



## Eight Y chromosome genes show copy number variations in horses

Haoyuan Han<sup>1,2</sup>, Xin Zhang<sup>1</sup>, Xiaocheng Zhao<sup>1</sup>, Xiaoting Xia<sup>1</sup>, Chuzhao Lei<sup>1</sup>, and Ruihua Dang<sup>1</sup>

<sup>1</sup>College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>2</sup>College of Animal Science and Technology, Henan University of Animal Husbandry and Economy, Zhengzhou, Henan 450046, China

**Correspondence:** Ruihua Dang (dangruihua@nwsuaf.edu.cn)

Received: 4 February 2018 – Revised: 18 April 2018 – Accepted: 25 May 2018 – Published: 2 July 2018

**Abstract.** Copy number variations (CNVs), which represent a significant source of genetic diversity on the Y chromosome in mammals, have been shown to be associated with the development of many complex phenotypes, such as reproduction and male fertility. The occurrence of CNVs has been confirmed on the Y chromosome in horses. However, the copy numbers (CNs) of *Equus caballus* Y chromosome (ECAY) genes are largely unknown. To demonstrate the copy number variations of Y chromosome genes in horses, the quantitative real-time polymerase chain reaction (qPCR) method was applied to measure the CNVs of the eukaryotic translation initiation factor 1A Y (*EIF1AY*), equine testis-specific transcript on Y 1 (*ETSTY1*), equine testis-specific transcript on Y 4 (*ETSTY4*), equine testis-specific transcript on Y 5 (*ETSTY5*), equine transcript Y4 (*ETY4*), ubiquitin activating enzyme Y (*UBE1Y*), sex determining region Y (*SRY*), and inverted repeat 2 Y (*YIR2*) across 14 Chinese domestic horse breeds in this study. Our results revealed that these eight genes were multi-copy; furthermore, some of the well acknowledged single-copy genes such as *SRY* and *EIF1AY* were found to be multi-copy in this research. The median copy numbers (MCNs) varied among different breeds for the same gene. The CNVs of Y chromosome genes showed different distribution patterns among Chinese horse breeds, indicating the impact of natural selection on copy numbers. Our results will provide fundamental information for future functional studies.

### 1 Introduction

The mammalian Y chromosome stands out from the rest of the genome because it is male specific, constitutively haploid, and exhibits unique structural and functional features (Skaletsky et al., 2003). This has led to a correspondingly unusual genomic landscape, rich in segmental duplications, which provide ample substrate for the generation of copy number variations (CNVs). CNVs, a major source of genetic variation between individuals, include deletions, duplications, and complex rearrangements typically larger than 50 base pairs to over several megabase pairs (Mb) in size. The male specific region of the Y chromosome (MSY) contains clusters of genes essential for male reproduction (Tüttelmann et al., 2011; Chang et al., 2013; Yue et al., 2013). In humans, CNVs of the testis-specific protein, Y-encoded (*TSPY*) have been found to be associated with semen quality and repro-

duction via the regulation of cell division in the process of spermatogenesis (Vodicka et al., 2007). In cattle, the CNVs of Y-linked genes also affect male fertility and play an important role in spermatogenesis (Hamilton et al., 2012; Chang et al., 2013; Yue et al., 2013). However, information about the annotation and transcriptome of horse MSY is still lacking.

Horses have played an instrumental role in transportation, agriculture, and warfare and have been faithful companions of humans since their domestication. Since the 1900's, due to the continuous development of combustion engine, the use of horses has gradually ceased. However, horses have not faded from human life. In many countries, horses have become domestic animals of both social and economic value (Yang et al., 2010). Although, the genetic variants that underlie the phenotypic diversification of horse breeds are poorly understood. Systematic discovery of the Y chromosome of *Equus caballus* (ECAY) genes started in 2004 (Raudsepp et

al., 2004). A detailed MSY gene catalogue was developed for the horse, with 37 horse MSY genes/transcripts being identified. The horse MSY harbors 20 X-degenerate genes and 17 acquired or novel genes (Paria et al., 2011); however, the specific CNVs of these genes have not been investigated.

In order to estimate the copy numbers of Y-linked genes in horses and compare the CNVs between different Chinese horse breeds, three X-degenerate genes, eukaryotic translation initiation factor 1A Y (*EIFIAY*), sex determining region Y (*SRY*), and ubiquitin activating enzyme Y (*UBEIY*), were chosen to have their copy numbers determined in this study. *SRY* is a well-known sex determination gene and the confirmation of its copy numbers will be helpful to form a better understanding of its structural and functional characters. *EIFIAY* and *SRY* genes were identified as single copy (Paria et al., 2011). Copy numbers of *UBEIY* orthologs were identified in other species (cats, pigs, and mice) (Mitchell et al., 1991; Quilter et al., 2002; Pearks Wilkerson et al., 2008), and the *UBEIY* gene was considered to be multi-copy in horses (Paria et al., 2011). However, the copy number range of *UBEIY* has not previously been provided. Five Y-ampliconic genes, equine testis-specific transcript on Y 1 (*ETSTY1*), equine testis-specific transcript on Y 4 (*ETSTY4*), equine testis-specific transcript on Y 5 (*ETSTY5*), equine transcript Y4 (*ETY4*), and inverted repeat 2 Y (*YIR2*), were all identified as multi-copy genes (Paria et al., 2011), but the ranges of copy numbers variations were inconclusive. Therefore, we investigated the CNVs of these eight Y chromosome specific genes in Chinese horses using the quantitative real-time polymerase chain reaction (qPCR) method. Our results will provide fundamental information regarding the copy numbers for horse Y chromosome genes, which will benefit future functional studies.

## 2 Materials and methods

### 2.1 Sample collection

Blood samples of 302 male horses were collected from 14 Chinese domestic breeds distributed in northwestern and southwestern China (Table 1). In addition, samples from two female horses were also collected to be used as female controls and water was used as a negative control; this was undertaken to verify the male specificity of the primers. The genomic DNA was extracted using a standard phenol-chloroform method (Sambrook and Russell, 2002). The DNA concentrations were diluted to  $20 \text{ ng} \mu\text{L}^{-1}$  with ultrapure water and stored at  $-20^\circ\text{C}$ .

### 2.2 Primer design

CNVs of eight Y chromosome specific genes, *EIFIAY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBEIY*, *SRY*, and *YIR2*, were investigated in this study. Because the sequence of horse Y chromosome is still lacking, copy numbers remain

**Table 1.** Sample information for the 14 chosen Chinese horse breeds.

Breed	Abbreviation	Sample size	Source region
Balikun	BLK	13	Balikun County, Xinjiang
Baise	BS	37	Baise County, Guangxi
Chaidamu	CD	17	Chaidamu, Qinghai
Chakouyi	CKY	30	Tianzhu County, Gansu
Debao pony	DB	23	Debao County, Guangxi
Guanzhong	GU	3	Fufeng County, Shaanxi
Guizhou	GZ	19	Guiyang, Guizhou
Hequ	HN	25	Maqu County, Gansu
Kazakh	HSK	14	Changji, Xinjiang
Ningqiang	NQ	6	Ningqiang County, Shaanxi
Datong	MY	7	Qilian County, Qinghai
Mogolia	MG	33	Chifeng, Mongolia
Yanji	YJ	30	Hejing County, Xinjiang
Yushu	YS	45	Yushu, Qinghai
Total		302	

uncertain for any Y-linked gene. Therefore, the two-copy gene, *beta-actin* (GenBank acc. no. NC\_009156) on an autosome was used as a reference. The PCR primers were designed for the *beta-actin* gene and the conserved region of the *SRY* gene, using the Primer Premier 5.0 program (<http://www.premierbiosoft.com/>). The other seven pairs of primers were obtained from Paria et al. (2011). The detailed information regarding the PCR primers and the predicted sizes of each amplicon is listed in Table 2. To confirm the Y chromosome-specificity of the designed primers, a routine PCR was performed using male and female horse genomic DNA as templates and water as a negative control. The PCR protocol was as follows: each  $12.5 \mu\text{L}$  reaction contained  $20 \text{ ng}$  of genomic DNA,  $5 \text{ pg}$  of each primer ( $10 \text{ pmol} \mu\text{L}^{-1}$ ),  $6.25 \mu\text{L}$  of  $2 \times$  PCR Mix buffer (including  $0.375 \text{ U}$  Taq DNA polymerase,  $2 \times$  PCR buffer,  $18.75 \mu\text{M}$   $\text{MgCl}_2$ , and  $2.5 \mu\text{M}$  dNTPs), and  $4.25 \mu\text{L}$  of distilled water. Thermocycling consisted of an initial denaturation at  $95^\circ\text{C}$  for 5 min, followed by 35 cycles at  $94^\circ\text{C}$  for 30 s,  $40 \text{ s}$  at annealing temperature (Table S1 in Supplement),  $72^\circ\text{C}$  for 30 s, a final extension at  $72^\circ\text{C}$  for 10 min, and then sample storage at  $4^\circ\text{C}$ . The PCR production of the male samples and female and negative controls were visualized on 1% native agarose gel. The images were acquired by a ChampGel™ 6000 Gel documentation and image analysis system and Lane 1D Gel imaging analysis software (Sagecreation, Beijing, China).

### 2.3 Quantitative real-time polymerase chain reaction

The quantitative real-time polymerase chain reaction (qPCR) method was used to measure the CNs of *EIFIAY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBEIY*, *SRY*, and *YIR2* in the samples using a Roche Lightcycler 480 system and SYBR PCR Master Mix (TAKARA, Dalian, China). Plates with 96 wells

**Table 2.** Correlation coefficient of the standard curve and the primer efficiency for eight horse Y chromosome genes.

Gene	Correlation coefficient ( $R^2$ )	Primer efficiency ( $E$ )
<i>beta actin</i>	0.9929	2.04
<i>EIF1AY</i>	0.9983	2.03
<i>ETSTY1</i>	0.9905	2.02
<i>SRY</i>	0.9947	2.11
<i>beta actin</i>	0.9918	1.98
<i>ETSTY5</i>	0.9910	1.93
<i>beta actin</i>	0.9900	1.99
<i>ETSTY4</i>	0.9901	2.00
<i>beta actin</i>	0.9967	2.12
<i>ETY4</i>	0.9903	1.99
<i>UBE1Y</i>	0.9998	2.06
<i>YIR2</i>	0.9908	1.91

were set up to run the qPCR. On each plate, wells were laid out for a calibrator, and a negative control (distilled water). Standard curves were generated from horse DNA diluted to 60, 40, 20, 10, 5, 2.5, and 1.25 ng  $\mu\text{L}^{-1}$  for eight pairs of primers. For the test samples, DNA was concentrated to 5 ng  $\mu\text{L}^{-1}$ . qPCR reactions with standard curve samples and test samples (including the calibrator and negative control) were run in triplicate. In this study, we ran a total of 302 horses on 176 plates (each plate was set up for 1 calibrator, 1 distilled water, and 14 testing samples) for the eight Y chromosome genes. Each reaction contained 10  $\mu\text{L}$  of SYBR Green PCR Master Mix, 0.8  $\mu\text{L}$  of primers (10 pmol  $\mu\text{L}^{-1}$ ), 6.8  $\mu\text{L}^{-1}$  of distilled water, and 1.6  $\mu\text{L}^{-1}$  of DNA template (5 ng  $\mu\text{L}^{-1}$ ). The qPCR was run with a program made up of the following steps: predenaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 5 s, and annealing at an appropriate temperature (Table S1) for 30 s. A melting curve was then generated by taking fluorescent measurements every 0.11 °C from 60 until 95 °C. Primer efficiencies were measured according to the equation  $E = 10^{(-1/\text{slope})}$ , and the slope was generated by a standard curve.

## 2.4 Copy number estimation

The CNs of *EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2* were estimated for test samples using the following three equations described in Hamilton et al. (2009):

$$\text{Copy number}_{\text{calibrator}} = \frac{(E_{\text{reference}})^{C_{T_{\text{reference}}}}}{(E_{\text{target}})^{C_{T_{\text{target}}}}}; \quad (1)$$

$$\text{Ratio} = \frac{(E_{\text{target}})^{\Delta C_{T_{\text{target}}(\text{calibrator-sample})}}}{(E_{\text{reference}})^{\Delta C_{T_{\text{reference}}(\text{calibrator-sample})}}}; \quad (2)$$

$$\text{Copy number}_{\text{test sample}} = (\text{Copy number}_{\text{calibrator}}) \times (\text{ratio}) \times 2. \quad (3)$$

In the above equations, the DNA sample of the horse Guizhou 59 was used as the calibrator. The cycle threshold ( $C_T$ ) value of the calibrator for each gene was determined by the average of 66  $C_T$  values obtained from 22 different plates for this particular sample. In equations 1–2,  $E$  = the PCR efficiency for the reference gene (*beta-actin*) or each target gene (*EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2*), and  $\Delta C_T = C_T$  of the calibrator –  $C_T$  of the test sample.

## 2.5 Quantitative real-time polymerase chain reaction data validation by TA cloning

To verify the accuracy of qPCR results, TA cloning was performed for the *EIF1AY* gene. PCR products from eight samples (three Baise (BS) horses, two Yanji (YJ) horses, one Debao (DB) horse, one Hequ (HN) horse, and one American Quarter (Q) horse) were purified using a Universal DNA Purification Kit (TIANGEN, Beijing, China), then ligated into the pGEM-T Easy cloning vector and transformed into *Escherichia coli* DH-5 $\alpha$  (CWBio, China). In total 108 clones (11–15 clones per sample) were picked and amplified using the PCR method. PCR products were sequenced on an ABI PRISM 377 DNA sequencer (Perkin-Elmer) (Shanghai Sangon Biotech Company, Shanghai, China).

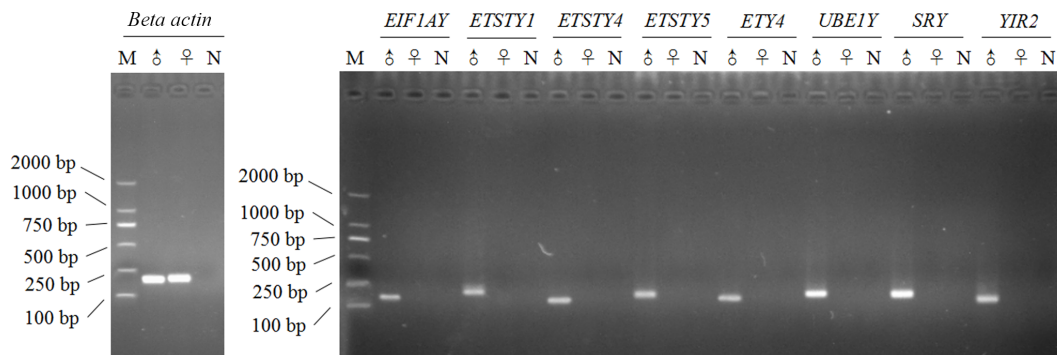
## 2.6 Statistical analysis

In order to minimize technical error and to obtain an accurate CN estimation, raw qPCR data that showed a coefficient of variation (CV) > 1 % between the duplicates were excluded from further analysis. The normality of the CN data was assessed with the Kolmogorov–Smirnov and Shapiro–Wilk normality tests (Shapiro and Wilk, 1965; Justel et al., 1997). Box plot analyses of the CN data were conducted to detect outliers in all the breeds as a whole. Multiple pair-wise comparisons of the median copy numbers (MCNs) between breeds were analyzed using a nonparametric Mann–Whitney  $U$  test (Mann and Whitney, 1947) with a Bonferroni correction (Dunn, 1961). MEGA 5.1 was used to align the cloning sequences (Tamura et al., 2011).

## 3 Results and discussion

### 3.1 Primer male-specificity

In order to validate the male specificity of the primers used in this study, a routine PCR was run using female DNA as the negative control. The results demonstrated that every primer pair for the target genes, *EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2*, amplified a male-specific band with the expected fragment size. This confirmed that the primers designed are male-specific and can be used for qPCR analysis in this study (Fig. 1).



**Figure 1.** Gel electrophoresis of the PCR products of the beta actin gene and the eight Y chromosome horse genes. M is the 2 kb DNA ladder; ♂ is male horse genomic DNA; ♀ is female horse genomic DNA; and N is the negative control (distilled water).

### 3.2 Standard curve and primer efficiency

The standard curves for the reference gene (*beta-actin*) and the eight target genes (*EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2*) were generated from horse DNA diluted to different concentrations; the correlation coefficients of the standard curves generated were all higher than 0.99. The resulting reactions had primer efficiencies higher than 1.90, demonstrating high amplification efficiencies. The correlation coefficients of the standard curves and the primer efficiencies for each primer are displayed in Table 2.

### 3.3 The copy number variations of eight genes on the equine Y chromosome

The gene copy numbers of tested horses were calculated using the calibrator as an adjustment based on Eqs. (2) and (3) (see Sect. 2) (for results see Table 3). Copy numbers determined by relative real-time PCR were considered to be approximations only, and not absolute copy numbers. As described in Hamilton et al. (2009), the copy number of the calibrator was estimated using the  $\Delta C_T$  method. The ratios relative to the calibrator were determined by the  $\Delta\Delta C_T$  method, which involves normalizing the samples to a calibrator to minimize the variation. Therefore, the relative copy numbers can be used to compare the relative amount of Y-linked genes between horses with confidence. Previous studies, have shown that relative real-time PCR can still produce a useful estimate of copy numbers (Yue et al., 2013; Hamilton et al., 2012, 2009).

The median copy numbers (MCNs) of the *EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2* genes were 1, 3, 8, 9, 26, 7, 1, and 12 with the CNs of 1–37, 1–25, 1–156, 1–80, 1–173, 1–77, 1–17, and 1–92, respectively (Table 3). Therefore, it is suggested that the eight horse Y chromosome genes were multi-copy. The number of private alleles for *EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *SRY*, *ETY4*, *UBE1Y*, and *YIR2* were 12, 7, 24, 12, 3, 47, 18, and 23 respectively. The one-copy alleles of the *EIF1AY* and *SRY*

genes were found in all 14 breeds, while other CNs were only shared in 2–13 breeds. We found that the CNs of the eight horse Y chromosome genes did not fit the normal distribution, with this relative small number of horses distributed to 14 breeds ( $P < 0.0001$ ), based on Kolmogorov–Smirnov and Shapiro–Wilk normality tests. As shown in Fig. 2, box plot analyses of the CNs data revealed that approximately 10 individuals ( $\sim 3\%$ ) were outliers, who displayed significantly higher CNs for each gene.

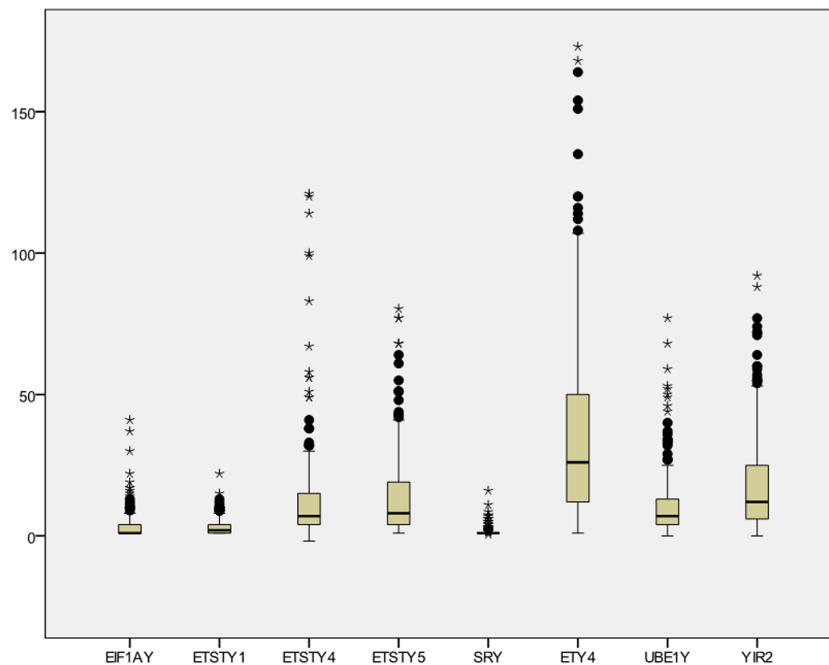
The eight horse Y chromosome genes studied were divided into X-degenerate genes (*EIF1AY*, *SRY*, and *UBE1Y*) and ampliconic genes (*ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, and *YIR2*) (Paria et al., 2011). Almost two thirds of the 20 X-degenerate genes found in horses are expressed ubiquitously and have a Y-linked homologue in mammalian species (Quiltere et al., 2002; Rohozinski et al., 2002; Skaletsky et al., 2003; Pearks Wilkerson et al., 2008; Hughes et al., 2010). Most equine Y-borne amplified sequences are expressed exclusively or predominantly in the testis (Paria et al., 2011), and presumably have a role in testicular function, which may be valuable in selecting stallions for breeding.

The *UBE1Y* gene is a X-degenerate gene, specifically expressed in testis (Paria et al., 2011). *UBE1Y* is conserved in most eutherians and marsupials, except that this gene is found as a pseudogene in some primate lineages and is absent in humans (Skaletsky et al., 2003). Our results showed that the CN of *UBE1Y* ranged from 1 to 77 among individuals (Table 3), which supports previous research stating that the horse is the only species where *UBE1Y* is a multi-copy gene (Paria et al., 2011). Orthologs in other species (cats, pigs, and mice) are single copy (Mitchell et al., 1991; Quilter et al., 2002; Pearks Wilkerson et al., 2008). The high copy number and testis-specific transcription of *UBE1Y* in horses, supports the hypothesis that the gene could be the gene responsible for regulating germ cell proliferation and, thus, male fertility (Lévy et al., 2000).

The *SRY* gene has a known function in sex determination at early stages of mammalian embryonic development (Wilhelm et al., 2007). It was considered to be a single-copy gene

**Table 3.** Copy number variations of eight Y chromosome genes in horses.

Breed	Sample size	Median (the range of copy number variations)							
		<i>EIF1AY</i>	<i>ETSTY1</i>	<i>ETSTY4</i>	<i>ETSTY5</i>	<i>ETY4</i>	<i>UBE1Y</i>	<i>SRY</i>	<i>YIR2</i>
BLK	13	1 (1–4)	3 (1–8)	6 (1–18)	35 (1–80)	22 (7–62)	4 (1–10)	1 (1–7)	12 (1–92)
BS	37	1 (1–4)	3 (1–9)	8 (1–58)	17 (3–68)	19 (1–95)	4 (1–46)	1 (1–17)	17 (2–72)
CD	17	2 (1–8)	1 (1–7)	6 (2–41)	3 (1–12)	16 (5–79)	9	1 (1–6)	8 (1–72)
CKY	30	3 (1–11)	2 (1–12)	13(1–32)	12 (2–43)	44 (6–151)	8 (2–52)	1 (1–3)	18 (4–71)
DB	23	1 (1–17)	4 (1–6)	12 (2–114)	7 (2–40)	30 (8–92)	5 (1–14)	1 (1–6)	7 (4–57)
GU	3	1	2 (2–3)	4 (2–4)	4 (2–4)	17 (7–26)	3 (1–5)	1	4 (3–4)
GZ	19	1 (1–3)	1 (1–6)	1 (1–9)	2 (1–11)	10 (4–50)	2 (1–6)	1	4 (2–39)
HN	25	1 (1–6)	3 (1–10)	7 (1–56)	7 (2–48)	21 (9–112)	7 (2–53)	1 (1–4)	16 (6–55)
HSK	14	2 (1–8)	4 (1–13)	6 (2–51)	7 (1–17)	30 (3–93)	5 (3–27)	1 (1–2)	16 (8–54)
NQ	6	2 (1–6)	1 (1–15)	8 (2–21)	14 (2–32)	76 (13–82)	34 (3–50)	1 (1–5)	26 (5–55)
MY	7	1 (1–2)	3 (1–6)	7 (3–13)	6 (2–12)	38 (18–103)	9 (6–15)	1	20 (12–29)
MG	33	5 (1–37)	4 (1–11)	11 (1–120)	13 (4–77)	53 (7–168)	13 (1–68)	1 (1–5)	18 (3–88)
YJ	30	1 (1–3)	3 (1–14)	9 (2–156)	5 (1–19)	23 (4–70)	7 (3–19)	1 (1–6)	12 (4–74)
YS	45	1 (1–24)	5 (1–25)	9 (1–78)	12 (2–78)	24 (4–173)	8 (1–77)	1 (1–4)	8 (2–49)
Total	302	1 (1–37)	3 (1–25)	8 (1–156)	9 (1–80)	26 (1–173)	7 (1–77)	1 (1–17)	12 (1–92)



**Figure 2.** Box plot analysis of eight genes CNs in horse. The outliers were indicated by a solid circle or an asterisk (extremely high CN).

in horses, which did not produce any signal by cDNA FISH (Paria et al., 2011). In our study, although the MCN of the *SRY* gene was fixed to 1 in all 14 breeds, the copy numbers of *SRY* ranged from 1 to 17 (Table 3). Therefore, we suggest that *SRY* is a multi-copy gene in horses. Considering the limited sensitivity of FISH, it is possible that the cDNA sequences of *SRY* may not have been long enough (> 1000 bp) to be examined using this method (Trask et al., 1993). *SRY* is present as a single copy in humans and mice, yet has been assumed

to be multi-copy in the cat and *Microtus cabreriae* and rabbit genomes (Fernández et al., 2002; Geraldés and Ferrand, 2006; Pearks Wilkerson et al., 2008). Equine *SRY* expression, similar to feline, is predominant in the testis, and can also show intermediate or broader expression (Pearks Wilkerson et al., 2008; Paria et al., 2011). There have been assertions that the *SRY* gene has more functions than just sex determination, and that the transcripts might also be actively needed in mature testis (Paria et al., 2011); this requires additional

investigation into the *SRY* copy number in other eutherian mammals.

*EIF1AY* was recognized as a single-copy gene with ubiquitous expression (Paria et al., 2011). Nine breeds have one MCN for the *EIF1AY* gene, but copy number variations (1–37) were detected (Table 3). Therefore, we assumed that *EIF1AY* was not a single-copy gene. This suggests that the population size and the method used in studies influences the result and final conclusion. Based on the 108 cloning sequences, 48 polymorphisms and 42 haplotypes were detected for the *EIF1AY* gene in this study (Table S2), which proved that the *EIF1AY* gene was multi-copy in horses. This was in accordance with our qPCR results, demonstrating that the CNVs results were credible and accurate using the qPCR method.

Compared with the X-degenerate genes, the ampliconic gene content is more diverse among lineages (Skaletsky et al., 2003). Five Y-ampliconic genes, (*ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, and *YIR2*), were all present in multiple copies in our study, which was consistent with the results from Paria et al. (2011). The MCN of the *ETSTY1* gene varied from 1 to 5 among the 14 breeds. The MCN of the *ETSTY4* gene was highest in Chakouyi breed (13) and lowest in Guizhou breed (1). For the *ETSTY5* gene, the Balikun breed possessed the highest MCN of 35, whereas the Guizhou breed only had a MCN of 2. The *ETY4* gene had the highest MCN tested with a range from 10 in Guizhou horses to 76 in Ningqiang horses. The MCN of the *YIR2* gene ranged from 4 to 26 (Table 3). The multi-copy portion of mammalian MSYs may share very little direct sequence homology between species, but is surprisingly consistent in function (Skaletsky et al., 2003; Hughes et al., 2010). Therefore, these multi-copy MSY genes, with testis-specific or predominantly testis related expression, are most frequently related to the testis and possibly spermatogenesis and male fertility related functions. For example, the *RBMY1* functional copy dosage is positively correlated with sperm motility, and dosage insufficiency is an independent risk factor for asthenozoospermia; therefore, comprehending the roll of CNVs in this gene is fundamental for understanding the cause of infertility (Chang et al., 2013; Yan et al., 2017). Other than the genes mentioned in this study, CNVs in different male-specific genes were found in mammals. Nine gene or gene families of human Y chromosome showed CNVs. These included the partial deletions of the TSPY cluster and the AZFc region which may influence spermatogenesis and a novel complex duplication of the AZFa region (Wei et al., 2015). Two Y-linked genes (*HSFY* and *ZNF280BY*) of swamp buffalo also showed abundant CNVs (Zhang et al., 2017).

The development of different modern horse breeds and various Y chromosome lineages is a reflection of human selection and environmental adaptation and occurs much later than the domestication of the species (Vila et al., 2001; Wallner et al., 2017). Variations on the Y chromosome are important tools to analyze both Y chromosome lineages and

domestication. In bulls, a genetic study on CNVs of Y-linked gene families in two ancestral Y-lineages has evaluated the effect of the number of Y-lineages on male reproduction and other traits (Yue et al., 2015). In horses, five MSY haplotypes have been identified by two Y-single nucleotide polymorphisms (SNPs) and one Y-indel (Han et al., 2015), and 42 other MSY haplotypes have been determined by 158 variants within domestic horses (Felkel et al., 2018); this suggests much higher diversity in Asian horses than in European breeds.

The CNV revealed a diverse distribution pattern among Chinese breeds in this study. The Guizhou was the breed which displayed the lowest CNV for the eight ECAY genes (Table S3–S10). Recent studies indicate that distribution of CNV regions may be shaped by natural selection (Cooper et al., 2007). This seems fitting, as most Guizhou horses are distributed in remote mountainous areas of Guizhou province and, due to the difficulty involved with transportation and occlusive conditions, rarely hybridize with other horse breeds. Therefore, it is possible that the gene copy numbers of Guizhou horses were relatively lower than other breeds before domestication. The results of this study support the assumption that CNVs might have been conserved for a long time and then passed on during the domestication of the horse (Metzger et al., 2013).

We suggest that other methods, such as array comparative genomic hybridization (array-CGH) (Wei et al., 2015; Shi et al., 2018) and the AccuCopy<sup>®</sup> assay method (Yan et al., 2017) could be applied to increase the accuracy of CNVs detection of ECAY genes. More functional analyses regarding the relationship between CNVs on ECAY and fertility should be investigated in the future.

#### 4 Conclusion

In this study, we first investigated the CNVs of eight Y chromosome genes in Chinese horses. The *EIF1AY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, *ETY4*, *UBE1Y*, *SRY*, and *YIR2* were multi-copy with MCNs of 1, 3, 8, 9, 26, 7, 1, and 12. The CNVs of Y chromosome genes showed different distribution patterns among Chinese horse breeds, indicating a natural selection effect on horse evolution and CNV formation.

**Data availability.** The measurement data involved in this study are available upon request to the authors.

**The Supplement related to this article is available online at <https://doi.org/10.5194/aab-61-263-2018-supplement>.**

**Author contributions.** RHD and HYH conceived the experiment; HYH carried out the experiment; HYH and XZ analyzed the data; HYH wrote the manuscript. X CZ, and XTX collected the samples and extracted genomic DNA. CZL revised the manuscript. All co-authors read and approved the final manuscript.

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

**Acknowledgements.** This work was supported by the National Natural Science Foundation of China (81270439).

Edited by: Steffen Maak

Reviewed by: three anonymous referees

## References

- Chang, T. C., Yang, Y., Retzel, E. F., and Liu, W. S.: Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development, *P. Natl. Acad. Sci. USA*, 110, 12373–12378, 2013.
- Cooper, G. M., Nickerson, D. A., and Eichler, E. E.: Mutational and selective effects on copy-number variants in the human genome, *Nat. Genet.*, 39, S22–S29, 2007.
- Dunn, O. J.: Multiple comparisons among means, *J. Am. Stat. Assoc.*, 56, 52–64, 1961.
- Felkel, S., Vogl, C., Rigler, D., Jagannathan, V., Leeb, T., Fries, R., Neuditschko, M., Rieder, S., Velie, B., Lindgren, G., Rubin C. J., Schlotterer C., Rattai T., Brem G., Wallner B.: Asian horses deepen the MSY phylogeny, *Anim. Genet.*, 49, 90–93, 2018.
- Fernández, R., Barragán, M. J. L., Bullejos, M., Marchal, J. A., Martínez, S., de la Guardia, R. D., and Sánchez, A.: Mapping the SRY gene in *Microtus cabrerai*: a vole species with multiple SRY copies in males and females, *Genome*, 45, 600–603, 2002.
- Geraldes, A. and Ferrand, N.: A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene, *Genet. Sel. Evol.*, 38, 313–320, 2006.
- Hamilton, C. K., Favetta, L. A., Di Meo, G. P., Floriot, S., Perucatti, A., Peippo, J., Kantanen, J., Eggen, A., Iannuzzi, L., and King, W. A.: Copy number variation of testis-specific protein, Y-encoded (TSPY) in 14 different breeds of cattle (*Bos taurus*), *Sex. Dev.*, 3, 205–213, 2009.
- Hamilton, C. K., Verduzco-Gómez, A. R., Favetta, L. A., Blondin, P., and King, W. A.: Testis-specific protein Y-encoded copy number is correlated to its expression and the field fertility of Canadian Holstein bulls, *Sex Dev.*, 6, 231–239, 2012.
- Han, H., Qin, Z., Gao, K., Yue, X., Zhang, T., Dang, R., Lan, X., Chen, H., and Lei, C.: Y-Single Nucleotide Polymorphisms Diversity in Chinese Indigenous Horse, *Asian Austral. J. Anim.*, 28, 1066–1074, 2015.
- Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., van Daalen, S. K., Minx, P. J., Fulton, R. S., McGrath, S. D., Locke, D. P., Friedman, C., Trask, B. J., Mardis, E. R., Warren, W. C., Repping, S., Rozen, S., Wilson, R., and Page, D. C.: Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content, *Nature*, 463, 536–539, 2010.
- Justel, A., Peña, D., and Zamar, R.: A multivariate Kolmogorov–Smirnov test of goodness of fit, *Stat. Probabil. Lett.*, 35, 251–259, 1997.
- Lévy, N., Navarro, A., Bishop, C. E., and Mitchell, M. J.: The ubiquitin-activating enzyme E1 homologous genes on the mouse Y chromosome (*Ube1y*) represent one functional gene and six partial pseudogenes, *Mamm. Genome*, 11, 164–168, 2000.
- Mann, H. B. and Whitney, D. R.: On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.*, 18, 50–60, 1947.
- Metzger, J., Philipp, U., Lopes, M. S., da Camara Machado, A., Felicetti, M., Silvestrelli, M., and Distl, O.: Analysis of copy number variants by three detection algorithms and their association with body size in horses, *BMC Genomics*, 14, 487, <https://doi.org/10.1186/1471-2164-14-487>, 2013.
- Mitchell, M. J., Woods, D. R., Tucker, P. K., Opp, J. S., and Bishop, C. E.: Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1, *Nature*, 354, 483–486, 1991.
- Paria, N., Raudsepp, T., Pearks Wilkerson, A. J., O'Brien, P. C., Ferguson-Smith, M. A., Love, C. C., Arnold, C., Rakestraw, P., Murphy, W. J., and Chowdhary, B. P.: A gene catalogue of the euchromatic male-specific region of the horse Y chromosome: comparison with human and other mammals, *PLoS One*, 6, e21374, <https://doi.org/10.1371/journal.pone.0021374>, 2011.
- Pearks Wilkerson, A. J., Raudsepp, T., Graves, T., Albracht, D., Warren, W., Chowdhary, B. P., Skow, L. C., and Murphy, W. J.: Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome, *Genomics*, 92, 329–338, 2008.
- PREMIER Biosoft International, CA, USA, available at: <http://www.premierbiosoft.com/>, last access: 20 June 2018.
- Quilter, C. R., Blott, S. C., Mileham, A. J., Affara, N. A., Sargent, C. A., and Griffin, D.: A mapping and evolutionary study of porcine sex chromosome genes, *Mamm. Genome*, 13, 588–594, 2002.
- Raudsepp, T., Santani, A., Wallner, B., Kata, S. R., Ren, C., Kata, S. R., Santani, A., Skow, L. C., and Chowdhary, B. P.: A detailed physical map of the horse Y chromosome, *P. Natl. Acad. Sci. USA*, 101, 9321–9326, 2004.
- Rohozinski, J., Agoulnik, A. I., Boettger-Tong, H., and Bishop, C. E.: Successful targeting of mouse Y chromosome genes using a site-directed insertion vector, *Genesis*, 32, 1–7, 2002.
- Sambrook, J. and Russell, D. W.: Molecular cloning: A laboratory manual, translated by: Huang, P. T., Science Press, Beijing, China, 2002.
- Shapiro, S. S., Wilk, M. B.: An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591–611, 1965.
- Shi, W., Massaia, A., Louzada, S., Banerjee, R., Hallast, P., Chen, Y., Bergström, A., Gu, Y., Leonard, S., Quail, M. A., Ayub, Q., Yang, F., Tyler-Smith, C., and Xue, Y.: Copy number variation arising from gene conversion on the human Y chromosome, *Hum. Genet.*, 137, 73–83, 2018.
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S. F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfling, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S. P.,

- Waterston, R. H., Wilson, R. K., Rozen, S., and Page, D. C.: The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, *Nature*, 423, 825–837, 2003.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.: MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, 28, 2731–2739, 2011.
- Trask, B. J., Allen, S., Massa, H., Fertitta, A., and Sachs, R.: Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization, *Cold Spring Harb. Sym.*, 58, 767–775, 1993.
- Tüttelmann, F., Simoni, M., Kliesch, S., Ledig, S., Dworniczak, B., Wieacker, P., and Röpke, A.: Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome, *PLoS One*, 6, e19426, <https://doi.org/10.1371/journal.pone.0019426>, 2011.
- Vila, C., Leonard, J. A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R. K., and Ellegren, H.: Widespread origins of domestic horse lineages, *Science*, 291, 474–477, 2001.
- Vodicka, R., Vrtel, R., Dusek, L., Singh, A. R., Krizova, K., Svacinova, V., Horinova, V., Dostal, J., Oborna, I., Brezinova, J., Sobek, A., and Santavy, J.: TSPY gene copy number as a potential new risk factor for male infertility, *Reprod. Biomed. Online*, 14, 579–587, 2007.
- Wallner, B., Palmieri, N., Vogl, C., Rigler, D., Bozlak, E., Druml, T., Jagannathan, V., Leeb, T., Fries, R., Tetens, J., Thaller, G., Metzger, J., Distl, O., Lindgren, G., Rubin, C.-J., Andersson, L., Schaefer, R., McCue, M., Neuditschko, M., Rieder, S., Schlotterer, C., Brem, G.: Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. *Curr. Biol. Cb*, 27, 2029–2035, 2017.
- Wei, W., Fitzgerald, T. W., Ayub, Q., Massaia, A., Smith, B. B., Dominiczak, A. A., Morris, A. A., Porteous, D. D., Hurler, M. E., Tyler-Smith, C., and Xue, Y.: Erratum to: Copy number variation in the human Y chromosome in the UK population, *Hum. Genet.*, 134, 789–800, 2015.
- Wilhelm, D., Palmer, S., and Koopman, P.: Sex determination and gonadal development in mammals, *Physiol. Rev.*, 87, 1–28, 2007.
- Yan, Y., Yang, X., Liu, Y., Shen, Y., Tu, W., Dong, Q., Yang, D., Ma, Y., and Yang, Y.: Copy number variation of functional RBMY1 is associated with sperm motility: an azoospermia factor-linked candidate for asthenozoospermia, *Hum. Reprod.*, 32, 1521–1531, 2017.
- Yang, H., Ma, Y. H., Li, B., and Dugarjaviin, M.: Progress on horse genome project, *Hereditas*, 32, 211–218, 2010.
- Yue, X. P., Chang, T. C., DeJarnette, J. M., Marshall, C. E., Lei, C. Z., and Liu, W. S.: Copy number variation of PRAMEY across breeds and its association with male fertility in Holstein sires, *J. Dairy Sci.*, 96, 8024–8034, 2013.
- Yue, X. P., Dechow, C., and Liu, W. S. A limited number of Y chromosome lineages is present in North American Holsteins, *J. Dairy Sci.*, 98, 2738–2745, 2015.
- Zhang, X., Han, H., Zhang, T., Sun, T., Xi, Y., Chen, N., Huang, Y., Dang, R., Lan, X., Chen, H., and Lei, C.: HSFY and ZNF280BY show copy number variations within 17 water buffalo populations, *Anim. Genet.*, 48, 221–224, 2017.