Copy number variations of five Y chromosome genes in donkeys

Haoyuan Han, Xiaocheng Zhao, Xiaoting Xia, Hong Chen, Chuzhao Lei, and Ruihua Dang

College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China

Correspondence to: Chuzhao Lei (leichuzhao1118@126.com)

Received: 21 July 2017 – Revised: 11 September 2017 – Accepted: 14 September 2017 – Published: 18 October 2017

Abstract. In mammals, the Y chromosome plays a pivotal role in male sex determination and is essential for normal sperm production. A number of studies were conducted on Y chromosome genes of various species and identified single-copy and multi-copy genes. However, limited studies about donkey Y chromosome genes have been done. In this study, 263 male samples from 13 Chinese donkey breeds were collected to analyze the copy number variations (CNVs) of five Y chromosome genes using the quantitative PCR (qPCR) method. These five genes (cullin 4 B Y (CUL4BY), equus testis-specific transcript y1 (ETSTY1), equus testis-specific transcript y4 (ETSTY4), equus testis-specific transcript Y 5 (ETSTY5), and sex-determining region Y (SRY)) were identified as multi-copy, whose median copy numbers (MCNs) were 5, 45, 2, and 2, and 13 with CNV ranges of 1–57, 1–227, 1–37, 1–86 and 1–152, respectively. The CNVs of these five genes were shared in different breeds. Compared to previous studies, the copy numbers of five genes showed some distinct consequences in this study. In particular, the well-known single-copy SRY gene showed CNVs in donkeys. Our results provided genetic variations of donkey Y chromosome genes.

1 Introduction

As a consequence of 230–300 million years of independent and non-recombining evolution, the majority (95 %) of the present-day Y chromosome is a male-specific region (MSY) that does not recombine with the X chromosome during meiosis to undergo homologous recombination (Graves, 1998; Quintana-Murci and Fellous, 2011). The MSY is enriched with multi-copy genes and copy number variations (CNVs) (Jobling, 2008; Skaletsky et al., 2003). CNVs are DNA segments of at least 50 bp in size that vary in copy number among individuals of a species (Mills et al., 2011). In many cases, CNV has been linked to gene dosage, gene regulation, gene expression, and phenotype (Cahan et al., 2009; Henrichsen et al., 2009; Redon et al., 2006; Toft and Ross, 2010; Zhang et al., 2009).

Y chromosome is notoriously difficult to be sequenced because of its distinctive and structurally repetitive genomic landscapes. Thus, studies on CNV do not include an analysis of variation on the Y chromosome generally and only focus on a few Y-linked genes (Skaletsky et al., 2003). The testis-specific protein Y-encoded (TSPY) gene shows 20–76 copies and the CNVs have been found to be associated with semen quality and reproduction by regulating cell division in the process of spermatogenesis in men (Nickkholgh et al., 2010; Vodicka et al., 2007). The TSPY gene shows 37–200 copies on the bovine Y chromosome (BTAY) (Hamilton et al., 2009; Jakubiczka et al., 1993). The heat-shock transcription factor, Y-linked (HSFY) gene contains around 70 copies in bulls (Hamilton et al., 2011). The data from Yue et al. (2013) confirmed that the copy number of the preferentially expressed antigen in melanoma, Y-linked (PRAMEY) gene family was highly variable, ranging from 2 to 31 across 15 cattle breeds, and plays an important role in spermatogenesis.

The sex-determining region Y (SRY) gene is known as the sex determination gene in mammalian development (Wilhelm et al., 2007). Thus, it is crucial to make clear of its CNV for a better understanding of SRY gene in donkeys. The testis-specific transcript Y (ETSTY) family is identified to consist of equine-specific genes (Paria et al., 2011); three of them (ETSTY1, ETSTY4, and ETSTY5) were chosen to analyze the CNVs in this study. The copy number of the cullin 4 B Y (CUL4BY) gene has been detected in other species, such as
cats and dogs (Murphy et al., 2006; Pearks Wilkerson et al., 2008). The \textit{CUL4BY} gene was involved in this study in order to determine its copy numbers in donkeys. Copy numbers reveal the basic genetic and structural information of Y chromosome genes. Because of the lack of assembling Y chromosome sequences in donkeys, it is difficult to analyze the CNVs of those genes at donkey Y chromosome by re-sequencing. As such, the previous study determining copy numbers on donkey Y chromosome genes was limited.

Population and genetic sources of Chinese native donkeys are abundant. There are 24 Chinese native donkey breeds, and all of which have important functions in transportation and tourism as well as in the production of meat, skin, and milk. In particular, Dezhou donkeys are famous for their skins, which are great materials for donkey-hide gelatin, and Guanzhong donkeys are famous for improving the body size of small donkeys and producing large mules. However, until now, few analyses of CNVs of donkey Y-linked genes have been reported. In this study, the quantitative PCR (qPCR) method was performed to determine the copy number variations of five Y chromosome genes (\textit{CUL4BY}, \textit{ETSTY1}, \textit{ETSTY4}, \textit{ETSTY5}, and \textit{SRY}) in 263 males from 13 Chinese donkey breeds, which would provide novel genetic variations of donkey genome.

2 Materials and methods

2.1 Sampling

Totaling 263 blood and ear tissue samples were obtained from 13 Chinese donkey breeds (Table 1). One additional female individual was used as female control in order to affirm the male specificity of the horse Y chromosome primers in donkeys. Genomic DNA was extracted from blood using a standard phenol–chloroform method (Sambrook and Russell, 2002). The DNA concentration was diluted to 20 ng \(\mu\text{L}^{-1}\) by ultrapure water and stored at \(-20^\circ\text{C}\).

2.2 Primer design

Primers of the \textit{beta-actin ACTB} gene were designed using primer 5.0. The male-specific primers of five Y chromosome genes – \textit{CUL4BY}, \textit{ETSTY1}, \textit{ETSTY4}, \textit{ETSTY5}, and \textit{SRY} – were acquired from Paria et al. (2011). Details of primers are shown in Table 2. Two males, one female control, and one negative control were amplified using a routine PCR to identify the male specificity of five pairs of primers in donkeys. The PCR protocol was as follows: each 12.5 \(\mu\text{L}\) reaction contained 20 ng of pooled genomic DNA, 5 pM each primer (10 pmol \(\mu\text{L}^{-1}\)), 6.25 \(\mu\text{L}\) of 2 \(\times\) PCR mix buffer (including 0.375 U Taq DNA polymerase, 2 \(\times\) PCR buffer, 18.75 \(\mu\text{M}\) MgCl\textsubscript{2} and 2.5 \(\mu\text{M}\) dNTPs), and 4.75 \(\mu\text{L}\) of distilled water. Thermocycling consisted of an initial denaturation at 95 \(^\circ\text{C}\) for 5 min, followed by 35 cycles of at 94 \(^\circ\text{C}\) for 30 s, at annealing temperature (Table 2) for 40 s, at 72 \(^\circ\text{C}\) for 30 s, and a final extension at 72 \(^\circ\text{C}\) for 10 min, before finally storing at 4 \(^\circ\text{C}\).

2.3 Copy number variations (CNVs) estimation

Standard curves and CNVs for each gene were measured by qPCR using Roche Lightcycler 480 system. qPCR was performed in plates with 96 wells, on which a calibrator, a negative control (distilled water), and diluted DNA (60, 40, 20, 10, 5, 2.5, and 1.25 ng \(\mu\text{L}^{-1}\) for standard curves) or testing samples (5 ng \(\mu\text{L}^{-1}\) for CNV estimation) were set up. qPCR reactions with DNA samples, calibrator, and negative control were run in triplicate. Each reaction contained 10 \(\mu\text{L}\) of SYBR PCR master mix (TaKaRa, Dalian, China), 0.8 \(\mu\text{L}\) of primers (10 pmol \(\mu\text{L}^{-1}\)), 6.8 \(\mu\text{L}\) of distilled water, and 1.6 \(\mu\text{L}\) of DNA template (5 ng \(\mu\text{L}^{-1}\)). The qPCR was run with a program of the following steps: predenaturation at 95 \(^\circ\text{C}\) for 10 min, followed by 40 cycles of denaturation at 95 \(^\circ\text{C}\) for 5 s and annealing at appropriate temperature (Table 2) for 30 s. A melting curve was then generated by taking fluorescent measurements every 0.11 \(^\circ\text{C}\) from 60 until 95 \(^\circ\text{C}\).

The cycle threshold (\(C_T\)) values generated from DNA of graded concentrations for reference (\textit{ACTB}) and targeted genes (\textit{CUL4BY}, \textit{ETSTY1}, \textit{ETSTY4}, \textit{ETSTY5}, and \textit{SRY}) were used to calculate the standard curves and their linear regression equations, the slopes of which were applied to measure primer efficiencies according to the equation \(E = 10^{\left(-1 / \text{slope}\right)}\). Then, relying on additional three equations (listed as follows) described in Hamilton et al. (2009), the CNVs of \textit{CUL4BY}, \textit{ETSTY1}, \textit{ETSTY4}, \textit{ETSTY5}, and \textit{SRY} were estimated. In the equations, the \(C_T\) values of the calibrator were determined by the average of all \(C_T\) values obtained in different plates for each gene, and \(\Delta C_T = C_T - C_T\) of the test sample.

\[
\text{Copy number}_{\text{calibrator}} = \frac{(E_{\text{reference}})^{C_T_{\text{reference}}}}{(E_{\text{target}})^{C_T_{\text{target}}}}
\]

\[
\text{Ratio} = \frac{(E_{\text{target}})\Delta C_T_{\text{target}}(\text{calibrator} - \text{sample})}{(E_{\text{reference}})\Delta C_T_{\text{reference}}(\text{calibrator} - \text{sample})}
\]

\[
\text{Copy number}_{\text{test sample}} = (\text{copy number}_{\text{calibrator}}) \times (\text{ratio}) \times 2
\]

2.4 Statistical analysis

In order to minimize technical error and to obtain an accurate copy number (CN) estimation, raw qPCR data showing a coefficient of variation (CV) > 1% between the duplicates were excluded from further analysis. The Kolmogorov–Smirnov and Shapiro–Wilks normality tests were used to analyze the normality of the CN data (Justel Peña and Zamar, 1997; Shapiro and Wilk, 1965). Box plot analysis of the CN data was conducted to detect the outliers in all the breeds as a whole. Multiple pair-wise comparisons of median copy
Table 1. Sample information of 13 donkey breeds in China.

<table>
<thead>
<tr>
<th>Breed</th>
<th>Abbreviation</th>
<th>Sample size</th>
<th>Collection location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qinghai donkey</td>
<td>QH</td>
<td>30</td>
<td>Minhe county, Qinghai Province</td>
</tr>
<tr>
<td>Gunsha donkey</td>
<td>GS</td>
<td>25</td>
<td>Yulin city, Shaanxi Province</td>
</tr>
<tr>
<td>Huaibei donkey</td>
<td>HL</td>
<td>6</td>
<td>Huaibei city, Anhui Province</td>
</tr>
<tr>
<td>Jiami donkey</td>
<td>JM</td>
<td>19</td>
<td>Mizhi county, Shaanxi Province</td>
</tr>
<tr>
<td>Liangzhou donkey</td>
<td>LZ</td>
<td>11</td>
<td>Wuwei city, Gansu Province</td>
</tr>
<tr>
<td>Kulun donkey</td>
<td>KL</td>
<td>6</td>
<td>Kulun county, Inner Mongolia Region</td>
</tr>
<tr>
<td>Qingyang donkey</td>
<td>QY</td>
<td>16</td>
<td>Qingyang city, Gansu Province</td>
</tr>
<tr>
<td>Guanzhong donkey</td>
<td>GU</td>
<td>9</td>
<td>Fufeng county, Shaanxi Province</td>
</tr>
<tr>
<td>Dezhou donkey</td>
<td>DZ</td>
<td>60</td>
<td>Dezhou city, Shandong Province</td>
</tr>
<tr>
<td>Biyang donkey</td>
<td>BY</td>
<td>39</td>
<td>Biyang county, Henan Province</td>
</tr>
<tr>
<td>Xiji donkey</td>
<td>XIJ</td>
<td>16</td>
<td>Xiji county, Ningxia Region</td>
</tr>
<tr>
<td>Xinjiang donkey</td>
<td>XJ</td>
<td>21</td>
<td>Yining city, Xinjiang Region</td>
</tr>
<tr>
<td>Changyuan donkey</td>
<td>CY</td>
<td>5</td>
<td>Changyuan county, Henan Province</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>263</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. QPCR primer information for the ACTB gene and five Y chromosome genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer F</th>
<th>Primer R</th>
<th>Tm (°)</th>
<th>Length (bp)</th>
<th>Accession number</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTB</td>
<td>GACGCTTTTGCTCAC</td>
<td>ACGGGCTTTTGCTCAC</td>
<td>60–65</td>
<td>203</td>
<td>NC_009156</td>
<td>this study</td>
</tr>
<tr>
<td>CUL4BY</td>
<td>TGGGGAATCGTGGTGAATA</td>
<td>CAAGGATCGTGGTCCTTAC</td>
<td>60</td>
<td>172</td>
<td>EU687546</td>
<td>Paria et al. (2011)</td>
</tr>
<tr>
<td>ETSTY1</td>
<td>GACGGACGACCTTGTGTTTT</td>
<td>CTAGTGGCAGTCTCTCTTGG</td>
<td>63</td>
<td>234</td>
<td>EU687549</td>
<td>Paria et al. (2011)</td>
</tr>
<tr>
<td>ETSTY4</td>
<td>GCAGTCTGAGAGTTGTCTTCT</td>
<td>CTCACTGAGCTGACACTTTG</td>
<td>65</td>
<td>175</td>
<td>EU687552</td>
<td>Paria et al. (2011)</td>
</tr>
<tr>
<td>ETSTY5</td>
<td>CAAAACCAAGAGGAGACCA</td>
<td>CTCCAGGAGCAAGCTACTCG</td>
<td>65</td>
<td>210</td>
<td>EU687553</td>
<td>Paria et al. (2011)</td>
</tr>
<tr>
<td>SRY</td>
<td>CCAACGCTTTATCTTTCGCA</td>
<td>CCATTCTCTTGTTTCACTCC</td>
<td>63</td>
<td>203</td>
<td>NM_001081810</td>
<td>this study</td>
</tr>
</tbody>
</table>

number (MCN) between breeds were analyzed using a non-parametric Mann–Whitney U test (Mann and Whitney, 1947) with a Bonferroni correction (Dunn, 1961).

3 Results and discussion

3.1 Male specificity of primers

The donkey Y chromosome sequence is not available. In an investigation of the equid Y chromosome, the horse and donkey MSY gene contents were compared. Comparative analysis of horse and donkey MSY revealed that the CUL4BY, ETSTY1, ETSTY4, ETSTY5, and SRY genes are shared in the genome of these two species; the sequence identity was 95–100% between the homologous horse and donkey transcripts according to Paria et al. (2011), who also verified that the primers for these five genes were male-specific in both horses and donkeys. In this study, the results of DNA fragments amplified with two males, one female control, and one negative control are shown in Fig. 1; ACTB showed a single band of expected size in male and female samples. The ETSTY4 and SRY genes displayed a male-specific band. The CUL4BY, ETSTY1, and ETSTY5 genes showed different sized fragments in females. However, the amplified band for each gene was specific and correctly sized in males; thus, it is reasonable to use the primers for our CNVs analysis.

Figure 1. Gel electrophoresis of PCR products of six donkey genes. The primers of these five Y-linked genes amplified male-specific bands with expected fragment size labeled above the band. M: 2 kb DNA ladder; ♂: male donkey genomic DNA; ♀: female donkey genomic DNA; N: negative control (distilled water).

Table 3. Correlation coefficient of standard curve and primer efficiency for ACTB and five Y chromosome genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Correlation coefficient ($R^2$)</th>
<th>Primer efficiency ($E$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTB</td>
<td>0.9995</td>
<td>2.14</td>
</tr>
<tr>
<td>ETSTY1</td>
<td>0.9913</td>
<td>1.90</td>
</tr>
<tr>
<td>SRY</td>
<td>0.9933</td>
<td>1.93</td>
</tr>
<tr>
<td>ACTB</td>
<td>0.9973</td>
<td>1.90</td>
</tr>
<tr>
<td>ETSTY4</td>
<td>0.9976</td>
<td>2.01</td>
</tr>
<tr>
<td>ETSTY5</td>
<td>0.9941</td>
<td>1.97</td>
</tr>
<tr>
<td>ACTB</td>
<td>0.9915</td>
<td>1.92</td>
</tr>
<tr>
<td>CUL4BY</td>
<td>0.9961</td>
<td>1.98</td>
</tr>
</tbody>
</table>

3.2 The CNVs of five Y chromosome genes in donkeys

According to the linear regression equations of standard curves and the equation $E = 10^{1-1/\text{slope}}$, correlation coefficients ($R^2$) were all higher than 0.99 and primer efficiencies ($E$) were higher than 1.90 for the ACTB, CUL4BY, ETSTY1, ETSTY4, ETSTY5, and SRY genes (Table 3). In total, the CNVs of 263 males from 13 Chinese donkey breeds were estimated. The CUL4BY, ETSTY1, ETSTY4, ETSTY5, and SRY were all multi-copy with median copy numbers (MCNs) of 5, 45, 2, 2, and 13, respectively. The CNVs ranges were 1–57, 1–227, 1–37, 1–86, and 1–152, respectively (Table 4). Based on Kolmogorov–Smirnov and Shapiro–Wilk normality tests, the copy numbers (CNs) of the five donkey Y chromosome genes did not fit the normal distribution in Chinese horses ($P<0.0001$). Box plot analyses of the CNs data revealed that approximately 10–15 individuals (4–6%) were outliers that had significantly higher CN for each gene (Fig. 2).

For the CUL4BY gene, the CNVs ranged from 1 to 57 with MCNs of 2–9 in 13 donkey breeds, while Guanzhong (GZ) donkeys had the highest MCN of 9 (Table 4). Dezhou donkeys showed significantly different MCN from nine other breeds for the CUL4BY gene (Table S1 in the Supplement).

Similarly, the highly amplified CUL4BY gene has been identified in the cats and dogs. (Murphy et al., 2006; Pearks Wilkerson et al., 2008). The horse CUL4BY showed intermediate expression being transcribed in testis, heart, spleen, and kidney (Paria et al., 2011); thus, the horse CUL4BY might be related to testis and male germ cell proliferation (Petroski and Deshaies, 2005). Considering the close relationship of donkeys and horses, we assumed that the donkey CUL4BY had an important function in male fertility.

The ETSTY family is novel and equid-specific, the orthologue of which is not found in non-equid species. The range of CNVs of the ETSTY1 gene was 1–227, with the MCNs varying from 12 (GZ donkey) to 99 (Changyuan donkey) in different breeds (Table 4), indicating that the ETSTY1 gene was a largely amplified gene with multiple copies. GZ donkeys exhibited significant difference of MCNs from 10 other
The copy numbers of five donkey genes on Y chromosome were determined in current study. The **CUL4BY**, **ETSTY1**, **ETSTY4**, **ETSTY5**, and **SRY** genes are multi-copy and have MCNs of 5, 45, 2, 2, and 13 with CNV ranges of 1–57, 1–227, 1–37, 1–86 and 1–152, respectively. We first investigated the copy numbers of donkey Y chromosome genes, providing basic genetic information for future studies in donkeys.

**4 Conclusions**

The copy numbers of five donkey genes on Y chromosome were determined in current study. The **CUL4BY**, **ETSTY1**, **ETSTY4**, **ETSTY5**, and **SRY** genes are multi-copy and have MCNs of 5, 45, 2, 2, and 13 with CNV ranges of 1–57, 1–227, 1–37, 1–86 and 1–152, respectively. We first investigated the copy numbers of donkey Y chromosome genes, providing basic genetic information for future studies in donkeys.

**Data availability.** The original data of the paper are available upon request from the corresponding author.
Author contributions. HH, CL, and RD designed this project; XZ and XX collected blood from animals; HH performed experiments in the laboratory; and HH wrote the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by National Natural Science Foundation of China (31072001, 31272399, 81270439).

Reviewed by: two anonymous referees
Edited by: Steffen Maak

References

H. Han et al.: CNVs of donkey Y-linked genes


