Computational Biology and Chemistry 70 (2017) 96–106

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

Biocomputational identification and validation of novel microRNAs predicted from bubaline whole genome shotgun sequences

H.K. Manku[a], J.K. Dhanoa[a], S. Kaur[b], J.S. Arora[a], C.S. Mukhopadhyay[a,*]

[a] School of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
[b] Department of Animal Genetics and Breeding, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form 1 July 2017
Accepted 7 August 2017
Available online 14 August 2017

Keywords:
microRNA
Whole genome shotgun sequencing
Prediction
Validation
Buffalo
Novel
Real time PCR

ABSTRACT

MicroRNAs (miRNAs) are small (19–25 base long), non-coding RNAs that regulate post-transcriptional gene expression by cleaving targeted mRNAs in several eukaryotes. The miRNAs play vital roles in multiple biological and metabolic processes, including developmental timing, signal transduction, cell maintenance and differentiation, diseases and cancers. Experimental identification of microRNAs is expensive and lab-intensive. Alternatively, computational approaches for predicting putative miRNAs from genomic or exomic sequences rely on features of miRNAs viz. secondary structures, sequence conservation, minimum free energy index (MFEI) *etc*. To date, not a single miRNA has been identified in bubaline (*Bubalus bubalis*), which is an economically important livestock. The present study aims at predicting the putative miRNAs of buffalo using comparative computational approach from buffalo whole genome shotgun sequencing data (INSDC: AWWX00000000.1). The sequences were blasted against the known mammalian miRNA. The obtained miRNAs were then passed through a series of filtration criteria to obtain the set of predicted (putative and novel) bubaline miRNA. Eight miRNAs were selected based on lowest E-value and validated by real time PCR (SYBR green chemistry) using RNU6 as endogenous control. The results from different trails of real time PCR shows that out of selected 8 miRNAs, only 2 (hsa-miR-1277-5p; bta-miR-2285b) are not expressed in bubaline PBMCs. The potential target genes based on their sequence complementarities were then predicted using miRanda. This work is the first report on prediction of bubaline miRNA from whole genome sequencing data followed by experimental validation. The finding could pave the way to future studies in economically important traits in buffalo.

1. Introduction

The eukaryotic microRNAs (miRNAs) belong to the family of single-stranded, non-coding, tiny RNAs of 19 to 25 nucleotide length, that act as post translational regulator of expression of several genes. Primarily, miRNAs negatively regulate gene expression at the post-transcriptional level by binding target mRNAs for cleavage or inhibition of translation (Kim et al., 2011). These micro-regulators have been reported to be located mostly within non-coding regions of genomes in animals, plants, fungi, viridiae and are usually transcribed by RNA polymerase II (Carrington and Victor, 2003). Mature miRNA is produced from the initial transcript of pri-miRNA through pre-miRNA with a characteristic hairpin structure (Yang et al., 2012). Hundreds of miRNA genes have been found in diverse animals, and many of these are phylogenetically conserved. Multidimensional role of microRNAs in developmental process, cell proliferation, cell death, patterning of the nervous system and haematopoiesis indicates that miRNAs are more numerous, and their regulatory impact is more pervasive than was previously anticipated (Victor, 2004).

In Asia, buffalo plays a pivotal role in socio-economic development through contributing milk, meat, hides and draft power for agricultural operations. Buffalo has been an integral part of livestock-rearing and agriculture in Asia for over 5000 years (Nanda and Toshihiko, 2003). There is plethora of scope for identification of novel biomarkers to impact faster genetic improvement in buffaloes. The post-transcriptional knocking down of gene has not been reported in buffaloes. In other domesticated species, miRNAs were initially discovered by genetic screening approaches, however, advanced experimental

* Corresponding author.
E-mail address: csmbioinfo@gmail.com (C.S. Mukhopadhyay).

http://dx.doi.org/10.1016/j.compbiolchem.2017.08.005
1476-9271/

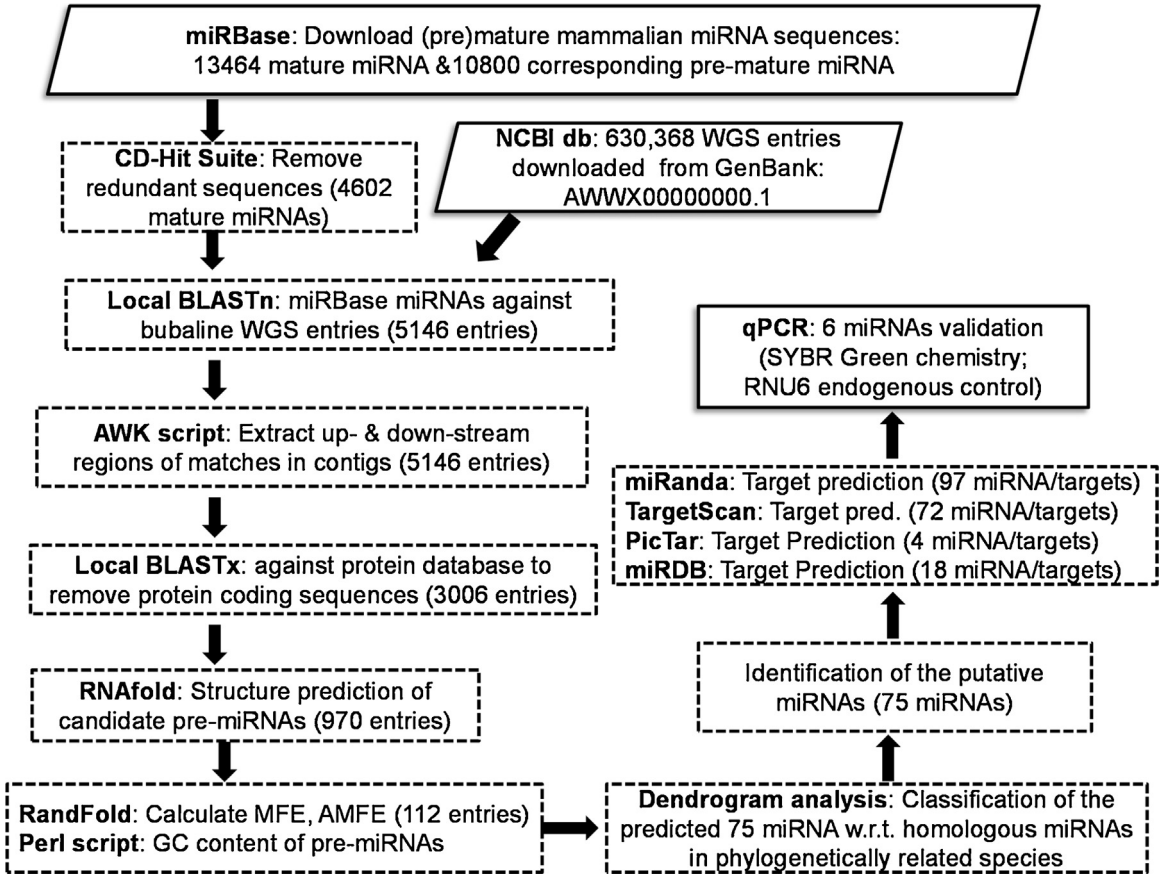

Fig. 1. Flow chart depicting the steps for predicting bubaline miRNAs from the whole genome shotgun sequencing contigs.

approaches were limited to species like cattle, sheep, swine *etc*., perhaps due to high cost involvement of next generation sequencing (NGS) of sRNAs. Alternatively, computational approaches have been developed which are based on matching certain thermodynamic (minimal free energy and a high minimal folding free energy index *etc*.), sequence characteristics (nucleotide content, G/C-content *etc*.) and structural features of miRNAs such as hairpin-shaped secondary structures, conservation of several miRNAs (Zhang et al., 2007; Arzuba et al., 2014).

In the present study we have applied an *in silico* approach to identify potential miRNA through homology based approach and to predict their possible targets in *Bubalus bubalis* as there were no miRNA reported for this species. Our work predicts *in silico* the bubaline miRNAs from a total of 6 30 368 whole genome shotgun reads of *Bubalus bubalis* (INSDC: AWWX00000000.1). Some of the identified miRNAs have been validated through real time PCR. The identified miRNAs can be experimentally tested as biomarkers to improve disease resistance, increase production and reproduction efficiency in buffalo.

2. Materials and methods

The steps followed during *in silico* prediction of *bubaline* miRNA have been described in the pipeline in Fig. 1.

2.1. Input data

The whole genome sequence (6 30 368 entries in FASTA format) of the water buffalo (*Bubalus bubalis*) genome (GenBank: AWWX00000000.1) was downloaded from http://www.ncbi.nlm.nih.gov/genome/?term=AWWX01 (Wang et al., 2005). The reported 13 464 miRNA and 10 800 pre-miRNA sequences of 30

Table 1
Sequence detail of the 8 selected miRNAs for primer designing based on their respective E-values and their uniqueness in the phylogenetic tree.

SN	WGS Fragment ID	miRNA-seq (5′–3′)	miRNA Name[a]	e-value[b]
1	AWWX01547453.1	gccaaaaaguucguucagguuu	bta-miR-2285e	6E-4
2	AWWX01546914.1	uguaugugugugguauguau	rno-miR-466b-3p	2E-3
3	AWWX01527129.1	auacguacauauauauauaua	hsa-miR-1277-5p	2E-3
4	AWWX01525429.1	aaaaucugaaugaacuuuuugg	bta-miR-2285b	4E-3
5	AWWX01524307.1	ccaaaaaguucguucugguuuu	bta-miR-2285j	2E-3
6	AWWX01476057.1	aagaaagcgaaguccauccucau	bta-miR-2322-5p	2E-4
7	AWWX01441341.1	aaaagcccaaaugaacuuuuugg	bta-miR-2285ad	3E-3
8	AWWX01428950.1	guccucaaggagcuucagucuagua	efu-miR-151	3E-5

[a] Best hit of the predicted miRNA against miRBase subject sequence.
[b] Expectation value (E-value) of the hit obtained by BLASTn.