

DATABASE

Open Access



BmncRNadb: a comprehensive database of non-coding RNAs in the silkworm, *Bombyx mori*

Qiu-Zhong Zhou¹, Bindan Zhang², Quan-You Yu¹ and Ze Zhang^{1*}

Abstract

Background: Long non-coding RNAs (lncRNAs) may play critical roles in a wide range of developmental processes of higher organisms. Recently, lncRNAs have been widely identified across eukaryotes and many databases of lncRNAs have been developed for human, mouse, fruit fly, etc. However, there is rare information about them in the only completely domesticated insect, silkworm (*Bombyx mori*).

Description: In this study, we systematically scanned lncRNAs using the available silkworm RNA-seq data and public unigenes. Finally, we identified and collected 6281 lncRNAs in the silkworm. Besides, we also collected 1986 microRNAs (miRNAs) from previous studies. Then, we organized them into a comprehensive and web-based database, BmncRNadb. This database offers a user-friendly interface for data browse and online analysis as well as the three online tools for users to predict the target genes of lncRNA or miRNA.

Conclusions: We have systematically identified and collected the silkworm lncRNAs and constructed a comprehensive database of the silkworm lncRNAs and miRNAs. This work gives a glimpse into lncRNAs of the silkworm and lays foundations for the ncRNAs study of the silkworm and other insects in the future. The BmncRNadb is freely available at <http://gene.cqu.edu.cn/BmncRNadb/index.php>.

Keywords: Silkworm, Long non-coding RNAs, RNA-seq, BmncRNadb

Background

The ENCODE project estimates that 62–75 % of the human genome are transcribed, but only 2 % of the transcripts can be translated to proteins [1, 2]. The GENCODE 22 release contains 19,814 protein-coding genes, 15,900 long non-coding RNA genes and 9894 small non-coding RNA genes [3]. These suggest that non-coding RNAs (ncRNAs) constitute a large fraction of the eukaryote transcriptome [4, 5].

Long non-coding RNAs (lncRNAs) are transcripts of DNA that are usually considered to be ≥ 200 nt (nucleotide) and do not have apparent coding capacity [6–10]. lncRNAs are widely present in the eukaryotic genomes [4, 11]. In the postgenomic era, since the development and application of next-generation sequencing

technologies, a large number of long non-coding RNAs have been identified in different species (e.g. human [12], mouse [13], fruit fly [14], etc.). Although the functions of most lncRNAs are still unclear, more and more evidence has proven that they play critical roles in various biological processes including cellular differentiation [15], epigenetics [16], transcriptional regulation [17] and immune response [18]. For example, in the placental mammals, *Xist* (X-inactive specific transcript) is a long non-coding RNA on the X-chromosome and takes part in inactivation of X-chromosome during the early developmental process of female embryo [8, 19]. In addition, thousands of lncRNAs have been reported in the insects and some of them show important roles in the life events of insects [14, 20–25]. *Acal* acts as a novel negative dorsal closure regulator during *Drosophila* embryogenesis and *Lncov1* is involved in the autophagic cell death of ovarioles in *Apis mellifera* [26, 27]. Therefore, lncRNAs are important functional elements in the genomes of higher organisms.

* Correspondence: zezhang@cqu.edu.cn

¹Laboratory of Evolutionary and Functional Genomics, School of Life Sciences, Chongqing University, Huxi Campus, No. 55 Daxuecheng South Rd, Shapingba, Chongqing 401331, China

Full list of author information is available at the end of the article



The domesticated silkworm, *Bombyx mori*, is one of important model organisms for Lepidoptera, more and more transcriptomic resources are available for the silkworm. The ncRNAs, especially microRNAs (miRNAs) were identified in the silkworm by Solexa sequencing [28]. In addition, the miRNAs are also reported that may take part in the fibroin synthesis and fibroin transport in the domesticated silkworm [29]. As one important member of the ncRNAs, lncRNAs also play key roles in the silkworm. The first silkworm lncRNA, *Fben-1* (female-brain expressed noncoding RNA-1) was identified in female-brain and may be involved in sexually dimorphic brain functions [30]. Although 11,810 silkworm lncRNAs are identified in different tissues with the loose standard, the loose threshold may lead to high false positive rate for lncRNA identification. Thus, it is still necessary to systematically identify the lncRNAs in the silkworm with more RNA-seq data and more stringent pipeline [30, 31].

Moreover, many databases on the information of lncRNA have been developed such as NONCODE, lncRNAdb, lncRBase, DeepBase [32–35], but the information of the silkworm lncRNA is almost blank in the present lncRNA databases [32–34, 36–39]. Currently, miRBase and microrna are two large databases containing miRNA information, however, the information of the silkworm miRNAs in the miRBase is rare and redundancy [40, 41]. Thus, in this study, we used a comprehensive approach to identify lncRNAs in the silkworm with all newly released RNA-seq data in the SRA (Sequence Read Archive) database and the unigene data [7, 13, 14, 25, 42–48]. The identified lncRNAs are organized into a database for user browser. In order to offer

more information about the silkworm ncRNAs, the available silkworm miRNAs and previously reported lncRNAs are also added to the database [28, 29, 40]. The database can be accessed at the website <http://gene.cqu.edu.cn/BmncRNAdb/index.php>.

Construction and content

Database architecture

The BmncRNAdb database implementation is based on the Gentoo Linux system with the tools of Apache 2.0 [49], PHP 5.4 (Personal home page Hypertext Pre-processor) [50], MySQL 5.16 [51], and Perl 5.12 [52]. The database architecture is illustrated in Fig. 1. Apache + PHP processes the user request and responds to user by the web browser. MySQL is used to create data model and data storage. The Perl script calls the background program to execute server request and returns the results to server by the CGI (Common Gateway Interface). Next, the web server will send the results of background program to BmncRNAdb user by the internet.

Data sets

New version of the silkworm genome sequence was downloaded from the silkworm genome database, SilkDB v2.0 [53]. The silkworm protein-coding genes were retrieved from Ensembl database (<http://meta-zoa.ensembl.org/>) [54]. All the silkworm RNA-seq data were downloaded from NCBI (National Center for Biotechnology Information) SRA databases (Additional file 1: Table S1) (<http://www.ncbi.nlm.nih.gov/sra>) [55–65]. The silkworm unigenes were downloaded from NCBI UniGene database [66]. Non-redundant protein (nr)

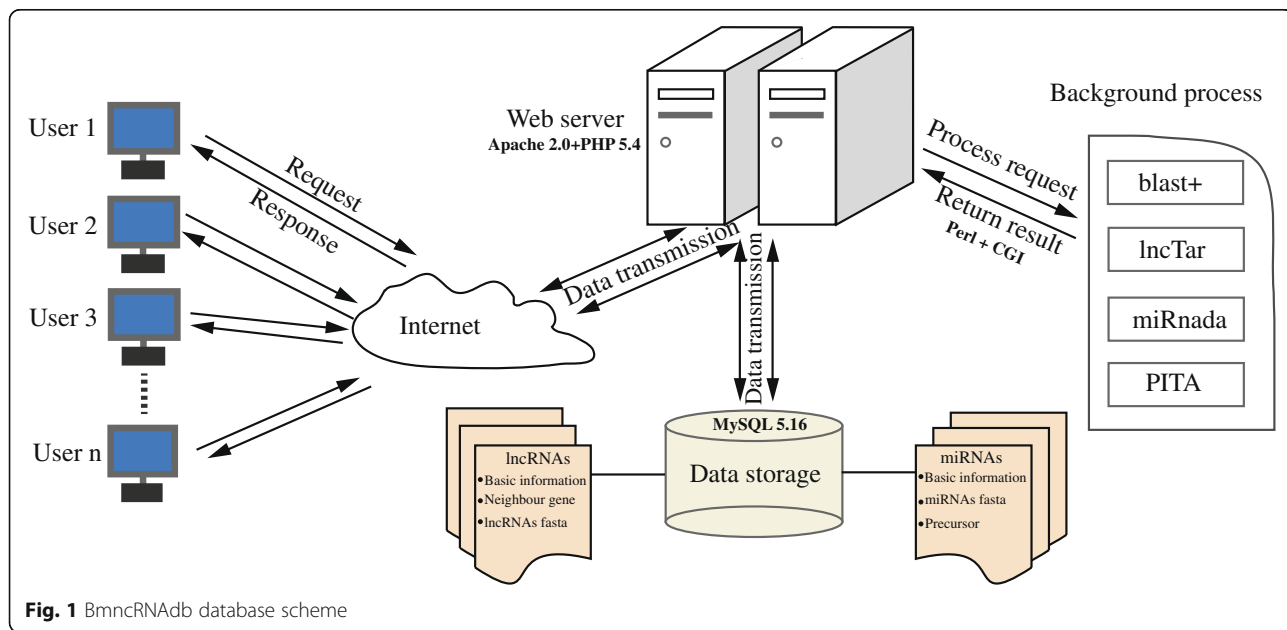


Fig. 1 BmncRNAdb database scheme

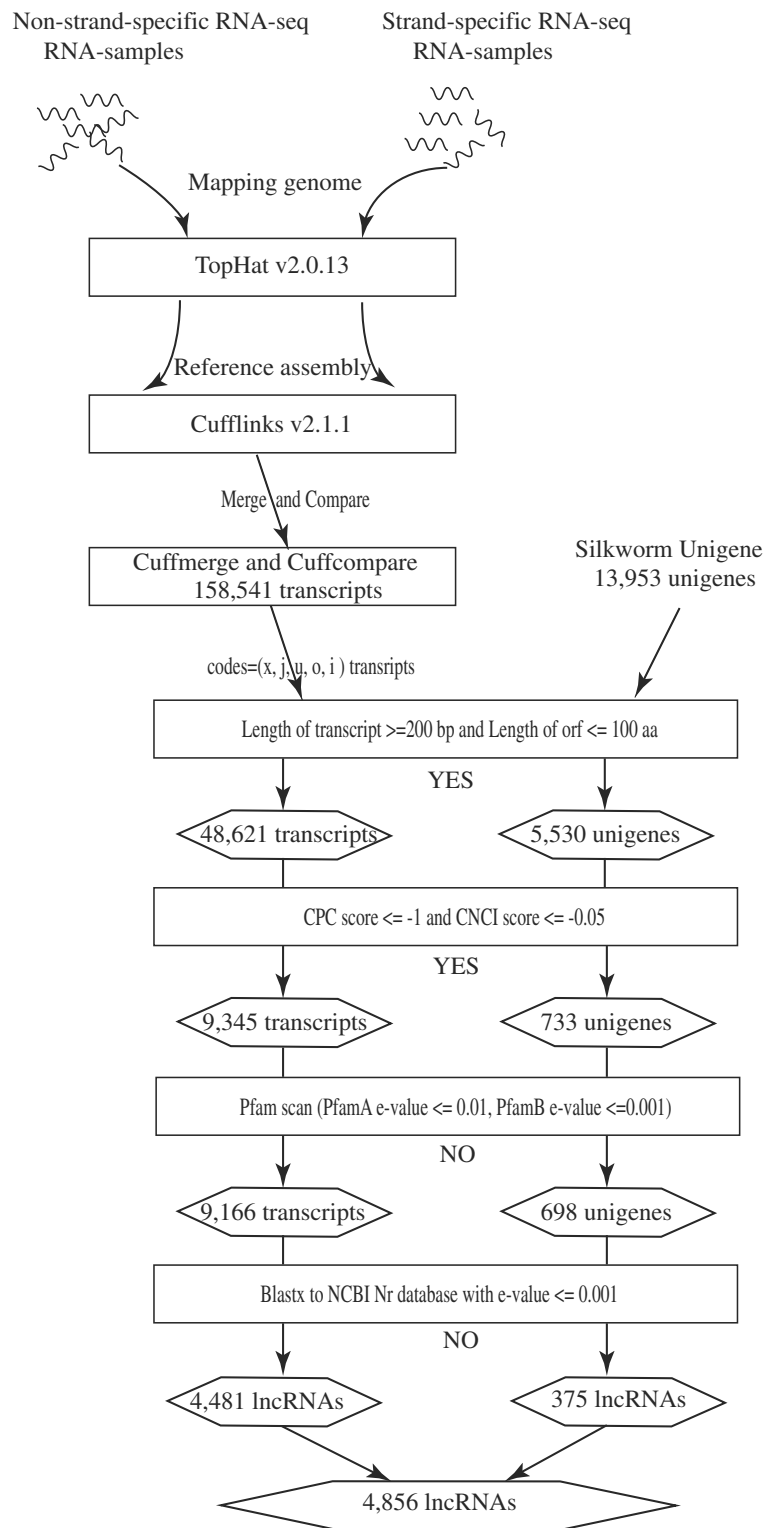


Fig. 2 Flowchart of lncRNAs identification in the silkworm. Left pipeline means the identification of lncRNAs from RNA-seq, and the right pipeline means the identification of lncRNAs from unigenes

sequences were also obtained from NCBI database [66]. A comprehensive protein database, Uniref100, was downloaded from UniProt databases (<http://www.uniprot.org/>) [67]. The current released (Pfam 28) Pfam-A and Pfam-B were obtained from EBI ftp website (<ftp://ftp.ebi.ac.uk/>) [68].

Genome-wide identification of lncRNAs in the silkworm

Two types of data from the silkworm were used for identification of the silkworm lncRNAs. The first is the silkworm RNA-seq data. Forty-one RNA-seq datasets were published by other research groups before January 15(th), 2015 and four RNA-seq datasets were produced by our laboratory (Additional file 1: Table S1) [55–65]. All the RNA-seq data are used to reconstruct the silkworm transcriptome using the software Tophat v2.0.13 and Cufflinks v2.1.1 [7, 25, 42, 43, 45, 46, 48, 69, 70]. The second is the silkworm unigenes. The unigene transcripts were assembled from EST (Expressed Sequence Tag) and some lncRNAs are also contained in the unigene transcripts [43]. Thus, the transcripts assembled from RNA-seq data and unigenes are used to identify lncRNAs in this study. The whole workflow to identify the silkworm lncRNAs is shown in Fig. 2.

RNA-seq short-reads assembly

QC (quality control) Toolkit of NGS (Next-Generation Sequencing) is used to control the reads quality of forty-five RNA-seq datasets [71]. High-quality RNA-seq reads are considered as clean reads data. The clean reads data were mapped to the newly assembled silkworm genome sequence with TopHat v2.0.13 [69]. Mapped reads for each sample were assembled using Cufflinks v2.1.1 with the protein-coding gene annotations separately [13, 69, 70]. All the sample assemblies are integrated into a merged assembly by Cuffmerge v2.1.1. We then used Cuffcompare v2.1.1 to generate different categories of the transcripts for the merged assembly [25, 43]. After that, 158,541 transcripts were generated from the transcriptome assembly. The five categories of the transcripts are retained including falling entirely within a reference intron (code='i'), sharing at least one splice junction with a reference transcript (code='j'), generic exonic overlap with a reference transcript (code='o'), unknown or intergenic

transcript (code='u') and exonic overlap with reference on the opposite strand (code='x') [13, 69]. These five categories of the transcripts and the silkworm unigenes are used to identify lncRNAs in the next step.

Protein-coding transcripts exclusion

lncRNAs are usually considered to have length $>=200$ bp and ORFs (open reading frame) $<=100$ aa (amino acids) [7, 42, 43, 70]. The assembled transcripts and unigene transcripts with the length <200 bp or ORFs >100 aa are excluded by the Perl Script, respectively. The retained 48,621 transcripts and 5530 unigenes are evaluated to the protein-coding potentiality for each transcript by the two tools, CPC (Coding Potential Calculator) and CNCI (Coding-Non-Coding Index) [42, 43, 70, 72–76]. In general, transcripts with protein-coding score <0 in the CPC or CNCI are regarded as non-coding potentiality [72, 73]. The CPC and CNCI can be complementary and improve the positive rate for lncRNA identification [72, 73]. Thus, we used two tools (CPC and CNCI) and set the protein-coding score -1 as threshold in the CPC and -0.05 as threshold in the CNCI [42, 43]. Only those transcripts have CPC score $<= -1$ and CNCI score $<= -0.05$ are retained. The retained 9345 transcripts and 733 unigenes are translated into the corresponding proteins by six frame translation and then the proteins were used to search against Pfam-A and Pfam-B databases. Transcripts that have significant hits against Pfam-A and Pfam-B will be removed [10]. At last, the blastx searches against NCBI Non-redundant protein (Nr) databases with the option e-value 0.001 were performed using retained transcripts [48, 77]. Transcripts that have a hit with Nr protein sequences were deleted in this process. In the end, 4856 lncRNAs were identified from the silkworm RNA-seq and unigenes (Fig. 2). The 95.65 % of lncRNAs belong to the 'u' (Unknown, intergenic transcript) category (Table 1). Moreover, in order to reduce the false positive rate for lncRNAs, 11,810 previously reported lncRNAs were re-identified by our stringent pipeline [31] and 1565 high-quality lncRNAs were retained, suggesting that the false positive rate for identification of the silkworm lncRNAs in previous study may be much higher. After removing the redundancy, 6821 lncRNAs were recorded in the BmncRNAdb database. A proven previously lncRNA,

Table 1 The summary of the silkworm lncRNAs identified by RNA-seq

Class code	Transcript number	Percentage	Description
j	16	0.36 %	At least one splice junction is shared with a reference transcript
o	26	0.58 %	Generic exonic overlap with a reference transcript
u	4286	95.65 %	Unknown, intergenic transcript
x	151	3.37 %	Exonic overlap with reference on the opposite strand
i	2	0.04 %	A transfrag falling entirely within a reference intron
total	4481	100 %	

Fben-1, is identified by our pipeline. This shows the reliability of our pipeline.

Characteristics of the silkworm lncRNAs

We surveyed the comprehensive characteristics of the silkworm lncRNAs including the length distribution, GC content, exon number distribution, link with transposable

elements, sequence conservation and correlation with neighbor protein genes (Fig. 3). The silkworm lncRNAs have shorter transcript length than the protein-coding genes (Fig. 3a). The lncRNAs also have lower GC content and less exon number than the protein-coding genes (Fig. 3b and c). However, the lncRNAs have a large degree of overlap of transposable elements in the silkworm

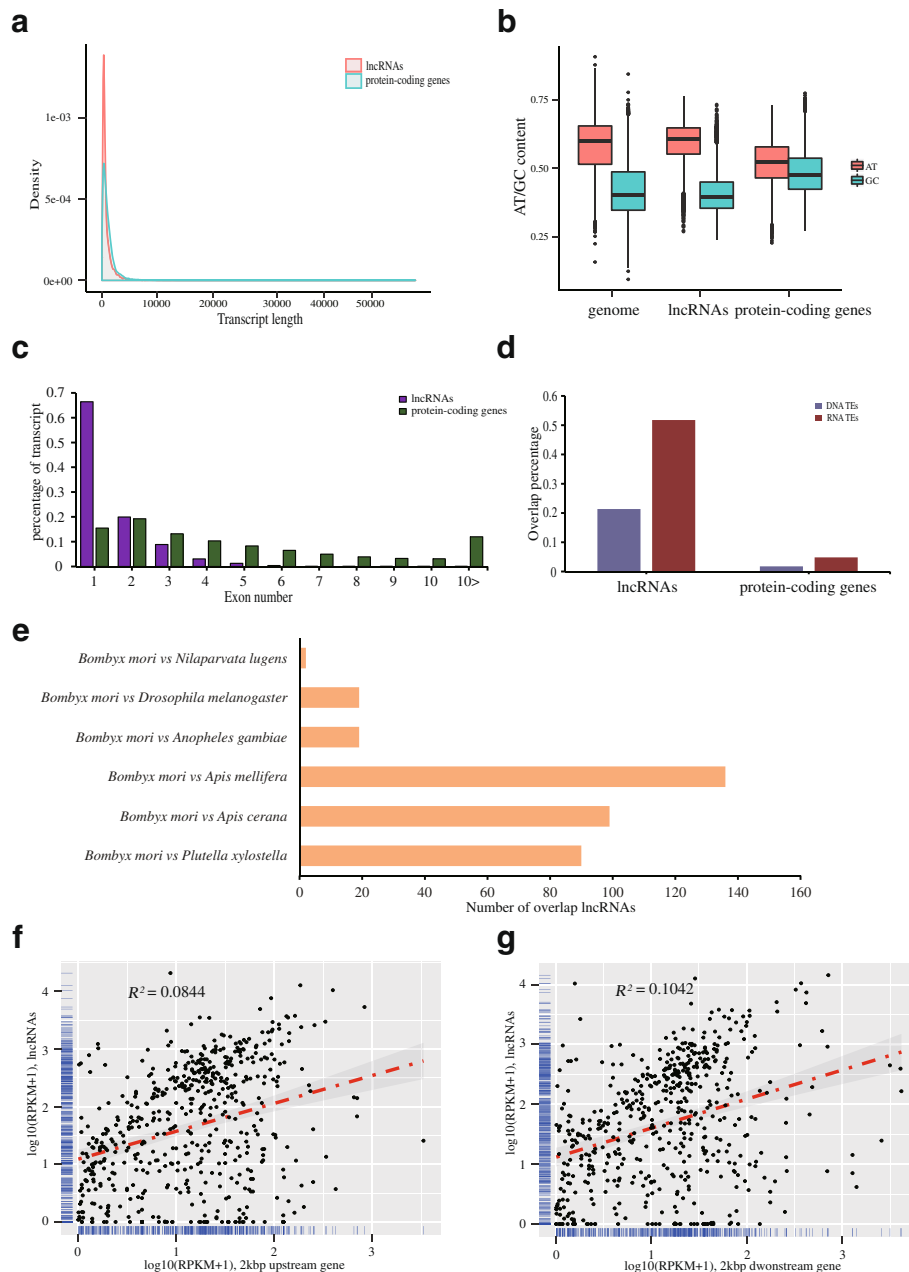


Fig. 3 Characteristics of silkworm lncRNAs. **a** Transcript length distribution in the lncRNAs and protein-coding genes. **b** AT/GC content among the silkworm genome, lncRNAs, and protein-coding genes. **c** Transcript exon number distribution. **d** Percentage of transcripts with at least 15 nt overlapping with transposable elements. **e** Sequence conservation between the silkworm (*Bombyx mori*) lncRNAs and other insects' lncRNAs (*Plutella xylostella*, *Apis cerana*, *Apis mellifera*, *Anopheles gambiae*, *Drosophila melanogaster*, *Nilaparvata lugens*). The x axis is the number of silkworm lncRNAs that overlap with other insects' lncRNAs at least 15 bp. **f** Expression scatter diagram in the lncRNAs and 2 kbp upstream protein-coding genes (Spearman test). **g** Expression scatter diagram in the lncRNAs and 2 kbp downstream protein-coding genes (Spearman test)

(Fig. 3d). The similar results were also reported in the previous studies [31, 70, 78]. The silkworm lncRNA that overlaps with other insect lncRNAs at least 15 bp is defined as sequence conservation [78]. Based on the standard, 136 silkworm lncRNAs show sequence conservation with the *Apis mellifera* (Hymenoptera) lncRNAs, the highest sequence conservation (Fig. 3e). And the silkworm lncRNAs also have relatively high sequence conservation with the *Plutella xylostella* (Lepidoptera) and *Apis cerana* (Hymenoptera) lncRNAs. However, the silkworm lncRNAs have low sequence conservation with the *Drosophila melanogaster*, *Anopheles gambiae* and *Nilaparvata lugens* lncRNAs. Furthermore, the expressions of the genes within 2 kbp neighbor regions (2 kbp upstream and 2 kbp downstream) of the putative silkworm lncRNAs are not significantly correlated with the expressions of lncRNAs (Spearman test) (Fig. 3f and g).

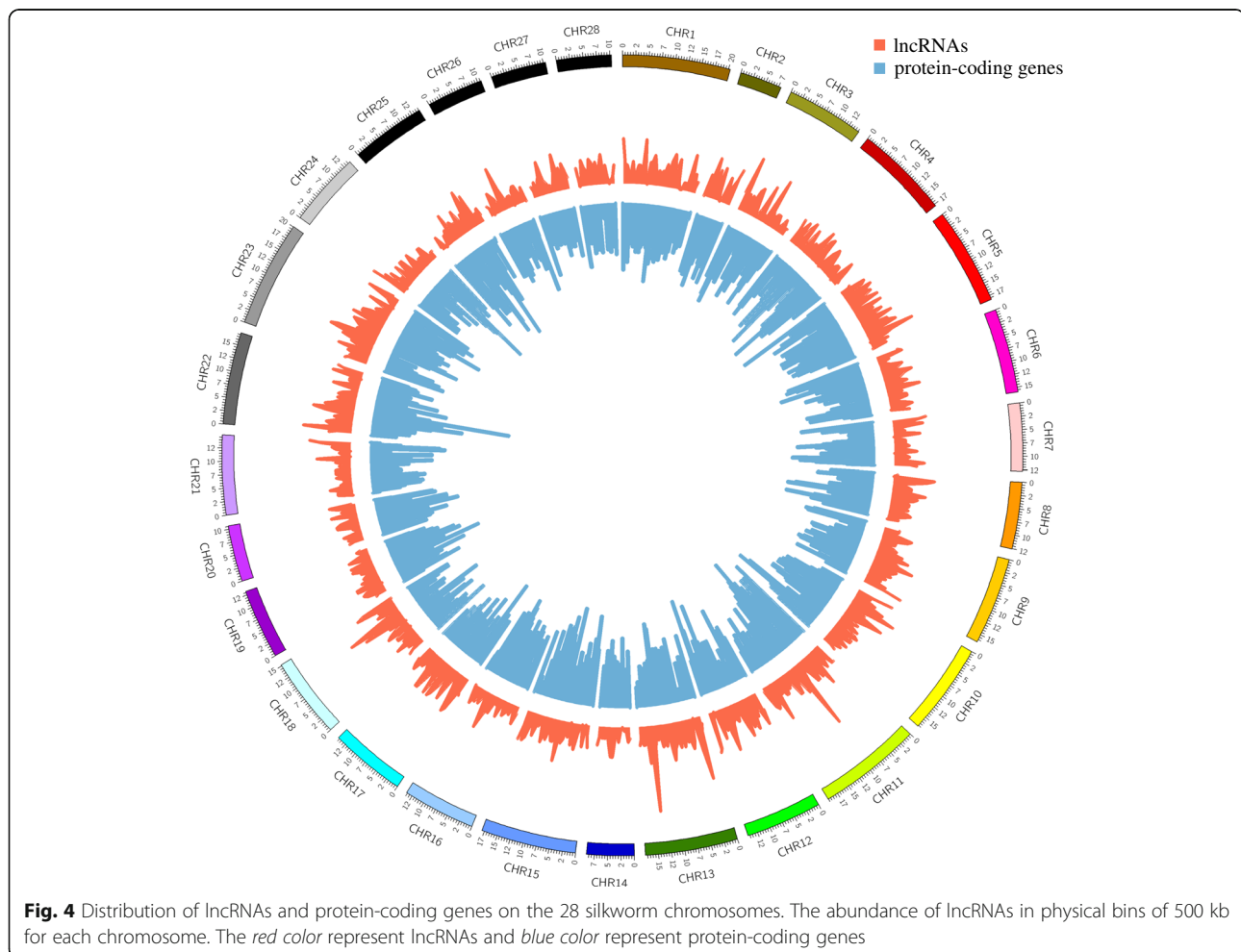
Collection of microRNAs in the silkworm

The silkworm microRNAs were comprehensively identified in the whole body, anterior or middle and posterior silk glands by next generation sequencing technology

[28, 29]. The datasets of the silkworm miRNAs were collected from miRBase and previous studies [28, 29, 40]. All miRNAs are compared by sequence pair-wise to remove redundancy and manual correction [79]. The formats of miRNAs are unified by the Perl Scripts.

Utility and discussion

Using the pipeline in Fig. 2, we identified and collected 6281 lncRNAs. About 58.67 % of lncRNAs can be located on the silkworm chromosomes and the rest lncRNAs are located in the scaffolds that cannot be mapped to the silkworm chromosomes. All the 28 chromosomes harbored lncRNAs. Interestingly, the chromosomal distribution of the lncRNAs is not significantly correlated with the protein-coding genes (Spearman $r = 0.017$, P -value = 0.62) (Fig. 4). This is consistent with the observation in the lncRNAs of human [12]. Moreover, we also collected 1986 miRNAs from previous studies and public databases [28, 29, 40]. In the end, we organized these silkworm lncRNAs and miRNAs into the BmncRNAdb database (<http://gene.cqu.edu.cn/BmncRNAdb/index.php>). The database contains six



functional sections, data browse, keywords search, Blast alignment, lncRNA target gene discovery, miRNA target gene discovery and data download.

Data browse

In the left navigation, clicking the 'Browse,' users can browse the information of lncRNAs including lncRNA name, scaffold, start position, end position, exon number and length (Fig. 5a). By clicking the lncRNA name, users can obtain the detail information about the lncRNAs such as the expression, max ORF length, coding potential score, neighbor genes and fasta sequence. Moreover, clicking the names of neighbor genes, users will obtain the corresponding genome annotation information. If users want to browse the

information of miRNA including miRNA name, miRNA sequence, 5p/3p class, miRNA length, they can choose the miRNA database and then click the 'Browse data' (Fig. 5b). By clicking the miRNA name, users can obtain the miRNA information such as miRNA length, reads count, confidence, fasta sequence and precursor information. In the search functional section, users can use keywords to search for lncRNA or miRNA in the BmncRNAdb to find the interesting entries. Although some databases (NONCODE, lncRNAdb, lncRBase, deepBase, etc.) also offer data browse for the lncRNAs, the information is mainly for human, mouse, fruit fly, etc. [32–35]. The BmncRNAdb provides not only the information for the silkworm lncRNAs but also for the lncRNAs neighbor genes and the silkworm miRNAs.

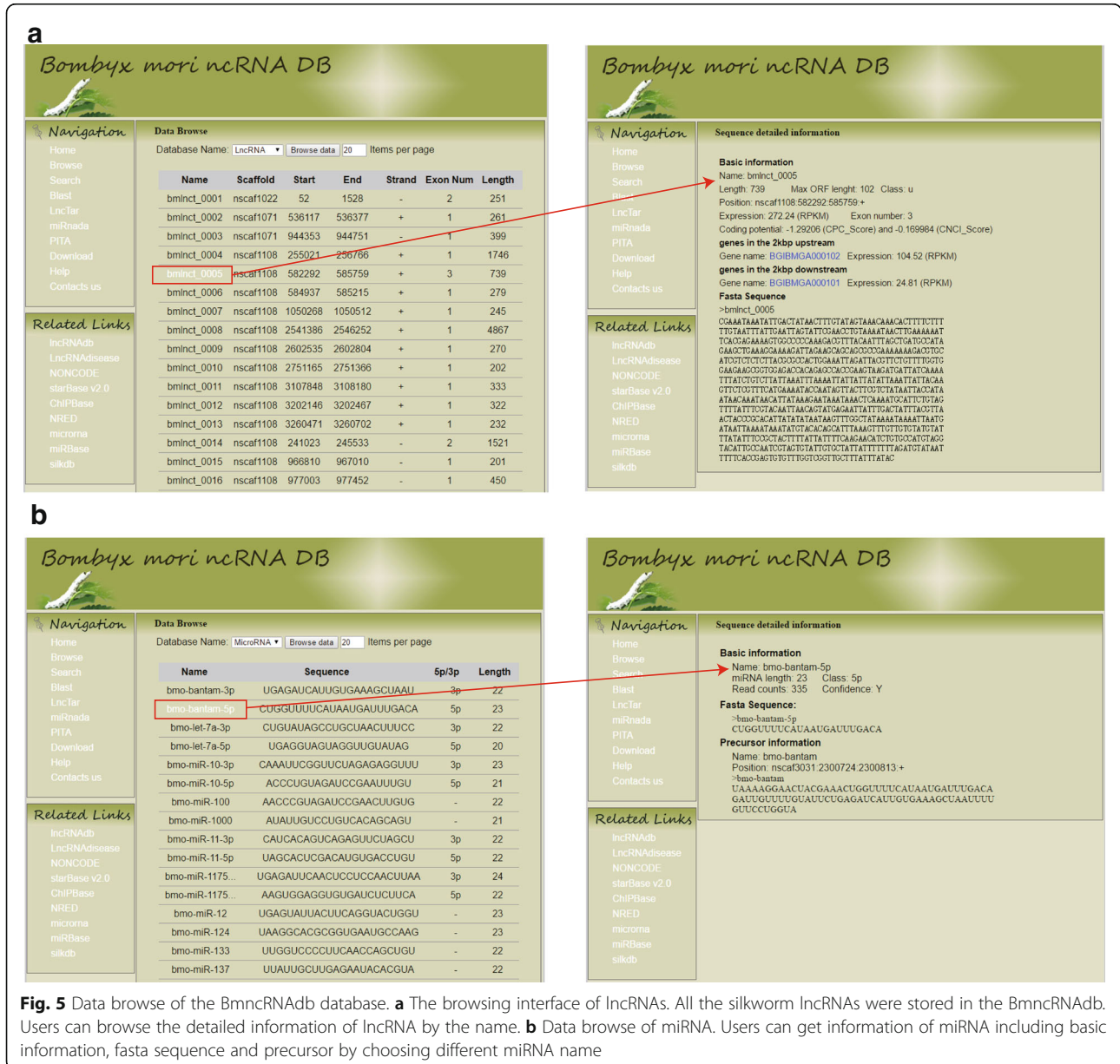


Fig. 5 Data browse of the BmncRNAdb database. **a** The browsing interface of lncRNAs. All the silkworm lncRNAs were stored in the BmncRNAdb. Users can browse the detailed information of lncRNA by the name. **b** Data browse of miRNA. Users can get information of miRNA including basic information, fasta sequence and precursor by choosing different miRNA name

miRanda result. In addition, BmncRNAdb also offers another online tool to find the target genes of a miRNA in the PITA functional section. The usage of PITA is very similar to the miRnada. All the online analysis tools are not only for the silkworm, but also can be used in other species. More help about the online tools is in the help section.

BmncRNAdb offers the download section for users to obtain all the silkworm lncRNA sequences, miRNA sequences and example data. In the help section, a guide manual is shown to help the users to learn how to better use the BmncRNAdb for their own research. In addition, under the left navigation, several useful or famous database resources about ncRNAs are collected in the BmncRNAdb related links. Our group will continue to collect more information on the silkworm ncRNAs and add more useful online tools about the functional research of ncRNAs to the BmncRNAdb in the future.

Conclusions

We have systematically identified and collected 6281 silkworm lncRNAs using the RNA-seq data and unigenes. We also collected 1986 silkworm miRNAs that were predicted by NGS. Integrating these lncRNAs and miRNAs data, we have constructed a comprehensive lncRNAs and miRNAs database (BmncRNAdb) for the silkworm (*Bombyx mori*). Through the BmncRNAdb database, users can browse and search for the detail information of lncRNAs and miRNAs in the silkworm. In addition, this database provides three online tools for users to find the target genes of an lncRNA and miRNA. BmncRNAdb will facilitate the ncRNA research of the silkworm and other insects in the future. Moreover, the availability of the complete set of lncRNAs from the silkworm will improve the comparative and evolutionary analyses of lncRNAs among different Lepidoptera or other insect species.

Availability and requirements

Database: BmncRNAdb

Database homepage: <http://gene.cqu.edu.cn/BmncRNAdb/index.php>

Operating system(s): Linux

Programming language: PHP, CGI, JavaScript, Perl

Other requirements: MySQL, Apache

The database is freely available without restrictions for use by academics and non-commercial researches. Inquiries concerning the database may be directed to zezhang@cqu.edu.cn or huaxia2033@126.com.

Additional file

Additional file 1: Table S1. The detail information of RNA-seq datasets. All the samples used to the identification of the silkworm lncRNAs. (DOC 44 kb)

Abbreviations

Aa: Amino acids; CGI: Common gateway interface; CNCI: Coding-non-coding index; CPC: Coding potential calculator; dG: Binding free energy; EST: Expressed sequence tag; *Fben-1*: Female-brain expressed noncoding RNA-1; lncRNAs: Long non-coding RNAs; miRNAs: microRNAs; NCBI: National Center for Biotechnology Information; ncRNAs: Non-coding RNAs; ndG: Normalized dG; NGS: Next-generation sequencing; Nr: Non-redundant protein; nt: Nucleotide; ORF: Open reading frame; PHP: Personal home page hypertext preprocessor; QC: Quality control; SRA: Sequence read archive; *Xist*: X-inactive specific transcript

Acknowledgements

We would like to thank previous researchers and communities to submit the RNA-seq data of the silkworm to NCBI SRA databases, specifically, thank Drs. Wei Sun, Min-Jin Han and Hong-En Xu for their insightful comments on our manuscript. We are grateful to the editor and anonymous reviewers for their comments and suggestions that have improved our manuscript greatly.

Funding

This work was supported by the National High Technology Research and Development Program of China (2013AA102507-2), the National Natural Science Foundation of China (31272363) and Chongqing Graduate Student Research Innovation Project (CYB14041).

Availability of data and materials

The public website of BmncRNAdb can be accessed at <http://gene.cqu.edu.cn/BmncRNAdb/index.php>. All sequences of the silkworm lncRNAs and miRNAs in fasta format are available in the download page.

Authors' contributions

ZZ and QYY participated in the research design. BZ, QYY and QZZ collected data. BZ, QZZ and QYY analyzed the data and implemented database. ZZ supervised and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

This is not applicable to this study.

Ethics approval and consent to participate

All data used in this study are in the public domain and have previously been published.

Endnotes

This is not applicable to this study.

Author details

¹Laboratory of Evolutionary and Functional Genomics, School of Life Sciences, Chongqing University, Huxi Campus, No. 55 Daxuecheng South Rd., Shapingba, Chongqing 401331, China. ²School of Economics and Business Administration, Chongqing University, Campus A, No. 174 Shazheng Rd., Shapingba, Chongqing 400044, China.

Received: 11 November 2015 Accepted: 8 September 2016

Published online: 13 September 2016

References

- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature*. 2012;489(7414):101–8.
- Serviss JT, Johnsson P, Grander D. An emerging role for long non-coding RNAs in cancer metastasis. *Front Genet*. 2014;5:234.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760–74.
- Carninci P, Hayashizaki Y. Noncoding RNA transcription beyond annotated genes. *Curr Opin Genet Dev*. 2007;17(2):139–44.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1 % of the

- human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799–816.
6. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10(3):155–9.
 7. Illott NE, Ponting CP. Predicting long non-coding RNAs using RNA sequencing. *Methods*. 2013;63(1):50–9.
 8. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, et al. The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. 1992;71(3):527–42.
 9. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7–21.
 10. Xiao H, Yuan Z, Guo D, Hou B, Yin C, Zhang W, et al. Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. *BMC Genomics*. 2015;16(1):749.
 11. Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene*. 2012;31(43):4577–87.
 12. Gibb EA, Vucic EA, Enfield KS, Stewart GL, Lonergan KM, Kennett JY, et al. Human cancer long non-coding RNA transcriptomes. *PLoS One*. 2011;6(10):e25915.
 13. Sun L, Zhang Z, Bailey TL, Perkins AC, Tallack MR, Xu Z, et al. Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse *Klf1* knockout study. *BMC Bioinformatics*. 2012;13:331.
 14. Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu JL, et al. Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biol Evol*. 2012;4(4):427–42.
 15. Ciaudo C, Servant N, Cognat V, Sarazin A, Kieffer E, Viville S, et al. Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genet*. 2009;5(8):e1000620.
 16. Hassan MQ, Tye CE, Stein GS, Lian JB. Non-coding RNAs: epigenetic regulators of bone development and homeostasis. *Bone*. 2015;doi:10.1016/j.bone.2015.05.026.
 17. Martens JA, Laprade L, Winston F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature*. 2004;429(6991):571–4.
 18. Peng X, Gralinski L, Armour CD, Ferris MT, Thomas MJ, Proll S, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *mBio*. 2010;1(5):e00206–10.
 19. Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome. *Annu Rev Genomics Hum Genet*. 2005;6:69–92.
 20. Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L. Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila metamorphosis*. *Sci Rep*. 2016;6:23330.
 21. Jenkins AM, Waterhouse RM, Muskavitch MA. Long non-coding RNA discovery across the genus *Anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. *BMC Genomics*. 2015;16:337.
 22. Etebari K, Furlong MJ, Asgari S. Genome wide discovery of long intergenic non-coding RNAs in Diamondback moth (*Plutella xylostella*) and their expression in insecticide resistant strains. *Sci Rep*. 2015;5:14642.
 23. Stanojic S, Gimenez S, Permal E, Cousserans F, Quesneville H, Fournier P, et al. Correlation of LNCrasiRNAs expression with heterochromatin formation during development of the holocentric insect *Spodoptera frugiperda*. *PLoS One*. 2011;6(9):e24746.
 24. Jayakodi M, Jung JW, Park D, Ahn YJ, Lee SC, Shin SY, et al. Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. *BMC Genomics*. 2015;16:680.
 25. Legeai F, Derrien T. Identification of long non-coding RNAs in insect genomes. *Curr Opin Insect Sci*. 2015;7:37–44.
 26. Rios-Barrera LD, Gutierrez-Perez I, Dominguez M, Riesgo-Escovar R. *acal* is a long Non-coding RNA in JNK signaling in epithelial shape changes during *Drosophila* dorsal closure. *PLoS Genet*. 2015;11(2):e1004927.
 27. Humann FC, Hartfelder K. Representational Difference Analysis (RDA) reveals differential expression of conserved as well as novel genes during caste-specific development of the honey bee (*Apis mellifera* L.) ovary. *Insect Biochem Mol Biol*. 2011;41(8):602–12.
 28. Liu S, Li D, Li Q, Zhao P, Xiang Z, Xia Q. MicroRNAs of *Bombyx mori* identified by Solexa sequencing. *BMC Genomics*. 2010;11:148.
 29. Li J, Cai Y, Ye L, Wang S, Che J, You Z, et al. MicroRNA expression profiling of the fifth-instar posterior silk gland of *Bombyx mori*. *BMC Genomics*. 2014;15:410.
 30. Taguchi S, Iwami M, Kiya T. Identification and characterization of a novel nuclear noncoding RNA, *Fben-1*, which is preferentially expressed in the higher brain center of the female silkworm moth, *Bombyx mori*. *Neurosci Lett*. 2011;496(3):176–80.
 31. Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, et al. Systematic identification and characterization of long Non-coding RNAs in the silkworm, *Bombyx mori*. *PLoS One*. 2016;11(1):e0147147.
 32. Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, et al. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res*. 2014;42(Database issue):D98–103.
 33. Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, et al. lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res*. 2015;43(Database issue):D168–73.
 34. Chakraborty S, Deb A, Maji RK, Saha S, Ghosh Z. lncRBase: an enriched resource for lncRNA information. *PLoS One*. 2014;9(9):e108010.
 35. Zheng LL, Li JH, Wu J, Sun WJ, Liu S, Wang ZL, et al. deepBase v2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res*. 2016;44(D1):D196–202.
 36. Volders PJ, Verheggen K, Menschaert G, Vandepoel K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res*. 2015;43(8):4363–4.
 37. Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS. NRED: a database of long noncoding RNA expression. *Nucleic Acids Res*. 2009;37(Database issue):D122–6.
 38. Park C, Yu N, Choi I, Kim W, Lee S. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics*. 2014;30(17):2480–5.
 39. Jiang Q, Wang J, Wu X, Ma R, Zhang T, Jin S, et al. lncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res*. 2015;43(Database issue):D193–6.
 40. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(Database issue):D68–73.
 41. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 2008;36(Database issue):D149–53.
 42. Nam JW, Bartel DP. Long noncoding RNAs in *C. elegans*. *Genome Res*. 2012;22(12):2529–40.
 43. Zhou ZY, Li AM, Adeola AC, Liu YH, Irwin DM, Xie HB, et al. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol Evol*. 2014;6(6):1387–92.
 44. Chu C, Spitale RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol*. 2015;22(1):29–35.
 45. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol*. 2015;22(1):5–7.
 46. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*. 2012;24(11):4333–45.
 47. Chen J, Quan M, Zhang D. Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq. *Planta*. 2014;241(1):125–43.
 48. Liao Q, Shen J, Liu J, Sun X, Zhao G, Chang Y, et al. Genome-wide identification and functional annotation of *Plasmodium falciparum* long noncoding RNAs from RNA-seq data. *Parasitol Res*. 2014;113(4):1269–81.
 49. Apache Group. Apache HTTP Server. 1997. <http://httpd.apache.org/>. Accessed 9 Jan 2013.
 50. PHP Group. PHP. 2001. <http://php.net/>. Accessed 9 Jan 2013.
 51. Oracle Group. MySQL. 1998. <https://www.mysql.com/>. Accessed 10 May 2012.
 52. Perl Group. Perl. 2002. <https://www.perl.org/>. Accessed 20 Mar 2012.
 53. Duan J, Li R, Cheng D, Fan W, Zha X, Cheng T, et al. SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res*. 2010;38(Database issue):D453–6.
 54. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. *Nucleic Acids Res*. 2015;43(Database issue):D662–9.
 55. Fang SM, Hu BL, Zhou QZ, Yu QY, Zhang Z. Comparative analysis of the silk gland transcriptomes between the domestic and wild silkworms. *BMC Genomics*. 2015;16:60.
 56. Nishida KM, Iwasaki YW, Murota Y, Nagao A, Mannen T, Kato Y, et al. Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in *Bombyx* germ cells. *Cell Rep*. 2015;10(2):193–203.

57. Shao W, Zhao QY, Wang XY, Xu XY, Tang Q, Li M, et al. Alternative splicing and trans-splicing events revealed by analysis of the *Bombyx mori* transcriptome. *RNA*. 2012;18(7):1395–407.
58. Xue J, Qiao N, Zhang W, Cheng RL, Zhang XQ, Bao YY, et al. Dynamic interactions between *Bombyx mori* nucleopolyhedrovirus and its host cells revealed by transcriptome analysis. *J Virol*. 2012;86(13):7345–59.
59. Ma L, Ma Q, Li X, Cheng L, Li K, Li S. Transcriptomic analysis of differentially expressed genes in the *Ras1(CA)*-overexpressed and wildtype posterior silk glands. *BMC Genomics*. 2014;15:182.
60. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010;328(5980):916–9.
61. Gong L, Chen X, Liu C, Jin F, Hu Q. Gene expression profile of *Bombyx mori* hemocyte under the stress of destruxin A. *PLoS One*. 2014;9(5):e96170.
62. Nie H, Liu C, Cheng T, Li Q, Wu Y, Zhou M, et al. Transcriptome analysis of integument differentially expressed genes in the pigment mutant (quail) during molting of silkworm, *Bombyx mori*. *PLoS One*. 2014;9(4):e94185.
63. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*. 2014;509(7502):633–6.
64. Shoji K, Hara K, Kawamoto M, Kiuchi T, Kawaoka S, Sugano S, et al. Silkworm HP1a transcriptionally enhances highly expressed euchromatic genes via association with their transcription start sites. *Nucleic Acids Res*. 2014;42(18):11462–71.
65. Cheng T, Fu B, Wu Y, Long R, Liu C, Xia Q. Transcriptome sequencing and positive selected genes analysis of *Bombyx mandarina*. *PLoS One*. 2015;10(3):e0122837.
66. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2015;43(D1):D6–17.
67. Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, et al. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(D1):D204–12.
68. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(Database issue):D222–30.
69. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–78.
70. Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*. 2014;15(2):R40.
71. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619.
72. Sun L, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e166.
73. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server issue):W345–9.
74. Wang Y, Xue S, Liu X, Liu H, Hu T, Qiu X, et al. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep*. 2016;6:20238.
75. Wang J, Yu W, Yang Y, Li X, Chen T, Liu T, et al. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. *Sci Rep*. 2015;5:16946.
76. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*. 2014;15:311.
77. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–91.
78. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*. 2013;9(4):e1003470.
79. Smalheiser NR, Torvik VI. Complications in mammalian microRNA target prediction. *Methods Mol Biol*. 2006;342:115–27.
80. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
81. Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, et al. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform*. 2015;16(5):806–12.
82. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in *Drosophila*. *Genome Biol*. 2003;5(1):R1.
83. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007;39(10):1278–84.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

