HeptD: A database describing epigenetic differences between Thoroughbred and Jeju horses

Jeong-An Gim a,b,1, Sugi Lee c,1, Dae-Soo Kim d,1, Kwang-Seuk Jeong a,e, Chang Pyo Hong f, Jin-Han Bae g, Jae-Woo Moon f, Yong-Seok Choi b,c, Byung-Wook Cho h, Hwan-Gue Cho i, Jong Bhak j, Heui-Soo Kim a,b,*

a Department of Biologic Sciences, College of Natural Sciences, Pusan National University, Busan 609-735, Republic of Korea
b Genetic Engineering Institute, Pusan National University, Busan 609-735, Republic of Korea
c Department of Statistics, College of Natural Sciences, Pusan National University, Busan 609-735, Republic of Korea
d Genome Resource Center, Korea Research Institute of Bioscience and Biotechnology (KRBIB), 111 Gwahangno, Yuseong-gu, Daejeon 305-806, Republic of Korea
e Institute of Environmental Technology & Industry, Pusan National University, Busan 609-735, Republic of Korea
f TBI, Theragen BIG Institute, TheragenEtes, Suwon 443-270, Republic of Korea
g Research Center, Dongnam Institute of Radiological and Medical Sciences (DIRAMS), Busan, Republic of Korea
h Department of Animal Science, College of Life Sciences, Pusan National University, Miryang 627-702, Republic of Korea
i School of Computer Science and Engineering, College of Engineering, Pusan National University, Busan 609-735, Republic of Korea
j BioMedical Engineering, UNIST, Ulsan, Republic of Korea

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form 19 December 2014
Accepted 23 January 2015
Available online 28 January 2015

Keywords:
Thoroughbred horse
Jeju horse
MeDIP-Seq
DNA methylation
Differentially methylated region

ABSTRACT

With the advent of next-generation sequencing technology, genome-wide maps of DNA methylation are now available. The Thoroughbred horse is bred for racing, while the Jeju horse is a traditional Korean horse bred for racing or food. The methylation profiles of equine organs may provide genomic clues underlying their athletic traits. We have developed a database to elucidate genome-wide DNA methylation patterns of the cerebrum, lung, heart, and skeletal muscle from Thoroughbred and Jeju horses. Using MeDIP-Seq, our database provides information regarding significantly enriched methylated regions beyond a threshold, methylation density of a specific region, and differentially methylated regions (DMRs) for tissues from two equine breeds. It provided methylation patterns at 784 gene regions in the equine genome. This database can potentially help researchers identify DMRs in the tissues of these horse species and investigate the differences between the Thoroughbred and Jeju horse breeds.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

DNA methylation at the cytosine of CpG dinucleotides is a biochemical modification that regulates gene expression in eukaryotic genomes. It is involved in many biological processes like genomic imprinting, development, and cell differentiation by epigenetic modification. Abnormal DNA methylation has been implicated in diseases including cancer (Bird, 2002; Robertson, 2005; Klose and Bird, 2006; Illingworth et al., 2008), With the advent of next generation sequencing (NGS) technologies, genome-wide DNA methylation data is now publicly available in many human disease samples. The ENCODE project provided researchers with the function of each human and mouse genomic region, and differentially methylated regions (DMRs) for tissues from two equine breeds. It provided methylation patterns at 784 gene regions in the equine genome. This database can potentially help researchers identify DMRs in the tissues of these horse species and investigate the differences between the Thoroughbred and Jeju horse breeds.

methylating data has been provided (Li et al., 2011; Sati et al., 2012), but these data are difficult to analyze and information is scattered across a wide number of sources. In order to overcome this deficiency, we have developed a DNA methylation database for Thoroughbred and Jeju horse cerebrum, lung, heart, and skeletal muscle tissues. The Horse Epigenetic Database (HeptD) provides the methylation peaks of each genomic region, and differentially methylated regions (DMRs) in four tissues of the two horse breeds.

Many bioinformatics tools provide methylation patterns for each genomic region in humans. Abnormal methylation patterns were confirmed by many databases based on the results of whole-genome methylation analysis (He et al., 2008; Hackenberg et al., 2011; Lv et al., 2012). Tissue-specific methylation patterns were also provided in humans (Xin et al., 2012; James et al., 2013). However, these databases concentrated solely on the human genome. Although genome-wide DNA methylation data in organisms, besides humans and plants, have been made accessible to the public, there have been few freely available databases until now.

Thoroughbred horses are intentionally selected for speed, stamina, and agility. Therefore, Thoroughbred horses have many specific genetic characteristics related to horse economic traits, like genotype, SNPs, and
transcripts (McGivney et al., 2009; Hill et al., 2010a; Hill et al., 2010b; Park et al., 2012). Jeju horses are the descendants of 160 Mongolian horses that have inhabited and been bred since 1276 on Jeju Island, South Korea (Nam, 1969). Compared to the Thoroughbred, the Jeju horse is hardy, small, and often used for racing or as a food source. Analysis of the Jeju horse is valuable for identifying Jeju-specific traits and conservation of the Korean genetic resource. This study is to provide information about DMRs between four tissues as well as between two different horse breeds.

Transposable elements (TEs) make up approximately 30–50% of the mammalian genome. The expression of TEs induces genome instability, thus the host genome maintains DNA methylation in order to inhibit TE expression (Carnell and Goodman, 2003; Girardot et al., 2006). Variation of DNA methylation state in TE regions could induce a change in the genome defense system. Comparing DMRs between the TEs could provide interesting research topics. This database provides TE location and insertion direction along with their DNA methylation state.

Providing genomic information and constructing a web-based database of DNA methylation patterns is generally useful for performing further functional studies as well as for elucidating tissue-specific or breed-specific traits. This database provides the relatively highly-methylated regions in the whole genome, and an additional selection option helps to compare the methylation of two equine breeds in four tissues. We also provide the ‘Gene index’ option for easily comparing gene regions between tissues of the two equine breeds. In order to enhance the further studies of database users, we have linked our results pages to the related UCSC Genome Browser and Ensembl database for each genomic region. Results are presented as a web page or in Microsoft Excel format. The HEpD will provide insight for the functional study of methylation and potential roles for finding tissue-specific and breed-specific methylation patterns.

2. Material and methods

2.1. Sample information

We performed methylated DNA immunoprecipitation followed by next-generation sequencing (MeDIP-Seq) in four tissues (cerebrum, lung, heart, and skeletal muscle) derived from genomic DNA from each individual of Thoroughbred and Jeju horses (Lee et al., 2014) (Table S1). All animal protocols (ethical procedures and scientific care) used in this study were reviewed and approved by the Pusan National University-Institutional Animal Care and Use Committee (PNU-IACUC; Approval Number PNU-2013-0411).

2.2. Data source generation of datasets

A total of eight genomic DNA were extracted from tissues, then sonicated and immunoprecipitated for MeDIP-Seq analysis. The DNA libraries were constructed by paired-end sequencing with a lead length of 50 bp via the Illumina HiSeq 2000. Sequence data were aligned to the horse reference genome (equCab2) by using the SOPAaligner ver. 2.21 (Li et al., 2009). The information on data production and alignment was indicated in Table S2, and genome coverage variation with sequence read depth was analyzed (Table S3). The peak scanning was carried out by MACS (version 1.4.2) with a cutoff P-value of $1 \times 10^{-4}$ (Zhang et al., 2008).

2.3. Identity methylation states

The location of a methylation peak was converted to a *.csv file and then stored in a MySQL database. The genomic regions and their corresponding methylation values were stored. Each methylation peak resolution is 50 bp. As shown in Fig. 1, we classified the methylation peak values into three categories for subsequent studies. First, Type I provides the methylation values and coordinates of peaks that pass a certain threshold set by the user (Fig. 1A). Second, the user can input a specific region in the horse genome; The Type II category will return the methylation peak values in the user-specified region (Fig. 1B). Lastly, the Type III search will identify regions with differential methylation between the Thoroughbred and Jeju horses. The degree of difference can be specified by the user. To compare the methylation peak values between two breeds, the Type III search is recommended (Fig. 1C).

Fig. 1. The three strategies used by HEpD for finding methylated regions. (A) This strategy provides the methylation peak values that pass a user-set threshold (Type I). (B) This strategy displays methylation peak values within a specified genomic region (Type II). (C) This strategy displays the differentially methylated peak values that pass a user-set threshold (Type III) of the Thoroughbred and Jeju horses.
Fig. 2. A screenshot of HEpD. (A) The main page of the database. The web interface retrieves and displays the selected methylation patterns from the total database. (B) The interface of the result page. This page contains the information about the methylation peak regions, breed-specific patterns, and tissue-specific patterns.
3. Results and discussion

In MeDIP-Seq analysis, a total of 21 to 24 million reads from eight samples were sequenced, and these data size was produced as 1.08 to 1.2 Gb (Table S2). After filtration, high-quality reads were mapped and aligned against the reference genome (Table S3). In order to confirm the DNA methylation pattern, analysis of methylated peaks in MeDIP-Seq is essential. Therefore, HEpD provides the methylation peaks from primary MeDIP-Seq data formatted as *.csv files. Each dataset was converted to a MySQL management system, and accessed via a PHP-based web interface. The web server is maintained by an Apache system. This database provides a user-friendly interface with multiple search options. Each genomic region found in this database can be easily accessed in the UCSC Genome Browser and Ensembl database. When additional genomic information becomes available for the equine genome, such as a new genome assembly or newly identified genes, this information can be applied to our database. We used two breeds of horse, and each individual was sacrificed in their golden age as racing ability (Gramm and Marksteiner, 2010). The methylation patterns can change as developmental state, sex, and age dependent (Zhang et al., 2011; Martino et al., 2013), and this database could provide a reference methylation state for users in further studies.

In this database, there are two retrieval steps. One is the main page (Fig. 2A), the other is the results section generated by the user’s query (Fig. 2B). Users can efficiently get data concerning methylation information for each genomic region. We present four methods to find specific methylation regions for each experimental aim in the MeDIP-Seq based database including gene index selection. The overview of the workflow in HEpD is shown in Fig. 3. Users can find the methylation patterns from four tissues from two horse breeds. In the results, each methylation degree is presented as a number, and the minimum resolution is 50 bp. The result is presented as a table, which contains the genomic location, degree of methylation, gene name, accession number, gene direction, TE family, TE insertion direction, and links to the UCSC Genome Browser or Ensembl database for each genomic region (web links are only provided in the web-based results). The methylation value is indicated color-wise; the more highly methylated regions are indicated by darker shaded color in the results page. The genic regions and TE-insertion regions are indicated by color, and their direction is separated by color. The intergenic regions and non-TE-inserted regions are indicated by NA (Not Applicable). This method is an ideal way to screen breed-specific or tissue-specific hypermethylated genomic regions.

3.1. Identifying highly methylated regions beyond threshold

This analysis method can be used to search for highly methylated regions in the equine genome. Fig. 1A shows a schematic genomic region and the corresponding methylation peaks. This analysis option can be selected in the ‘Chromosome & Regions’ query section in the web interface. Users can search interesting gene names or TE families by inputting the names in the text box. The result consists of eight columns (seven columns in the Excel results table). The results table provides the location, the methylation value, gene name, gene accession number, gene direction, TE family, TE direction, and the link to the UCSC Genome Browser or Ensembl database for each genomic region (web links are only provided in the web-based results). The methylation value is indicated color-wise; the more highly methylated regions are indicated by a darker shaded color in the results page. The genic regions and TE-insertion regions are indicated by color, and their direction is separated by color. The intergenic regions and non-TE-inserted regions are indicated by NA (Not Applicable). This method is an ideal way to screen breed-specific or tissue-specific hypermethylated genomic regions.

3.2. Confirming the methylation state of a specific genomic region

This method can help one to confirm the methylation degree of a specific genomic region. The user can select the genomic regions of interest on the chromosome (Fig. 1B), and the results will show all genomic information including the methylation values found in this region. DNA methylation in promoter regions is known to regulate gene expression levels, therefore it is important to identify the methylation degree in the promoters of interesting genes (Barres et al., 2012). We provide the option to find the gene name, and alternatively, the user can search for methylation data by gene name. This information can help the user to identify highly methylated genomic regions. The user can click each link to the UCSC Genome Browser or Ensembl database in order to get detailed genomic information as well as the equine

![Fig. 3](image-url) A flow diagram representing the overall process for searching the methylation state in four tissues from two horse breeds.
Fig. 4. A screenshot of the ‘Gene index’ section. The table allows the user to retrieve the methylation state of specific genes and their location (A). The methylation state of the TNF gene in Thoroughbred horse skeletal muscle (B) and Jeju horse skeletal muscle (C). This section provides the results as a new tab or window where users can easily compare methylation state between tissues or horses.
reference sequence information. By using this method, users can retrieve the methylation state of specific regions in horse genome, then apply to design primer sets for bisulfite sequencing. We confirmed tissue-specific regions revealed in the previous study by this method (Lee et al., 2014), and this database provides similar methylation patterns (Fig. S1).

3.3. Confirm tissues of breed-specific DMRs

This query system identifies DMRs between the Thoroughbred and Jeju horses. This system provides the DMRs and their genomic information. At the whole genome level, comparison of DNA methylation may provide a clue to the epigenetic regulation of these different equine traits. In order to select the comparison of the corresponding four tissues between the Thoroughbred and Jeju horses, the user must arbitrarily select breeds, and chromosome as mandatory. The differences and threshold (the minimum methylation degree) are set by the user. The user can retrieve the DMR and genomic information between the two breeds for the selected tissues.

3.4. Gene index

The ‘Gene index’ section provides the methylation state of a particular gene in the various tissues of the two equine breeds. This section consists of a table, and allows the user to search by gene name or gene accession number. We provide a total of 784 genes identified in NCBI RefSeq database based on the equCab2 horse genome assembly [http://www.primate.or.kr/hepd/TissueDB_geneindex.html]. Each gene is displayed by chromosome number, and then ordered by genomic location (Fig. 4A). In order to compare the methylation state of gene promoter regions, the user can get the promoter genomic location and direction, then search for the methylation state of the specific genomic region. The user can also compare methylation states between tissues of two equine breeds. The user selects the tissues and genes of interest, and then the system provides the results in a new tab or window (depending on the web browser). For example, we compared DNA methylation at the TNF gene in Thoroughbred and Jeju horses and identified different methylation states in the skeletal muscle of the two breeds (Fig. 4B, C). This section can provide target genes to explain the mechanism of breed-specific gene expression (Geisen et al., 2014).

4. Conclusion

HEpD provides methylation data for four tissues from two horse breeds. This system helps one to select the highly-methylated genomic regions beyond a user-set threshold, presents the methylation state of a specific genomic region, compares the methylation state between two breeds, and uses a gene index for comparing the methylation state in a gene region. Users can see the detailed methylation state in tissues of two horse breeds displayed on the results page via four methods. Our work could provide insight into the role of DNA methylation in regulating horse traits and function. HEpD is freely available at the URL [http://www.primate.or.kr/hepd]. Any questions and advice regarding the use of this database are always welcome.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2015.01.047.

Acknowledgment

This work was supported by a grant from the Next Generation BioGreen 21 Program (No. PJ0081062011), Rural Development Administration, Republic of Korea.

References


Carnell, A.N., Goodman, J.L., 2003. The long (LINEs) and the short (SINEs) of it: altered DNA methylation as a precursor to toxicity. Toxical Sci. 75, 229–235.


