Apidologie (2018) 49:660–670

DOI: 10.1007/s13592-018-0593-z

Original article

First identification of long non-coding RNAs in fungal parasite *Nosema ceranae*

Rui Guo[1], Dafu Chen[1], Cuiling Xiong[1], Chunsheng Hou[2], Yanzhen Zheng[1], Zhongmin Fu[1], Qin Liang[1], Qingyun Diao[2], Lu Zhang[1], Hongquan Wang[1], Zhixian Hou[1], Dhiraj Kumar[3]

[1]College of Bee Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, China
[3]School of Biology and Basic Medical Science, Soochow University, Suzhou 215123, China

Received 4 September 2017 – Revised 10 July 2018 – Accepted 8 August 2018

Abstract – *Nosema ceranae* is a unicellular fungal parasite of honey bees and causes huge losses for apiculture. Until present, no study on *N. ceranae* long non-coding RNAs (lncRNAs) was documented. Here, we sequenced purified spores of *N. ceranae* using strand-specific library construction and high-throughput RNA sequencing technologies. In total, 83 novel lncRNAs were predicted from *N. ceranae* spore samples, including lncRNAs, long intergenic non-coding RNAs (lincRNAs), and sense lncRNAs. Moreover, these lncRNAs share similar characteristics with those identified in mammals and plants, such as shorter length and fewer exon number and transcript isoforms than protein-coding genes. Finally, the expression of 12 lncRNAs was confirmed with RT-PCR, confirming their true existence. To our knowledge, this is the first evidence of lncRNAs produced by a microsporidia species, offering novel insights into basic biology such as regulation of gene expression of this widespread taxonomic group.

RNA-seq / long non-coding RNA / *Nosema ceranae* / honey bee

1. INTRODUCTION

Nosema ceranae is a unicellular fungal parasite that infects midgut epithelial cells of honey bees (Fries et al. 1996). *N. ceranae* spores germinate in response to proper pH and ionic condition in the midgut and inject the infective sporoplasm into the host cells through extruded polar filament. Cell cycle includes stages of meronts, sporonts, and sporoblast (Bailey and Ball 1991). Infected cells are filled with spores during the late stage of infection and the cell may burst to release the spores, which could infect neighboring epithelia cells or be expelled through feces to infect new honeybee individuals (Bailey and Ball 1991; Smith 2012).

N. ceranae was first described from colonies of *Apis cerana* that were sympatric with *Apis mellifera* colonies in China (Fries et al. 1996). However, it is reported that a host switch from *A. cerana* to *A. mellifera* occurred relatively recently (Higes et al. 2006; Huang et al. 2007; Fries 2010). *N. ceranae* is now widespread and can be found on all continents where beekeeping is practiced (Chen et al. 2008; Williams et al. 2008; Giersch et al. 2009; Invernizzi et al. 2009). *N. ceranae* has the potential to dramatically reduce colony strength and productivity (Botías et al. 2012). It can also interact with other environmental stressors weakening colony health (Alaux et al. 2010; Pettis et al. 2012). Some studies showed that a close relationship may

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13592-018-0593-z) contains supplementary material, which is available to authorized users.

Corresponding author: D. Chen, dfchen826@fafu.edu.cn
Rui Guo, Dafu Chen, and Cuiling Xiong contributed equally to this work.
Handling editor: Yves Le Conte

exist between *N. ceranae* and colony collapse disorder (CCD), a disease causing serious honeybee colony losses in the past several years in America, Asia, and Europe (Coxfoster et al. 2007; Bromenshenk et al. 2010).

In the past decade, an increasing number of studies demonstrated that eukaryotic genomes encode a great amount of functional transcripts of non-coding RNAs (ncRNAs) (Wilusz et al. 2009; Yang et al. 2014). ncRNAs are arbitrary classified into two types based on their sizes. One is small RNAs, which are shorter than 200 nucleotides (nt), including but not limited to microRNAs (miRNAs), transfer RNAs (tRNAs), Piwi-interacting RNAs (piRNAs), and small nucleolar RNAs (snoRNAs). The other type is long non-coding RNAs (lncRNAs), which are longer than 200 nt and lack protein-coding potential (Laurent et al. 2015). LncRNA functions by acting on protein-coding genes via the *cis*-acting element and *trans*-acting factor (Zhang et al. 2017). Currently, lncRNAs have been found to play a critical role during biological process including the regulation of gene expression via chromatin modification (Rinn et al. 2007; Gupta et al. 2010) or transcriptional processing (Martianov et al. 2007; Wang et al. 2008), and X-chromosome dosage compensation (Chow et al. 2005), genomic imprinting (Sleutels et al. 2012), epigenetic regulation (Khalil et al. 2009), as well as cell pluripotency and differentiation (Dinger et al. 2008). Transcription from intergenic loci generates long intergenic non-coding RNAs (lincRNAs) (Young et al. 2012). The discovery of lncRNAs was largely impeded due to their low expression levels (Bergmann and Spector 2014). RNA sequencing (RNA-seq) data are very useful resources to identify lncRNAs (Wang et al. 2009). More than 9000 lincRNA genes were discovered in the human genome (Xing et al. 2014; Ganegoda et al. 2015; Xue et al. 2015), and many lncRNAs were also found in mouse (Karlic et al. 2017), sheep (Bakhtiarizadeh et al. 2016), pig (Zhou et al. 2014), chicken (Li et al. 2012), *Anopheles gambiae* (Jenkins et al. 2015), and *Drosophila melanogaster* (Young et al. 2012). However, very little is known about ncRNAs in microsporidia including *N. ceranae* at present. The genome of *N. ceranae* (Cornman et al. 2009) was sequenced and published in 2009, which greatly facilitates molecular studies of this significant fungal parasite. Recently, Huang and Evans (2016) identified six microRNA-like small RNAs from *N. ceranae* by using deep sequencing and RT-qPCR methods. In the current study, we sequenced *N. ceranae* spores using Illumina sequencing technology, identified 83 lncRNAs associated with this important fungal pathogen of honey bee, and further confirmed the existence of 12 lncRNAs via RT-PCR.

2. MATERIAL AND METHODS

2.1. *N. ceranae* spore purification

Apis mellifera ligustica workers infected with *N. ceranae* were selected from an apiary in Fuzhou, Fujian province, China. The infected honey bees were first kept in – 20 °C for 20 min to kill them. Subsequently, fresh spores of *N. ceranae* were isolated and purified following the method described by Higes et al. (2007) and Cornman et al. (2009) with some modifications. Briefly, midguts of infected workers were removed and crushed in sterile water and filtered with four-layer gauze to remove tissue debris. The filtered suspension was centrifuged at 8000 rpm for 5 min and the supernatant was discarded. The re-suspended pellet was further purified on a discontinuous Percoll gradient (Solarbio, China) consisting of 5 mL each of 25, 50, 75, and 100% Percoll solution. The spore suspension was overlaid onto the gradient and centrifuged at 14,000 rpm for 90 min at 4 °C. The spore pellet was carefully extracted with a syringe followed by centrifugation again on a discontinuous Percoll gradient to obtain clean spores. Following the above-mentioned method, three spore samples termed as Nc-1, Nc-2, and Nc-3 were prepared. A bit of spores were subjected to PCR identification and confirmed to be mono-specific using previously described primers (Chen et al. 2008). The cleaned spores were frozen in liquid nitrogen and stored at – 80 °C until RNA-seq and RT-PCR.

2.2. Deep sequencing

Total RNA of *N. ceranae* spore samples was extracted using Trizol (Life Technologies Inc., USA) according to the manufacturer's instructions, and checked via 1% agarose gel

eletrophoresis. Subsequently, rRNAs were removed using Ribo-Zero™ kit (Epicenter, USA) to retain mRNAs and ncRNAs. The enriched mRNAs and ncRNAs were fragmented into short fragments by using divalent cations under elevated temperature in NEBNext reaction buffer (NEB, USA) and reverse transcripted into cDNA with random primers. Second-strand cDNA were synthesized by DNA polymerase I, RNase H, dNTP (dUTP instead of dTTP), and second-strand synthesis reaction buffer (NEB, USA). Next, the cDNA fragments were purified with QiaQuick PCR extraction kit (QIAGEN, Germany), end repaired, poly (A) added, and ligated to Illumina sequencing adapters. Then UNG (Uracil-*N*-glycosylase) was used to digest the second-strand cDNA. The digested products were size selected by agarose gel electrophoresis, PCR amplified, and sequenced using Illumina HiSeq Xten by Biomarker Technology Co. (Beijing, China). Raw sequencing data have been uploaded to the National Centre for Biotechnology Information (NCBI) as Bioproject PRJNA395264.

2.3. Long non-coding RNA analysis

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. After removing reads containing adapter, reads containing ploy-N and low-quality reads from raw data, clean data (clean reads) were obtained. Meanwhile, Q20, Q30, GC-content, and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

The transcriptome was assembled using the StringTie based on the reads mapped to the reference genome of *N. ceranae* (https://www.ncbi.nlm.nih.gov/genome/?term=nosema%20ceranae). The assembled transcripts were annotated using the gffcompare program. The unknown transcripts were used to screen for putative lncRNAs. Four computational approaches include CNCI (Sun et al. 2013), CPC (Kong et al. 2007), Pfam (Finn et al. 2013), and CAPT (Wang et al. 2013) were combined to sort non-protein-coding RNA candidates from putative protein-coding RNAs in the unknown transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts with lengths more than 200 nt and have more than two exons were selected as lncRNA candidates and further screened using the above-mentioned four software that have the power to distinguish the protein-coding genes from the non-coding genes. The different types of lncRNAs including lincRNA, intronic lncRNA, and antisense lncRNA were selected using cuffcompare. The detailed flow of novel lncRNA prediction is presented in Figure 1. StringTie (1.3.1) was used to calculate FPKMs (fragments per kilobase of transcript per million mapped reads) of both lncRNAs and protein-coding genes in each sample. Gene FPKMs were computed by summing the FPKMs of transcripts in each gene group. FPKM means fragments per kilo base of exon per million fragments mapped, calculated based on the length of the fragments and reads count mapped to this fragment.

2.4. RT-PCR validation

Total RNA of the above-mentioned purified *N. ceranae* spores was isolated using the AxyPre RNA extraction kit (Axygen, America), according to the manufacturer's instructions. Total RNA recovered was immediately used to generate first-strand cDNA using the PrimeScript cDNA synthesis kit (TAKARA, China), according to the manufacturer's instructions. Resultant cDNA synthesized were stored at – 20 °C.

To verify putative lncRNAs in *N. ceranae* spores, 13 lncRNAs were selected for RT-PCR assay. On the basis of sequences of lncRNAs MSTRG.3360.2, MSTRG.5334.1, MSTRG.5107.2, MSTRG.3360.1, MSTRG.4106.1, MSTRG.2986.2, MSTRG.1802.2, MSTRG.6512.1, MSTRG.5015.1, MSTRG.6654.2, MSTRG.3636.1, MSTRG.3946.1, and MSTRG.3946.2, the specific primers were designed using DNAMAN (Lynnon Biosoft, America) and synthesized by Sangon Biotech (Shanghai, China) Co., Ltd. The primer sequences with successful amplification were shown in Table I. PCR was carried out on an T100 thermo cycler (BIO-RAD, America) using Premix (TAKARA, China) following the manufacturer's protocol under the conditions: pre-denaturation step at 94 °C for 5 min; 30 amplification cycles of denaturation at 94 °C for 50 s, annealing at 55 °C for 30 s, and

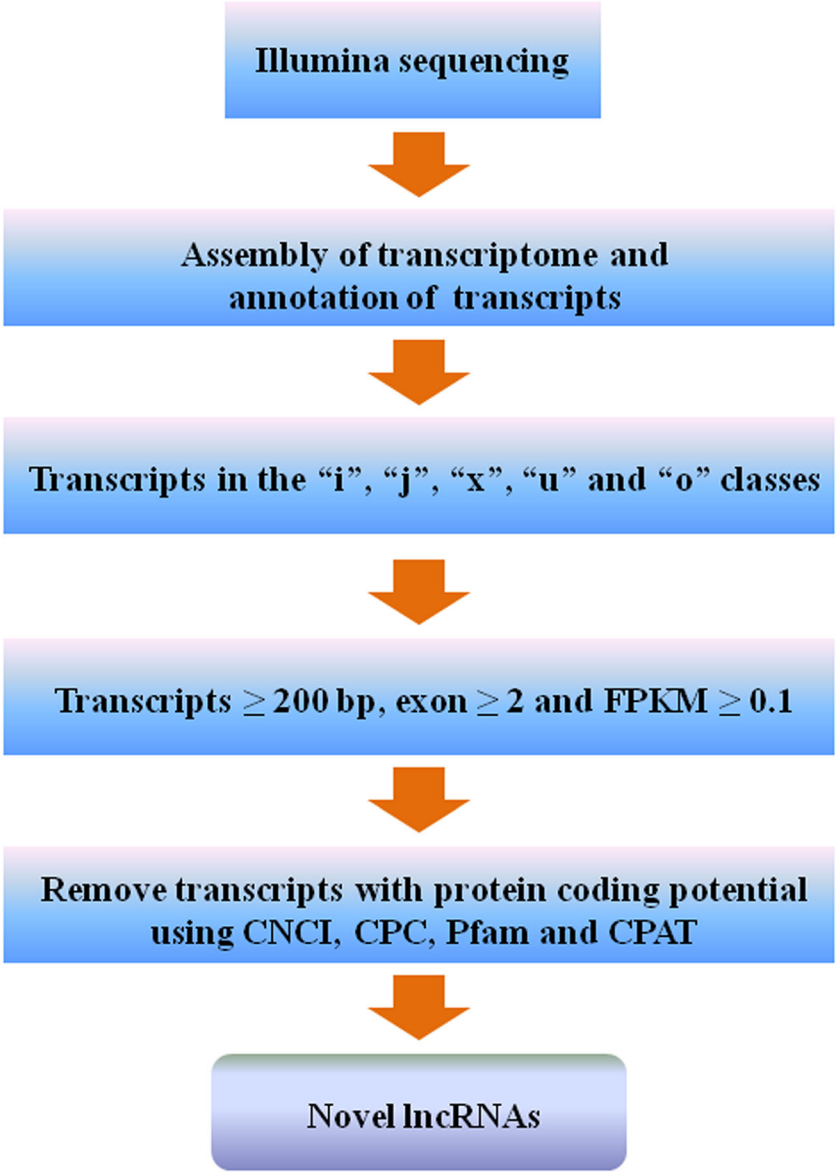

Figure 1 The pipeline of novel lncRNA prediction**.** The purified *N. ceranae* spore samples were sequenced using Illumna HiSeq technology. Raw reads were filtered to gain clean reads, which were used to assemble the transcriptome. The assembled transcripts were annotated and the unknown ones were subjected to lncRNAs screening. Transcripts with lengths more than 200 nt and FPKM ≥ 0.1 and have more than two exons were selected as lncRNA candidates and further screened using a combination of CNCI, CPC, Pfam, and CAPT software. Finally, the various types of lncRNAs including lincRNA, sense lncRNA, and antisense lncRNA were distinguished using cuffcompare. "i" class is a union of transcripts within reference transcripts' intron domain. "j" class is a union of potential novel transcripts or transcript-derived fragments. "x" class is a union of transcripts owning intersection between transcripts' antisense strands. "u" class is a union of transcripts within reference transcripts' intergenic region. "o" class is a union of transcripts owning some intersection with reference transcripts' exons.

Table I. Primers used in this study

Primer name	Primer sequence	Expected size of the PCR product (bp)
MSTRG.3360.2-F	CACGAAAGAGTGGCAACCT	235
MSTRG.3360.2-R	AGTTCCTCCTCACATTCCAA	
MSTRG.5334.1-F	CCATAAAGACTAAGACCTGCG	283
MSTRG.5334.1-R	TGGAACAACAAGGGAACAC	
MSTRG.5107.2-F	GCTCCAATGGCAAGTTTG	762
MSTRG.5107.2-R	CGAAGGAATGGGACAAGTA	
MSTRG.3360.1-F	ACACAAAGCACGAAAGAGTG	243
MSTRG.3360.1-R	AGTTCCTCCTCACATTCCAA	
MSTRG.4106.1-F	TTGCGTTGTCCTGCTGAA	463
MSTRG.4106.1-R	CGTTCCTAACCTATCATCATCG	
MSTRG.2986.2-F	GGGCGGTTGTTATGGAAGT	667
MSTRG.2986.2-R	GCCGTTTACATTCGCTTACAC	
MSTRG.1802.2-F	CAGATGTATTGATGGGCTCC	152
MSTRG.1802.2-R	AAGGTGGCGACCGATTAT	
MSTRG.5015.1-F	GGCTGCCACCAATCAGTTA	774
MSTRG.5015.1-R	CGGCTACTACAAACGAAACG	
MSTRG.6654.2-F	GCAACAACTTCTGGTGTCTCT	143
MSTRG.6654.2-R	GCTCTTATCATTTCTGGCACAG	
MSTRG.3636.1-F	AGCCTACTCTATCCCGTTTG	85
MSTRG.3636.1-R	CGCACAAGCCTAAAGACAT	
MSTRG.3946.1-F	CGAAATGTTCAAGGCTGC	738
MSTRG.3946.1-R	TGATGCTGTGGCTAAGACA	
MSTRG.3946.2-F	CGAAATGTTCAAGGCTGC	711
MSTRG.3946.2-R	TGATGCTGTGGCTAAGACA	

elongation at 72 °C for 1 min; followed by a final elongation step at 72 °C for 10 min. The PCR products were subjected to 1.5% agarose gel and Genecolor (Gene-Bio, China) staining.

3. RESULTS

In this study, total RNA of purified *N. ceranae* spores (Figure 2a) was extracted and checked via

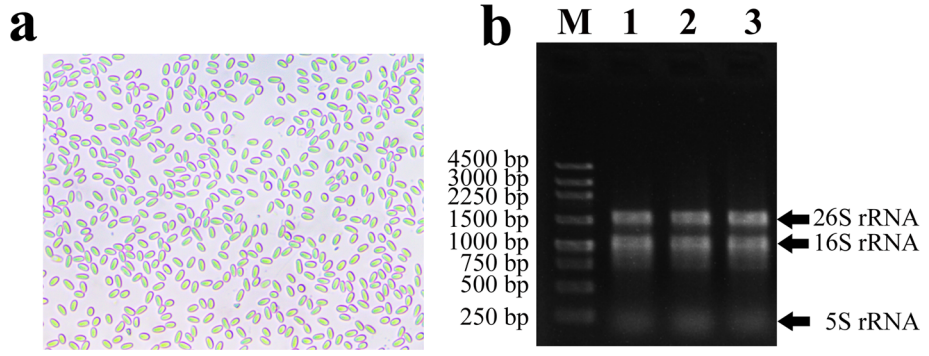

Figure 2 Agarose gel electrophoresis of the extracted RNA from purified *N. ceranae* spores. Spores of *N. ceranae* were purified and examined under optical microscope (× 400) (**a**). The total RNA of *N. ceranae* spores was extracted and analyzed via agarose gel electrophoresis (**b**). Lane 1 total RNA of Nc-1 sample. Lane 2 total RNA of Nc-2 sample. Lane 3 total RNA of Nc-3 sample. Lane M 250 bp DNA ladder.

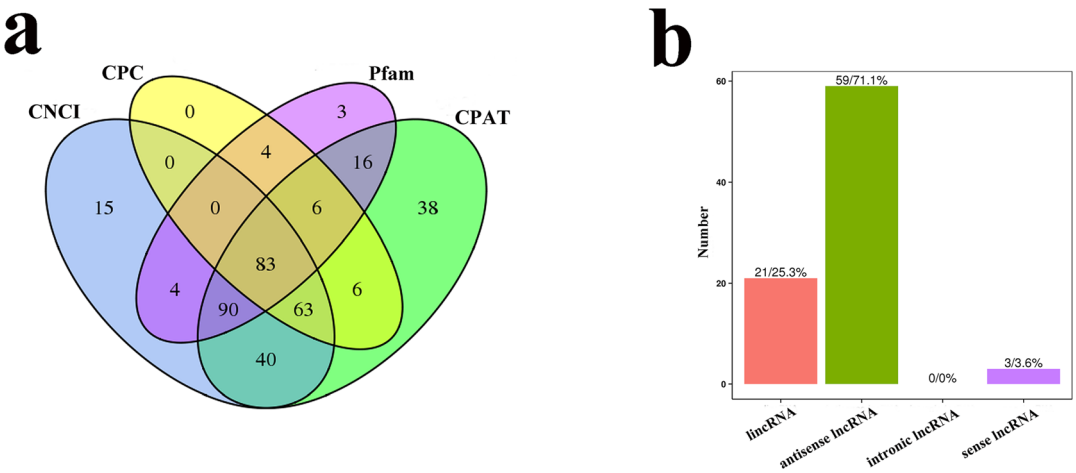

Figure 3 Prediction of novel lncRNA of *N. ceranae*. **a** Venn analysis of the predicted novel lncRNAs from CNCI, CPC, Pfam, and CAPT software. **b** Classification of the predicted novel lncRNAs.

agarose gel electrophoresis, the result showed clear bands of 26S, 16S, and 5S rRNAs (Figure 2b). In total, over 218.47 million raw reads were produced from three cDNA libraries of *N. ceranae* spores after removing adaptor sequences and low-quality reads (Table SI). For Nc-1, Nc-2, and Nc-3, 107,123,113 (79.75%), 117,688,499 (78.84%), and 116,776,007 (76.15%) reads were mapped to the reference genome of *N. ceranae*, respectively (Table SII). A total of 83 lncRNAs were predicted with CNCI, CPC, Pfam, and CAPT (Figure 3a), of which 59, 21, and 3 were antisense lncRNAs, lincRNAs, and sense lncRNAs, respectively (Figure 3b). Intriguingly, no intronic lncRNA was observed.

LncRNAs were found to be shorter in length than protein-coding genes in *N. ceranae* due to the lower number of exons. About 60.24% of *N. ceranae* lncRNAs have a length of 400 bp, which is nearly three times the ratio of 19.63% detected in protein-coding genes (Figure 4a, b). Consistent with their counterparts in the mammals, *N. ceranae* lncRNAs had fewer exons than protein-coding genes. Approximately 85.54% of *N. ceranae* lncRNAs had only two exons, which is much higher than those (3.87%) observed in protein-coding genes (Figure 4c, d). Then, 9.62% of lncRNA open reading frames (ORFs) have a length of 100 bp, which is higher than the ratio (2.14%) found in protein-coding genes (Figure 4e, f). Interestingly, 88.49% of total ORFs of lncRNAs have a length of 50bp (Figure 4e). In addition, lncRNA loci possess fewer transcript isoforms than protein-coding mRNA loci (Figure 4g), implying that lncRNAs are less complex than protein-coding genes. Moreover, the expression level of each transcript was estimated and the result showed the levels of lncRNAs are lower than those of mRNAs (Figure 4h).

To verify the reliability of the predicted lncRNAs, we randomly selected 13 lncRNAs (included in the total predicted 83) for RT-PCR validation, and the result displayed signal bands could be successfully amplified from 12 lncRNAs (Figure 5), which indicated that a high percentage of lncRNAs identified in our study are reliable in terms of expression.

4. DISCUSSION

Though transcriptomic studies on honey bees in response to *N. ceranae* challenge were previously reported (Aufauvre et al., 2014; Badaoui et al., 2017), however, omics researches on *N. ceranae* are extremely limited. Thanks to the development of high-throughput techniques, thousands of lncRNA genes were identified from a variety of species (Lee et al. 2012; Jalali et al. 2015). In recent years, more and more studies indicated that lncRNAs play key roles in biological processes such as the regulation of gene expression (Martianov et al. 2007; Rinn et al. 2007; Wang et al. 2008; Gupta et al. 2010). To the best of our knowledge, no study on microsporidia lncRNAs

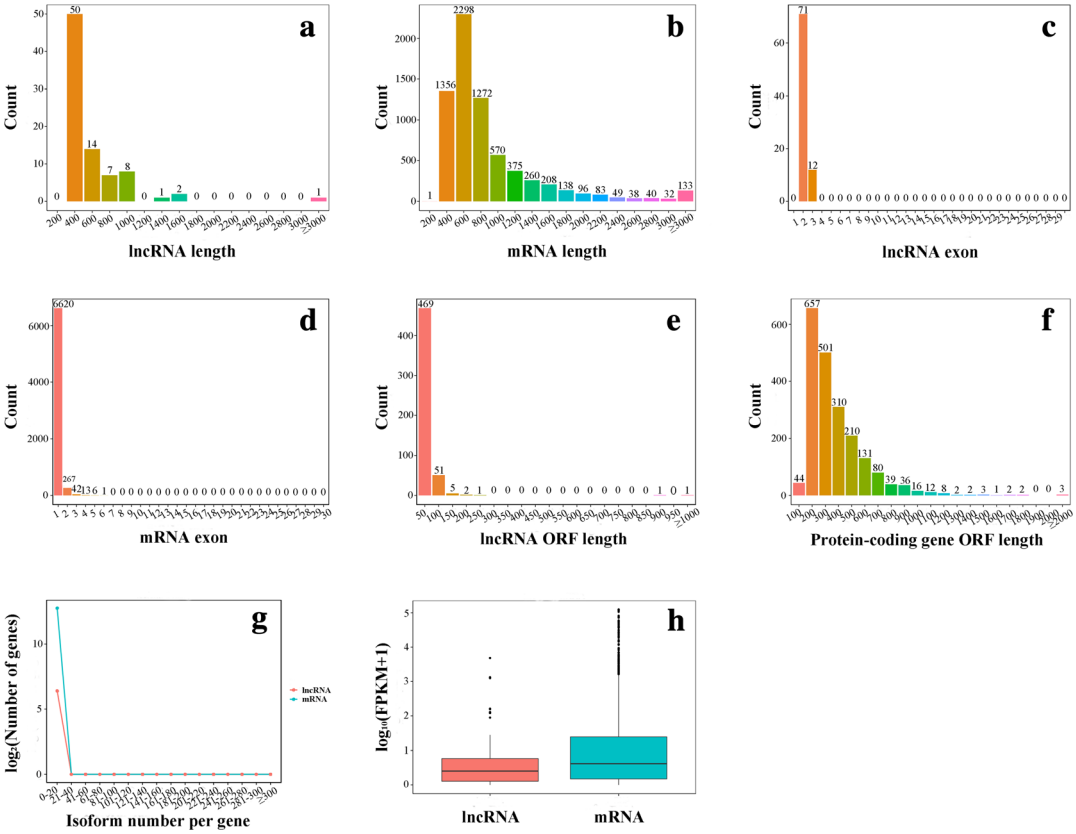

Figure 4 Structural feature and expression analysis of *N. ceranae* lncRNAs. **a** The length of lncRNAs. **b** The length of protein coding genes. **c** The exon number of lncRNAs. **d** The exon number of protein coding genes. **e** The ORF length of lncRNAs. **f** The ORF length of protein coding genes. **g** The alternatively spliced isoforms of lncRNAs and protein coding genes. **h** The expression level of lncRNAs and protein coding genes.

was reported until now. LncRNAs' function in microsporida including *Nosema* is largely unknown at present. In the present study, in order to investigate lncRNAs in *N. ceranae*, purified spores of *N. ceranae* were sequenced using RNA-seq technology. Based on a rigorous filtering criteria, a total of 83 lncRNAs including antisense lncRNAs, lincRNAs, and sense lncRNAs were predicted, bridging the gap of *N. ceranae* lncRNA. These lncRNAs share similar characteristics including

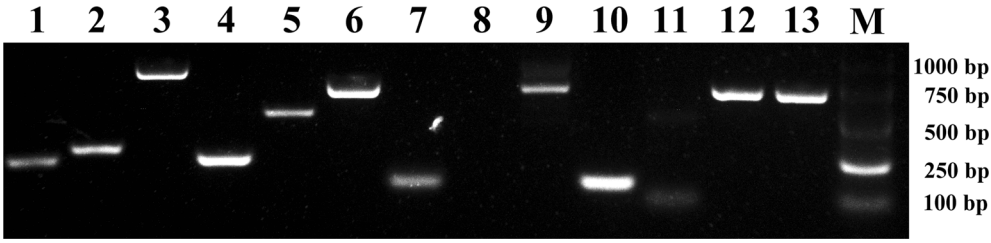

Figure 5 RT-PCR validation of *N. ceranae* lncRNAs. Lane 1–13 different *N. ceranae* lncRNAs selected for molecular identification. Lane 1 MSTRG.3360.2. Lane 2 MSTRG.5334.1. Lane 3 MSTRG.5107.2. Lane 4 MSTRG.3360.1. Lane 5 MSTRG.4106.1. Lane 6 MSTRG.2986.2. Lane 7 MSTRG.1802.2. Lane 9 MSTRG.5015.1. Lane 10 MSTRG.6654.2. Lane 11 MSTRG.3636.1. Lane 12 MSTRG.3946.1. Lane 13 MSTRG.3946.2. The specific PCR primers are displayed in Table I.

shorter length and fewer exon number than protein-coding genes with those identified in other species (Trapnell et al. 2010; Xiao et al. 2015; Wang et al. 2016). Furthermore, we conducted RT-PCR assays for 13 randomly selected lncRNAs and the expression of 12 lncRNAs was verified from *N. ceranae* spore samples, suggesting the true existence of majority of predicted lncRNAs. *N. ceranae* lncRNAs identified in this study can provide some reference information of importance for other microsporidia species. To survive outside of the host, microsporidia form environmentally resistant spores that are protected by a thick two-layered wall, and are capable of maintaining metabolic activity (Vavra and Larsson 1999). *Encephalitozoon cuniculi* and *Antonospora locustae* spores have been shown to possess transcripts different from prokaryotic operons (Williams et al. 2005. Corradi et al. 2008). Liu et al. (2016) performed sequencing and analysis of ungerminated and germinated spores of *Nosema bombycis* using RNA-seq, respectively assembled 2756 and 2690 unigenes based on more than 60 million transcript reads. In our study, over 218.47 million raw reads were generated from three cDNA libraries of *N. ceranae* spores, showing transcriptional signals from dormant spores. Once coming into contact with a host cell or triggered by appropriate stimuli, the spores rapidly evert the polar tube, through which the infective sporoplasm is transferred into the host cytoplasm after penetrating the host cell membrane, and following entering the host cell, the parasite begins to divide and proliferate (moronic stage) (Schottelius et al. 2000). Because microsporidia spores are resistance stages with very low cellular activity, it is inferred that the number of lncRNAs in *N. ceranae* meront is more than that in *N. ceranae* spore. One way to identify lncRNAs in *N. ceranae* meront is to carry out deep sequencing of *N. ceranae*-infected honey bee gut and gain the transcriptome of *N. ceranae* via filtering out host data.

Previously, DiMaria et al. (1996) characterized an RNA homologous to U2 RNA and a single copy gene encoding the RNA homolog in *Vairimorpha necatrix*, which is the first report of ncRNAs identified in a microsporidia species. Fast's group first isolated and sequenced the genes encoding both U6 and U2 snRNAs from the intracellularly parasitic microsporidia *Nosema locustae*, and further found both genes are expressed and both RNAs are capable of being folded into secondary structures typical of other known U6 and U2 snRNAs, demonstrating that U6 and U2 snRNAs are likely to be functional components of an active splicesome in *N. locustae* (Fast et al., 1998). More recently, Belkorchia and colleagues developed a dedicated in silico approach based on genome prediction and transcriptome validation to survey the presence of ncRNAs in the genomes of four *Encephalitozoon* species *E. cuniculi*, *Encephalitozoon intestinalis*, *Encephalitozoon hellem*, and *Encephalitozoon romaleae*, and identified ten novel ncRNAs in the *Encephalitozoon* spp., including nine snoRNA-like genes and one U1 snRNA (Belkorchia et al. 2017). They further investigated the genomes of three *Nosema* species *N. ceranae*, *Nosema apis*, and *N. bombycis*, and identified most of the above-mentioned ncRNAs in *Encephalitozoon* spp. (Belkorchia et al. 2017). However, the number of known ncRNAs in micrsporidia is very limited, and no report of lncRNAs in microsporidia was documented until now. In the current study, for the first time, a total of 83 lncRNAs including lincRNAs, sense lncRNAs, and antisense lncRNAs were detected in *N. ceranae* spore, which enlarge the reservoir of ncRNAs in microsporidia. Whether these lncRNAs are shared with other microsporidia species needs further investigation. Microsporidia genome is highly reduced and compact (Peyretaillade et al. 2011; Corradi et al. 2010), and it is also found intron poor (Lee et al. 2010; Peyretaillade et al. 2012). It is predicted that only 16 introns are within 15 genes in *E. cuniculi* genome (Katinka et al. 2001. Vivares et al. 2002), and there is no intron at all in the genome of *Enterocytozoon bieneusi* (Akiyoshi et al. 2009). *N. ceranae* genome has a size of 7.86 M, which is predicted to have five tRNA genes containing introns and six genes with short introns (Cornman et al. 2009). Here, no lncRNA generated from introns was found using bioinformatics pipeline, which may partly due to very limited introns in *N. ceranae* genome.

Taken together, using next-generation sequencing technology and bioinformatic method, lncRNAs in *N. ceranae* spore were predicted, analyzed, and identified in the current study. This is

the first evidence of lncRNAs generated in *N. ceranae*; our findings provide novel insights into basic biology involving this widespread fungal parasite of honeybee. In the future, more studies are required to clarify the role of *N. ceranae* lncRNAs in gene expression regulation and pathogenicity.

ACKNOWLEDGEMENTS

We thank L Li, B Li, TH Yan, HY Zhao, TY Zhuang, and HZ Chen for their assistance in spore sample preparation, and thank all editors and reviewers for their invaluable comments that improved our work.

AUTHOR CONTRIBUTIONS

DF Chen, R Guo, and Q Liang contributed to the conceptualization and design of this study. R Guo, CL Xiong, CS Hou, and QY Diao performed data analysis. YZ Zhen, ZM Fu, L Zhang, HQ Wang, ZX Hou, and D Kumar carried out experiments. DF Chen and R Guo contributed to writing of the manuscript. R Guo and D Kumar performed language polish.

FUNDING

This work was funded by the Earmarked Fund for Modern Agro-industry Technology Research System (CARS-44-KXJ7), the Science and Technology Planning Project of Fujian Province (2018J05042), the Teaching and Scientific Research Fund of Education Department of Fujian Province (JAT170158), the Science and Technology Innovation Fund of Fujian Agriculture and Forestry University (CXZX2017342), and the Open Fund of Key Laboratory of Pollinating Insect Biology of Ministry of Agriculture and Rural Affairs (2017MFNZS02).

Première identification de longs ARNs non codants dans le parasite *Nosema ceranae*

ARN-seq/long ARN non codant / *Nosema ceranae* / abeille

Erste Identifizierung von langen non-coding RNA's beim Bienenparasiten *Nosema ceranae*

RNA-seq/lange non-coding RNA / *Nosema ceranae* / Honigbiene

REFERENCES

Akiyoshi, D.E., Morrison, H.G., Lei, S., Feng, X., Zhang, Q., et al. (2009) Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. PLoS Pathog. 5(1), e1000261.

Alaux, C., Brunet, J.L., Dussaubat, C., Mondet, F., Tchamitchan, S., et al. (2010) Interactions between *Nosema* microspores and a neonicotinoid weaken honeybees (*Apis mellifera*). Environ. Microbiol. 12(3), 774–782.

Aufauvre, J., Mismeaucouturier, B., Viguès, B., Texier, C., Delbac, F., et al. (2014) Transcriptome analyses of the honeybee response to *Nosema ceranae* and insecticides. PLoS One 9(3), e91686.

Badaoui, B., Fougeroux, A., Petit, F., Anselmo, A., Gorni, C., et al. (2017) RNA-sequence analysis of gene expression from honeybees (*Apis mellifera*) infected with *Nosema ceranae*. PLoS One 12(3), e0173438.

Bailey, L., Ball, B.V. (1991) Honey bee pathology. 2nd edition. London: Academic Press.

Bakhtiarizadeh, M.R., Hosseinpour, B., Arefnezhad, B., Shamabadi, N., Salami, S.A. (2016) *In silico* prediction of long intergenic non-coding RNAs in sheep. Genome 59(4), 263.

Belkorchia, A., Pombert, J.F., Polonais, V., Parisot, N., Delbac, F., et al. 2017 Comparative genomics of microsporidian genomes reveals a minimal non-coding RNA set and new insights for transcription in minimal eukaryotic genomes. DNA Res. 24(3), 251.

Bergmann, J.H., Spector, D.L. (2014) Long non-coding RNAs: modulators of nuclear structure and function. Curr. Opin. Cell Biol. 26(1), 10.

Botías, C., Martín-Hernández, R., Días, J., García-Palencia, P., Matabuena, M., et al. (2012) The effect of induced queen replacement on *Nosema* spp. infection in honey bee (*Apis mellifera iberiensis*) colonies. Environ. Microbiol. 14(4), 845–859.

Bromenshenk, J.J., Henderson, C.B., Wick, C.H., Stanford, M.F., Zulich, A.W., et al. (2010) Iridovirus and microsporidian linked to honey bee colony decline. PLoS One 5(10), e13181.

Chen, Y., Evans, J.D., Smith, I.B., Pettis, J.S. (2008) *Nosema ceranae* is a long-present and wide-spread microsporidian infection of the European honey bee (*Apis mellifera*) in the United States. J. Invertebr. Pathol. 97(2), 186–188.

Chow, J.C., Yen, Z., Ziesche, S.M., Brown, C.J. (2005) Silencing of the mammalian X chromosome. Annu. Rev. Genomics Hum. Genet. 6, 69–92

Cornman, R.S., Chen, Y.P., Schatz, M.C., Street, C., Zhao, Y., et al. (2009) Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. PLoS Pathog. 5(6), e1000466.

Corradi, N., Gangaeva, A., Keeling, P.J. (2008) Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora*

locustae and *Encephalitozoon cuniculi*. Genomics 91, 388–393.

Corradi, N., Pombert, J.F., Farinelli, L., Didier, E.S., Keeling, P.J. (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. Nat. Commun. 1, 77.

Coxfoster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318(5848), 283–287.

Dimaria, P., Palic, B., Debrunnervossbrinck, B.A., Lapp, J., Vossbrinck, C.R. (1996) Characterization of the highly divergent U2 RNA homolog in the microsporidian *Vairimorpha necatrix*. Nucleic Acids Res. 24(3), 515.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., et al. (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 18, 1433–1445.

Fast, N.M., Roger, A.J., Richardson, C.A., Doolittle, W.F. (1998) U2 and U6 snRNA genes in the microsporidian *Nosema locustae*: evidence for a functional spliceosome. Nucleic Acids Res. 26(13):3202–3207.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., et al. (2013) Pfam: the protein families database. Nucleic Acids Res. 42(Database issue), 222–30.

Fries, I. (2010) *Nosema ceranae* in European honey bees (*Apis mellifera*). J. Invertebr. Pathol. 103(1), S73-S79.

Fries, I., Feng, F., Silva, A.D., Slemenda, S.B., Pieniazek, N.J. (1996) *Nosema ceranae* n. sp. (Microspora, Nosematidae), morphological and molecular characterization of a microsporidian parasite of the Asian honey bee *Apis cerana* (Hymenoptera, Apidae). Eur. JProtistol. 32(3), 356–365.

Ganegoda, G.U., Li, M., Wang, W., Feng, Q. (2015) Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations. IEEE Trans Nanobioscience 14(2), 175–183.

Giersch, T., Berg, T., Galea, F., Hornitzky, M. (2009) *Nosema ceranae* infects honey bees (*Apis mellifera*) and contaminates honey in Australia. Apidologie 40(2), 117–123.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., et al. (2010) Long noncoding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464,1071–1076.

Higes, M., Martin-Hernandez, R., Meana, A. (2006) *Nosema ceranae*, a new microsporidian parasite in honeybees in Europe. J. Invertebr. Pathol. 92, 93–95.

Higes, M., Garcíapalencia, P., Martínhernández, R., Meana, A. (2007) Experimental infection of *Apis mellifera* honeybees with *Nosema ceranae* (Microsporidia). J. Invertebr. Pathol. 94(3), 211–217.

Huang, Q., Evans, J.D. (2016) Identification of microRNA-like small RNAs from fungal parasite *Nosema ceranae*. J. Invertebr. Pathol. 133, 107–109.

Huang, W.F., Jiang, J.H., Chen, Y.W., Wang, C.H. (2007) A *Nosema ceranae* isolate from the honeybee *Apis mellifera*. Apidologie 38, 30–37.

Invernizzi, C., Abud, C., Tomasco, I.H., Harriet, J., Ramallo, G., et al. (2009) Presence of *Nosema ceranae* in honeybees (*Apis mellifera*) in Uruguay. J. Invertebr. Pathol. 101(2), 150–153.

Jalali, S., Kapoor, S., Sivadas, A., Bhartiya, D., Scaria, V. (2015) Computational approaches towards understanding human long non-coding RNA biology. Bioinformatics 31(14), 2241–2251.

Jenkins, A.M., Waterhouse, R.M., Muskavitch, M.A. (2015) Long non-coding RNA discovery across the genus anopheles reveals conserved secondary structures within and beyond the Gambiae complex. BMC Genomics 16(1), 1–14.

Karlic, R., Ganesh, S., Franke, V., Svobodova, E., Urbanova, J., Suzuki, Y., et al. (2017) Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. DNA Res. 24(2), 2019–2220.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414, 450–453.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA. 106, 11667–11672.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., et al. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35(Web Server issue), W345-W349.

Laurent, G.S., Wahlestedt, C., Kapranov, P. (2015) The landscape of long noncoding RNA classification. Trends Genet. 31(5), 239–251.

Lee, R.C.H., Gill, E.E., Roy, S. W, Fast, N.M. (2010) Constrained intron structures in a microsporidian. Mol. Biol. Evol. 27(9), 1979–1982.

Lee, T.L., Xiao, A., Rennert, O.M. (2012) Identification of novel long noncoding RNA transcripts in male germ cells. Methods Mol. Biol. 825, 105–114.

Li, T., Wang, S., Wu, R., Zhou, X., Zhu, D., et al. (2012) Identification of long nonprotein coding RNAs in chicken skeletal muscle using next generation sequencing. Genomics 99(5), 292–298.

Liu, H., Li, M., He, X., Cai, S., He, X., et al. (2016) Transcriptome sequencing and characterization of ungerminated and germinated spores of *Nosema bombycis*. Acta Biochim. Biophys. Sin. 48(3), 246.

Martianov, I., Ramadass, A., Serra, B.A., Chow, N., Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445, 666–670.

Pettis, J.S., vanEngelsdorp, D., Johnson, J., Dively, G. (2012) Pesticide exposure in honey bees results in increased levels of the gut pathogen *Nosema*. Naturwissenschaften 99(2), 153–158.

Peyretaillade, E., El, A. H, Diogon, M., Polonais, V., Parisot, N., et al. (2011) Extreme reduction and compaction of microsporidian genomes. Res. Microbiol. 162(6), 598.

Peyretaillade, E., Parisot, N., Polonais, V., Terrat, S., Denonfoux, J., et al. (2012) Annotation of microsporidian genomes using transcriptional signals. Nature Commun. 3(4), 1137.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7), 1311–1323.

Schottelius, J., Schmetz, C., Kock, N.P., Schuler, T., Sobottka, I., et al. (2000) Presentation by scanning electron microscopy of the life cycle of microsporidia of the genus *Encephalitozoon*. Microbes Infect. 2, 1401–1406.

Sleutels, F., Zwart, R., Barlow, D.P.. (2012) The noncoding Air RNA is required for silencing autosomal imprinted genes. Nature 415(6873), 810–813.

Smith, M.L. (2012) The honey bee parasite *Nosema ceranae*: transmissible via food exchange? PLoS One 7(8), e43319.

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., et al. (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 41(17), e166.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G. et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28(5), 511–515.

Vavra, J., and Larsson, J. I. R. (1999). Structure of the microsporidia, in Wittner, M., Weiss, L.M. (Eds.) The Microsporidia and Microsporidiosis. Washington, DC: ASM Press, 7–84.

Vivares, C.P., Gouy, M., Thomarat, F., Metenier, G. (2002) Functional and evolutionary analysis of a eukaryotic parasitic genome. Curr. Opin. Microbiol. 5, 499–505.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., et al. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. Nature 454(7200), 126–130.

Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10(1), 57–63.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P., et al. (2013) CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 41(6): e74-e74.

Wang, Y., Xue, S., Liu, X., Liu, H., Hu, T., et al. (2016) Analyses of long noncoding RNA and mRNA profiling using RNA sequencing during the preimplantation phases in pig endometrium. Sci. Rep. 6, 20238.

Williams, B.A., Slamovits, C.H., Patron, N.J., Fast, N.M., Keeling, P.J. (2005) A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc. Natl. Acad. Sci. USA 102, 10936–10941.

Williams, G.R., Shafer, A.B.A., Rogers, R.E.L., Shutler, D., Stewart, D.T. (2008) First detection of *Nosema ceranae*, a microsporidian parasite of European honey bees (*Apis mellifera*), in Canada and central USA. J. Invertebr. Pathol. 97(2), 189–192.

Wilusz, J.E., Sunwoo, H., Spector, D.L. (2009) Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23(13): 1494–1504.

Xiao, H., Yuan, Z., Guo, D., Hou, B., Yin, C., et al. (2015) Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. BMC Genomics 16(1), 1–16.

Xing, D., Liang, J.Q., Li, Y., Lu, J., Jia, H.B., et al. (2014) Identification of long noncoding RNA associated with osteoarthritis in humans. Orthop. Surg. 6(4), 288–293.

Xue, Y., Ma, G., Gu, D., Zhu, L., Hua, Q., et al. (2015) Genome-wide analysis of long noncoding RNA signature in human colorectal cancer. Gene 556(2), 227–234.

Yang, L., Froberg, J.E., Lee, J.T. (2014) Long noncoding RNAs: fresh perspectives into the RNA world. Trends Biochem. Sci. 39(1), 35–43.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., et al. (2012) Identification and properties of 1119 candidate lincRNA loci in the *Drosophila melanogaster* genome. Genome Biol. Evol. 4(4), 427–442.

Zhang, T., Zhang, X., Han, K., Zhang, G., Wang, J., et al. (2017) Genome-wide analysis of lncRNA and mRNA expression during differentiation of abdominal preadipocytes in the Chicken. G3: Genes|Genomes|Genetics 7(3), 953–966.

Zhou, Z.Y., Li, A.M., Adeola, A.C., Liu, Y.H., Irwin, D.M., et al. (2014) Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. Genome Biol. Evol. 6(6), 1387–1392.