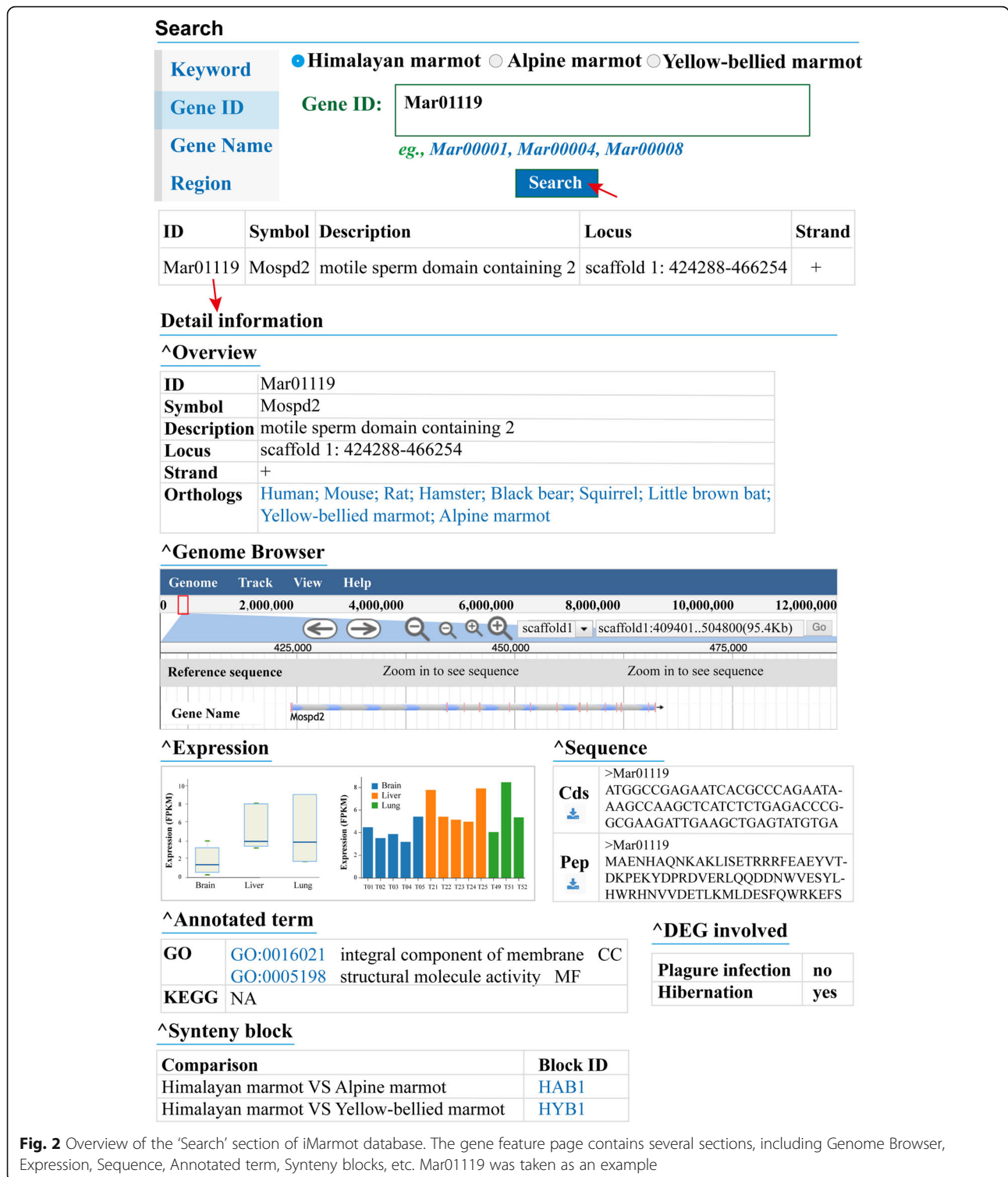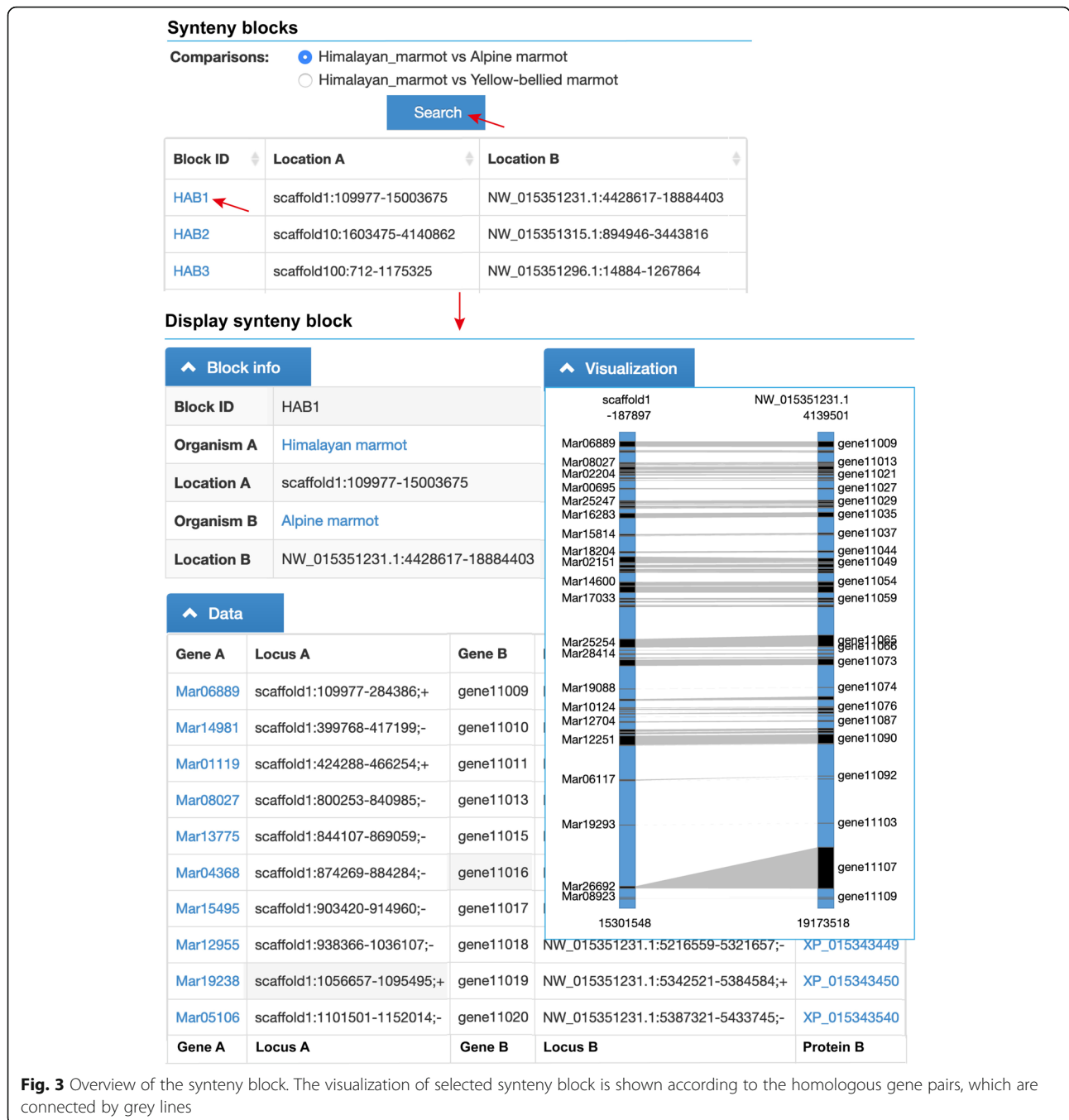
Genomic variations

The iMarmot database currently contains high quality SNP data of 5 marmots, including an albinistic Himalayan marmot, Long-tailed marmot, Gray marmot, Yellow-bellied marmot, and Mongolian marmot. The raw sequence data of 5 marmots were derived from our previous work. Clean reads were mapped to the Himalayan marmot reference genome using BWA [22] and sorted by SAMtools [23]. We then marked duplicates with Picard tools (v1.94). SNP were then called following a standard procedure using the Genome Analysis

Toolkit (GATK 3.0) with default parameters [24]. Variants were removed if they meet any of the following criteria: 1) an overall quality score of < 30; 2) a mapping quality score of < 60; 3) a phred-scaled p-value > 60; 4) a variant quality by depth score < 2; 5) genotype quality score < 5. In total, 24,192,293 SNPs of 5 marmots were retained. Users can visualize SNPs annotated in the Himalayan marmot genome in JBrowse and access each marmot's SNPs in Download page.

Expression data of hibernators and plague hosts

To understand the biological features of marmots, we have collected the expression profiles derived from RNA-seq or microarray data of several hibernators and plague hosts obtained from the Gene Expression Omnibus (GEO) database of NCBI. For RNA-seq data, the SRA

Fig. 2 Overview of the 'Search' section of iMarmot database. The gene feature page contains several sections, including Genome Browser, Expression, Sequence, Annotated term, Synteny blocks, etc. Mar01119 was taken as an example

toolkit (version 2.9.6) was used to convert the data from SRA to FASTQ format and then applied cutadapt (version 1.18) [25] with the following criteria for the quality control: 1) remove adaptors; 2) base quality > 20. After obtaining the high-quality data, mRNA sequencing reads were analyzed by unified protocol. First of all, HISAT was used to map the raw RNA-seq reads to their corresponding genomes [26]. Subsequently, the alignments were delivered to StringTie [27], which would assemble the full-length transcripts representing multiple splice variants for each

Fig. 3 Overview of the synteny block. The visualization of selected synteny block is shown according to the homologous gene pairs, which are connected by grey lines

gene locus. The merged full-length transcripts were given to StringTie again for obtaining better quantification of transcript richness. At last, Ballgown obtained all transcripts and abundances from StringTie and classified them according to the sample group [28]. For microarray, the raw data was downloaded from the GEO repository. Then R package Limma was applied to perform the data standardization and assess the differential expression via pairwise comparison. If multiple probes existed for one gene, the mean expression of these probes is considered as the expression of this gene. The criterion is *p*-value less than 0.05. Totally, 6 plague infection projects and 14 hibernation projects were retained, containing 536 samples, covering 14 tissues and 10 species.

Utility and discussion
Search and browse
To facilitate the query for gene information, we built the index for various types of annotation data, such as gene symbol, GO terms and KEGG pathway, and

Liu *et al. BMC Genomics*   (2020) 21:266

Page 6 of 8

provided a basic search form for the Himalayan marmot genome, Alpine marmot and Yellow-bellied marmot. The users only need to input a single or batch of gene IDs, gene symbols or genome regions for retrieving the specific gene information. Each gene has a detailed information page, which is divided into subsections including the general information, gene structure implemented in JBrowse, sequence of protein and CDS, annotated term, expression pattern, as well as synteny block (Fig. 2). Moreover, users could also browse the GO or KEGG category and synteny blocks for the Himalayan marmot, Alpine marmot and Yellow-bellied marmot. Genes within each annotated category or synteny block will be displayed and further linked to Genome browser. In addition, we also provided a search function aside the main menu, which allows global keyword queries against all the records stored in the database.

Genome browser

In iMarmot database, we have implemented JBrowse, a commonly used Genome browser, to display the genome sequences, gene models, genomic variations [29]. Currently, de novo assemblies of 3 marmots and genomic variations of 5 marmots were imported into JBrowse [29]. Particularly, there is a Himalayan marmot genome track embedded in the Genome Browser subsection in the gene feature page. This subsection provides a graphical and informative view of the gene structure of the searched gene model (Fig. 2). In addition, the JBrowse has several supporting tracks of non-coding RNA, pseudogene, repetitive sequence and GC content for the Himalayan marmot. In addition, JBrowse also provides access to easily add the genomic information of other marmots.

DEGs visualization

In order to infer the molecular mechanism underlying the biological features of marmots, we have developed the DEGs module for visualization the differential expression profiles derived from the transcriptome data of several hibernators and plague hosts. Users can easily filter the DEGs within one project by tissues, comparisons, up or down regulation status, fold change value, and *p*-value. The result page of differential expression analysis contains the parameters used for statistical analysis, project description, significant DEGs list, and linkages to download file containing the differential expression analysis results. The file format of DEGs directly conform to the request of Online tools, such as enrichment analyses, PCA and correlation heatmap.
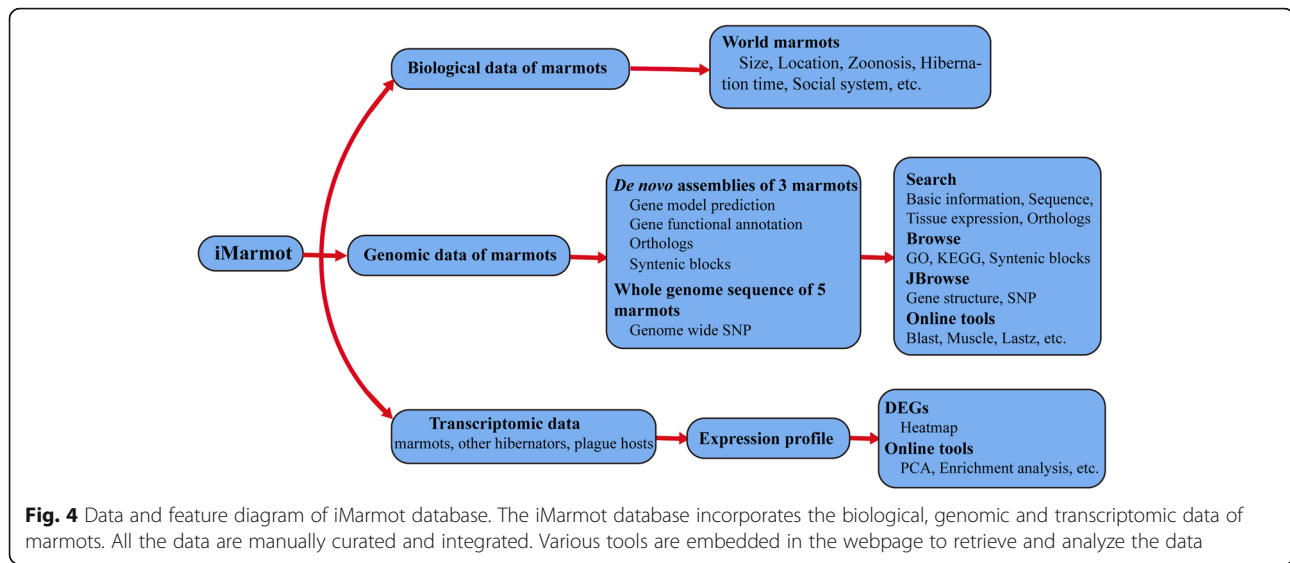
Online tools

Numerous candidate genes were proposed in the comparative genomic and transcriptomic studies. The designating of those genes into specific functional categories would definitely facilitate the perception of the potential molecular mechanism under certain pathophysiological status. By enrichment analysis, we are able to detect the significantly overrepresented functional classes in a given gene list, which provide a useful tool to infer the major influenced functional categories or biochemical pathways in organisms with specific experimental treatments. In the iMarmot database, we have embedded the two commonly used tools: GO and KEGG tools, which facilitate the users to find the significantly enriched GO terms and KEGG pathways by uploading an interesting gene list. The 'GO tool' was executed by involving the GO:: TermFinder Perl module. The 'Pathway tool' was operated based on the PathwayTools [30]. The hypergeometric distribution test was applied to determine the significance of enriched gene class in either GO or KEGG tool [31]. We have also integrated some other frequently used genomic and transcriptomic tools, such as Blast [17], Muscle [32], Lastz [33], Genewise [34], PCA, hCluster and correlation heatmap. Users can upload their own data, fill in the parameters and submit the task. Then a result page will appear automatically.

Other functions

In the Download page, we provide linkages for batch data download. The users can directly download all the processed data, including the whole genome sequence, protein and CDS sequences, gene function annotation, SNP data and expression profile data. Meanwhile, iMarmot users are also welcome to share their own data with us. In the Submit page, users could find the format requirements of each data type that can be integrated to the database. In addition, a "How to" page and a "FAQ" page give the instructions of main functions to the iMarmot users and help them easily manage the database.

Conclusions

In conclusion, we have developed the iMarmot database, which serves as a central portal for marmot species. iMarmot stores the biological information, genome sequences, gene information, synteny blocks, orthologous of three marmots, SNP data of five marmots and transcriptomic data from various hibernators and plagues hosts (Fig. 4). As iMarmot is a manually integrated database, all the raw data collected/submitted would be reprocessed by our own customized pipelines. For doing this, different results may generate while compared to the original published papers. The processed data is

Liu *et al. BMC Genomics*   (2020) 21:266

Page 7 of 8

Fig. 4 Data and feature diagram of iMarmot database. The iMarmot database incorporates the biological, genomic and transcriptomic data of marmots. All the data are manually curated and integrated. Various tools are embedded in the webpage to retrieve and analyze the data

further related the retrieval and analysis module. The retrieval and analysis module, including various query, visualization and analysis tools, execute the main functions of the iMarmot database (Fig. 4). We believe that this database will not only broaden the understanding of genetic mechanism underlying marmots' biological features, but also attract more and more attentions on the conservation of endangered marmot species.

Abbreviations
ADW: Animal Diversity Web; DEGs: Differentially Expressed Genes; GEO: Gene Expression Omnibus; GO: Gene Ontology; IUCN Redlist: International Union for Conservation of Nature Redlist; KEGG: Kyoto Encyclopedia of Genes and Genomes; NCBI: National Center for Biotechnology Information; Nr: Nonredundant protein database; Nt: Nonredundant nucleotide sequence database

Authors' contributions
BL performed the data analysis. BL, QY, FH and JW constructed the database. SZ, RW, WW and YT collected the data and prepared the basic dataset. BL and LB wrote the manuscript. JF and EL conceived the work and revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets generated and analyzed in the current study are freely available on the Download page of iMarmot database with the web link: http://www.marmotdb.org/MARDB/download/toIndex.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Laboratory Animal Center, Xi'an Jiaotong University Health Science Center, No.76, Yanta West Road, Xi'an 710061, Shaanxi, China. [2]Research Institute of Atherosclerotic Disease, Xi'an Jiaotong University Cardiovascular Research Center, Xi'an 710061, Shaanxi, China. [3]Laboratory Animal Center, Ningxia Medical University, Yinchuan 750004, Gansu, China. [4]Qinghai Institute for Endemic Disease Prevention and Control, Xining 811602, Qinghai, China. [5]Department of Molecular Pathology, Faculty of Medicine, Interdisciplinary Graduate School of Medicine, University of Yamanashi, Yamanashi 409-3898, Japan.

References
1.   Hoffmann RS, Anderson CG, Thorington RW, Heaney LR. Family sciuridae. In: Wilson DERDM, editor. Mammal species of the world. Washington DC: Smithsonian Institution Press; 1993. p. 419–65.
2.   Cardini A. The geometry of the marmot (rodentia: sciuridae) mandible: phylogeny and patterns of morphological evolution. Syst Biol. 2003;52(2): 186–205.
3.   Yan L, Kudej RK, Vatner DE, Vatner SF. Myocardial ischemic protection in natural mammalian hibernation. Basic Res Cardiol. 2015;110(2):9.
4.   Armitage KB, Blumstein DT, Woods BC. Energetics of hibernating yellow-bellied marmots (*Marmota flaviventris*). Comp Biochem Physiol A Mol Integr Physiol. 2003;134(1):101–14.
5.   Cardini A, Thorington RW, Polly PD. Evolutionary acceleration in the most endangered mammal of Canada: speciation and divergence in the Vancouver Island marmot (Rodentia, Sciuridae). J Evol Biol. 2007;20(5):1833–46.
6.   Sariyeva G, Abdel Z, Shabunin A, Sagiyev Z, Abdikarimov S, Bazarkanova G, et al. Current status of the Sari-Dzhas natural focus of plague, Kyrgyzstan: epizootic activity and marmot population. Vector Borne Zoonotic Dis. 2018; 18(10):524–32.
7.   Ge P, Xi J, Ding J, Jin F, Zhang H, Guo L, et al. Primary case of human pneumonic plague occurring in a Himalayan marmot natural focus area Gansu Province, China. Int J Infect Dis. 2015;33:67–70.
8.   Suntsov VV. Sympatric speciation of the plague microbe *Yersinia pestis*: monohostal specialization in the host-parasite marmot-flea (*Marmota sibirica-Oropsylla silantiewi*) system. Izv Akad Nauk Ser Biol. 2016;2:117–27.
9.   Armitage KB, Woods BC. Group hibernation does not reduce energetic costs of young yellow-bellied marmots. Physiol Biochem Zool. 2003;76(6):888–98.
10. Bai L, Liu B, Ji C, Zhao S, Liu S, Wang R, et al. Hypoxic and cold adaptation insights from the Himalayan marmot genome. iScience. 2019;11:519–30.

Liu *et al. BMC Genomics*   (2020) 21:266

Page 8 of 8

11. Stasulli NM, Eichelberger KR, Price PA, Pechous RD, Montgomery SA, Parker JS, et al. Spatially distinct neutrophil responses within the inflammatory lesions of pneumonic plague. MBio. 2015;6(5):e01530–15.
12. Comer JE, Sturdevant DE, Carmody AB, Virtaneva K, Gardner D, Long D, et al. Transcriptomic and innate immune responses to *Yersinia pestis* in the lymph node during bubonic plague. Infect Immun. 2010;78(12):5086–98.
13. Hammamieh R, Muhie S, Borschel R, Gautam A, Miller SA, Chakraborty N, et al. Temporal progression of pneumonic plague in blood of nonhuman primate: a transcriptomic analysis. PLoS One. 2016;11(3):e0151788.
14. Ou J, Ball JM, Luan Y, Zhao T, Miyagishima KJ, Xu Y, et al. iPSCs from a hibernator provide a platform for studying cold adaptation and its potential medical applications. Cell. 2018;173(4):851–863.e816.
15. Fedorov VB, Goropashnaya AV, Stewart NC, Toien O, Chang C, Wang H, et al. Comparative functional genomics of adaptation to muscular disuse in hibernating mammals. Mol Ecol. 2014;23(22):5524–37.
16. Riemondy KA, Gillen AE, White EA, Bogren LK, Hesselberth JR, Martin SL. Dynamic temperature-sensitive A-to-I RNA editing in the brain of a heterothermic mammal during hibernation. RNA. 2018;24(11):1481–95.
17. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
18. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33(Web Server issue):W116–20.
19. Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O'Donovan C, Martin MJ, et al. The UniProt-GO annotation database in 2011. Nucleic Acids Res. 2012;40(Database issue):D565–70.
20. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(Web Server issue):W182–5.
21. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
22. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–60.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
25. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):3.
26. Kim D, Langmead B, Salzberg SL. HISAT:a fast spliced aligner with low memory requirements.pdf. Nat Methods. 2015;12(4):357.
27. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.
28. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol. 2015;33(3):243–6.
29. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, et al. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 2016;17:66.
30. Karp PD, Paley S, Romero P. The pathway tools software. Bioinformatics. 2002;18(Suppl 1):S225–32.
31. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO:: TermFinder--open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. Bioinformatics. 2004;20(18):3710–5.
32. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
33. S. H.R. Improved pairwise alignment of genomic DNA. The Pennsylvania State University. 2007.
34. Birney E, Durbin R. Using GeneWise in the Drosophila annotation experiment. Genome Res. 2000;10(4):547–8.

Publisher's Note