Population genetic copy number variation of \textit{CHKB, KLF6, GPC1} and \textit{CHRM3} genes in Chinese domestic yak (\textit{Bos grunniens}) breeds

Habtamu Abera Goshu, Wu Xiao Yun, Min Chu, Bao Pengjia and Ping Yan
Population genetic copy number variation of CHKB, KLF6, GPC1 and CHRM3 genes in Chinese domestic yak (Bos grunniens) breeds

Habtamu Abera Goshu¹,²*, Wu Xiao Yun¹, Min Chu¹, Bao Pengjia¹ and Ping Yan¹*

Abstract: Single nucleotide polymorphisms (SNPs) are common markers used for population genetics studies. Recently, copy number variations (CNVs) have been used to study genetic and phenotypic diversity within and among populations. Therefore, studies of the genetics of CNVs are important in the context of understanding evolutionary changes and genomic selection based upon genetic and phenotypic variation. The aim of this study was to analyse the distribution of the choline kinase beta (CHKB), Krüpple-like factor 6 (KLF6), glypican 1 (GPC1) and cholinergic receptor muscarinic 3 (CHRM3) genes in five representative Chinese domestic yak breeds. The data generated by qPCR was transformed into log₂ ratio and analysed using GraphPad (PRISM). The results show that the CNVs of CHKB, KLF6, and GPC1 genes presented more copy number losses in Tianzhu, Gannan and Plateau yak populations compared with the Datong and polled yak having relatively more copy number gains. However, the CHRM3 CNV showed more copy number gains in five yak populations. Therefore, these results indicate that there are relatively more copy number losses (deletion) in the yak populations; supporting the hypothesis that log₂ ratio is more powerful at detecting loss than gain in copy number types. Taken together, these data provide information on the application.
genome CNVs in population genetics and suggest that the CNVs of the genes could exert a significant effect on phenotypic differences in yak populations.

**Subjects:** Agriculture & Environmental Sciences; General Science; Bioinformatics; Developmental Biology; Endocrinology; Genetics; Molecular Biology; Agriculture; Forensic Science; Food Science & Technology

**Keywords:** CNVs; population genetics; genes; domestic yak; breeds

1. **Introduction**

Morphological and molecular markers are often used to distinguish different mammalian species. However, characterization of the available genomic resources is significantly important in the study of genetic diversity and for selection of superior yak breeds and possibly other bovid species. Yaks (*Bos grunniens*) are of the same genus to the two domesticated cattle, *Bos taurus* and *Bos indicus* with chromosomal number of 60. The domestic yak (*B. grunniens*) is common in the larger and harshest highlands of the “the roof of the globe” Qinghai-Tibet Plateau (QTP) (Wiener, Han, & Long, 2003). The long hair and large horns are peculiar phenotypic characteristics of the yak breeds (Liu et al., 2013). Indeed the yak has become strongly integrated into Tibetans’ socio-cultural life (Liang et al., 2016), and it occupies a significant economic role in these mountainous regions of Asia (Medugorac et al., 2017); providing milk, meat, transportation (“boat of the plateau”), hair, draught power, and fuel for Tibetans and nomadic dwellers (Wiener, Han, & Long, 2003). Yaks have a long history of domestication and selective breeding (Wiener et al., 2003). Also, a unique physiological feature of yak breeds is their ability to survive extremely harsh environments at high altitudes, overcoming hypoxia, severely cold winters, and cool moist summers, and using limited grazing resources (Wiener et al., 2003). A major issue in the understanding of the mechanisms behind the development of the yak breeds is lack of genome DNA sequence information; therefore, a study elucidating the population genetics of copy number variations in yaks could lead to significant economic gains with respect to the breeding objectives. A population genetics survey based on CNVs has been reported for cattle breeds (Liu et al., 2010; Xu et al., 2016). Several studies have revealed that the copy number of variable regions in the genome can shape genetic networks and regulate gene expression, thereby contributing to inter-individual phenotypic variability (Conrad et al., 2010; Freeman et al., 2006). Notably, the CNVs encompassed more actual nucleotide content per genome compared to SNPs, further underscoring the importance of CNVs in the evolution of genetic diversity (Redon et al., 2006).

Therefore, the population genetics of CNVs is important in the understanding of evolutionary changes and the application of genomic selection based on genetic and phenotypic variation within a given population. Copy number variation contributes to the creation of novel genes which may allow genotypic and phenotypic diversification of the repertoire of such genes in the response to changing environments (Iskow, Gokcumen, & Lee, 2012). For instance, a study by Lee et al. (2013) showed that there are a significant copy number variations present for the olfactory receptor (OR) genes in cattle, which is significant in determining individual and/or breed-specific differences in olfaction capacity. Human Amylase1 (*AMY1*) CNVs are thought to be driven by natural selection and may have potential roles in driving evolutionary biological changes (Fernández & Wiley, 2017). For instance, *AMY1* copy number is observed to be higher in populations that consume high-starch diets which are thought to have resulted from positive or directional selection and relatively lower numbers of copies are found in populations consuming lower starch diets due to genetic drift (Perry et al., 2007). Sjödin and Jakobsson (2012) in their review suggested that recombination, mutation, selection, and demography are major forces in population genetics of CNVs. Evolutionary processes, domestication and artificial selection are thought to be the main factors responsible for genomic changes including CNVs and phenotypic divergence in animal productions (Kijas et al., 2009; Novembre & Ramachandran, 2011; Parsch & Ellegren, 2013). The study of CNVs has become important and they are an abundant source of genetic and phenotypic
diversity (Liu et al., 2010), which is associated to complex disease and phenotypes (Stankiewicz & Lupski, 2010; Yang et al., 2007; Zhang, Youji, Yong, & Xingxu, 2015).

Genomic copy number variation of large segments DNA (at least 1 kb in size), are composed of both gains (insertions or duplications) and losses (deletions or null genotypes) in comparison with a designated reference genome sequence (Scherer et al., 2007). Genic CNV regions (CNVRs) have the potential to affect phenotypes variation, gene expression, and adaptation through the disruption of gene dosages (Weischenfeldt, Symmons, Spitz, & Korbel, 2013) and altering transcription levels of genes (Pollex & Hegele, 2007). CNVs have been well studied and linked to various phenotypic traits and diseases in humans and rodents (Almal & Padh, 2012; Cook & Scherer, 2008; Girirajan et al., 2013). The association between the CNVs and production traits of economic interest had been reported in swine (Chen et al., 2012). Iafrate et al. (2004) revealed that copy number variations are an important genome resource in humans.

In the past, little research has been conducted on yak CNVs, focusing mostly on surveying CNVRs using array comparative genomic hybridization (aCGH) (Zhang, Jia, et al., 2014), BovineHD Genotyping Bead Chip Array (Quanwe et al., 2014) and next-generation sequencing (NGS) (Zhang et al., 2016). In recent years, relative copy number distributions were determined for the CNVs of myosin heavy chain 3 (MYH3), mitogen-activated protein kinase 10 (MAPK10) and leptin receptor (LEPR) genes and revealed considerable diversity in genetic copy numbers in Chinese cattle (Liu et al., 2016; Shi et al., 2015; Xu et al., 2013). Previous studies have using a read depth approach in the genome resequencing analysis of CNVs for choline kinase beta (CHKB), Krüppel-like factor 6 (KLF6), glypican 1 (GPC1) and cholinergic receptor muscarinic 3 (CHRM3) genes has identified copy number variable regions in yak breeds (Zhang et al., 2016). Zhang et al. (2016) also suggested that the CNVRs are an important genetic resource for phenotypic and genotypic variations for both yak and other bovid species, and these earlier studies have provided the raw materials for the current study. To date, there is no report made on the study on population genetics of copy number variations of the CHKB, KLF6, GPC1 and CHRM3 genes in Chinese domestic yak breeds by using real-time PCR technique. The main aim this study was to analyse copy number variations of the CHKB, KLF6, GPC1 and CHRM3 genes in Chinese domestic yak breeds. The results obtained from this work will be an essential step towards deciphering the molecular mechanisms underlying trait variation among the domesticated yaks in Qinghai-Tibet Plateau (QTP) of China.

2. Materials and methods

2.1. Study areas and sample collections

The main goal of Chinese domestic yak breeding comprises enhanced production of meat and milk, whilst maintaining and conserving genetic diversity of yak populations. Five representative Chinese domestic yak breeds were selected to study, namely, the polled yak (Datong Yak Farm in Qinghai Province, altitude 3200 m), Tianzhu white yak (Tianzhu Tibetan Autonomous County in Gansu Province, altitude 3000 m), plateau yak (northern and southern Qinghai Province, altitude 3700–4700 m), Datong yak (Datong Yak Farm in Qinghai Province, altitude 3200 m), and Gannan yak (Gansu Province, bordering Sichuan and Qinghai, altitude 3300–4400 m). The origins of the five breeds are shown in Figure 1.

Fresh blood samples were collected from a total of 477 yaks (222 Datong yaks, 165 polled yaks, 30 Tianzhu yaks, 30 Gannan yaks, and 30 plateau yaks). The blood samples were taken from the jugular vein and collected in vacuum tubes (5 mL) containing anticoagulant citrate. All blood samples were stored at ~80°C prior to DNA extraction. The animals were in good physical condition and assumed to not be genetically related. All yaks were grazing on natural pastures without feed supplementation, under similar feeding conditions and management.
2.2. Genomic DNA isolation
Genomic DNA (gDNA) was purified using clot blood DNA purification kit (CWBIO, China, Beijing) and cleaned using the genomic DNA clean up (>10kb) of E.Z.N.A. microelute DNA clean-up Kit (Omega Bio-tek, Georgia, USA) following the manufacturer’s instructions. The concentration and quality of the DNA was determined using a NanoDrop Biophotometer 2000 (Thermo Fisher Scientific Inc.) and through electrophoresis on 1% agarose gels and stained with ethidium bromide.

2.3. Primer design and PCR amplification
Using bovine genomic sequences, primers were designed for CNV of KLF6, CHKB, GPC1 and CHRM3 genes using the Primer v5.0 software (Shi et al., 2015) and the qPCR primers designed (HTTP://bitesizebio.com/10041/a-step-by-step-guide-to-designing-qPCR-primers/) (Table 1). The primers

![Figure 1. Geographic map of the five yak breeds.](image)

**Table 1. Primers designed for qPCR based on bovine reference genome**

| Accession no. | Name of gene | Primers sequence (5’→3’) | Product length (bp) |
|---------------|--------------|---------------------------|---------------------|
| XM_005901807.2 | KLF6         | F: ATGCTCATGGGAAGGGTGTG   | 82                  |
|               |              | R: CTTGGCACCAGTGTCCTTC    |                     |
| NM_001206094.1| CHKB         | F: GCAGTCTCGGTTCAGTTCT    | 90                  |
|               |              | R: AATGCAAGGAGTCGGAGGTG   |                     |
| XM_014440020.1| GPC1         | F: GCTTTGCCTTTCCCTACGGG  | 86                  |
|               |              | R: ACATGTTGCGGTGTTTG     |                     |
| NM_174270.2   | CHRM3        | F: AGGGCTTCAAGATGGAAGCG  | 149                 |
|               |              | R: AGCTAGGATGCCCAGGGTTG  |                     |
| NM_001034034.2| BTF3         | F: AACCCAGGAAACTGCAAA   | 166                 |
|               |              | R: TCGGTGAAATGCCCTCTCG   |                     |
were designed so that the PCR product/amplicon size was <200bp (80–200 bp), melting temperature with ranges between 57 and 63°C (with a maximum difference of 3°C in the Tm’s of the two primers), length of primers (18–22 bp), GC content (50–60%) and having low self complementarily. Polymerase chain reactions (PCR) used for validating the amplification primers were performed in a total volume of 25 μL containing 50 ng/μL gDNA, GoTaq® Green master mix, 2×, 10 μM primer and nuclease-free water (ddH₂O). Thermal cycling conditions were as follows: an initial denaturaton step of 95°C for 2 min, then 35 cycles at 95°C for 1 min, 55–60°C for 1 min, 72°C for 1 min, 72°C for 5 min final extension and at 4°C final holding temp. PCR products were loaded directly into 1% agarose gels or non-denaturing TBE polyacrylamide gels. The amplification primers were also assessed using qPCR melting curve analysis.

### 2.4. Copy number variation analysis of KLF6, CHKB, GPC1 and CHRM3 genes

Primers designed for copy number variation is listed in Table 1. The relative copy number variations of the bovine KLF6, CHKB, GPC1 and CHRM3 genes were examined using quantitative real-time polymerase chain reaction (qPCR). Bovine basic transcription factor 3 (BTF3) was used as a normal (two copy or diploid) internal reference gene for genomic qPCR because in the Database of Genomic Variants, neither CNV nor segmental duplication has been described (Bickhart et al., 2012). The qPCR reaction was carried out using SYBR® Premix Ex Taq™ II (Tli RNase H Plus) in 25 μL total reaction mixture containing 50 ng of gDNA, 12.5 μL SYBR Premix Ex Taq II (2×) and 10 pmol of primers (TAKARA BIO INC.). The thermal cycling profile of qPCR experiments was: one cycle of 95°C for 1 min followed by 39 cycles at 95°C for 10 s, annealing at 60°C for 30 s, extension at 68°C for 10 s. During qPCR thermal cycles, primers were also examined using melting curve (dissociation) steps and no template control (a minus sample control) included. Experiments were repeated three times, and data used for statistical analyses was obtained using the average mean values and standard deviation.

### 2.5. Statistical analysis

Copy number was calculated based on average threshold cycle (ΔΔCt) methods (Schmittgen & Livak, 2008), related to reference gene BTF3 (Liu et al., 2010). The final result of the copy number was defined using the formula 2 × 2−ΔΔCt (Bae et al., 2010), where ΔΔCt was the difference in threshold cycles for each sample normalized to reference gene (Bae et al., 2010). The normal value of DNA copy number in tested animals as diploid and all measured target DNA copy number values were adjusted relative to the two copies (Yim et al., 2011). The populations genetics of different CNV types of KLF6, CHKB, GPC1 and CHRM3 genes were grouped into gain (>0.5), loss (<−0.5) and normal (<±0.5) of the CNVs based on log₂ ratio (log₂2−ΔΔCt) relative to control sample previously described elsewhere (Liu et al., 2016; Shi et al., 2015; Xu et al., 2013). The pairwise comparison and scatter plots were determined using GraphPad (PRISM5) statistical software using the value obtained from the log2 ratio (log₂2−ΔΔCt).

### 3. Results

Yak breeding is an economically important sector and supports millions of people for milk and beef; the breeding of yak has been supported intensively by farm animal’s genetic resources, an institute of Chinese academy of agricultural science. Recently, much attention has been paid to the investigation of genomic markers related to population genetics due to a vital role in the livestock selection. As shown in Figure 2A, the CNV of CHKB gene in the Tianzhu, Gannan and plateau yak individuals had greater copy number loss while Datong and polled yak individuals had greater copy number gains. Significant differences were found in the pairwise comparison between Polled versus Tianzhu (P = 0.0001), Gannan (P = 0.0001); Datong versus Tianzhu, Gannan, and Gannan versus Plateau yak (P = 0.0001, P < 0.01). No significant differences were detected between Datong versus Polled (P = 0.35), Plateau (P = 0.013); Polled versus Plateau (P = 0.08) and Tianzhu versus Gannan (P = 0.22), Plateau (P = 0.039) yak in terms of the CNVs of the CHKB gene (Table 3). Using both GraphPad (PRISM) and log₂ (ratio) analysis, the CNV of CHKB gene differentiates the yak breeds in two distinct groups based on provincial locations or sites, which include Datong, polled and plateau yak in Qinghai province and Tianzhu and Gannan Gansu province. For all breeds, an overall frequency of CHKB
copy number loss (copy number 0 or 1), normal level (copy number 2), and gain (more than two copies) was 40.8%, 29.5%, and 29.7%, respectively (shown in Figure 2B). Approximately 80% of Tianzhu yaks, 46.7% of plateau yaks and 80% of Gannan yaks breeds exhibited loss copy number genotypes, while about 41% of Datong and 23% of polled yak demonstrated copy number gain genotypes (Figure 2B, Table 2). The normal copy number genotype was found at higher frequency in the polled (35%) and Datong (29.5%) yak populations.

As presented in Figure 3A, the CNV of the KLF6 gene showed greater copy number loss in Tianzhu, Gannan and plateau yak individuals, whilst greater copy number gain was observed in Datong and polled yak individuals. There were highly significant differences in the pairwise comparison between Polled versus Tianzhu, Gannan, Datong, and Plateau yak ($P = 0.0001$, Table 5), however, no significant

### Table 2. Distribution of the CHKB gene CNV types among five yak breeds sampled from across China

| Breeds          | Sample size | CNV types     |
|-----------------|-------------|---------------|
|                 |             | Loss | Gain | Normal |
| Datong yak      | 222         | 29.3% (65) | 41% (91) | 29.7% (66) |
| Polled yak      | 163         | 41.1% (67) | 23.9% (39) | 35% (57) |
| Tianzhu yak     | 30          | 80.0% (24) | 10.0% (3) | 10.0% (3) |
| Plateau yak     | 30          | 46.7% (14) | 23.3% (7) | 30.0% (9) |
| Gannan yak      | 30          | 80.0% (24) | 3.3% (1) | 16.7% (5) |
| Overall         | 475         | 40.8% (194) | 29.7% (141) | 29.5% (140) |
difference was observed between Datong versus Plateau ($P = 0.10$), Tianzhu ($P = 0.35$), Tianzhu versus Gannan ($P = 0.085$), plateau ($P = 0.57$) and Gannan versus plateau ($P = 0.20$) yak breeds (Table 5). However, the frequency of copy number loss in KLF6 stood at 31.7% (0 or 1 copy number), normal 28.5% (2 copy numbers) and gain 39.8% (greater two copy numbers) as presented in Figure 3B and Table 4. Among the populations, 56.7% of Tianzhu, 56.7% of plateau and 60% of Gannan yak breeds contained only loss copy number genotypes. However, 36.5% of Datong and 62.1% of Polled yak contained gain copy number genotypes (Figure 3B and Table 4).

As shown in Figure 4A, the GPC1 gene CNV in the Tianzhu, Gannan and Plateau yak individuals exhibited greater copy number loss while polled yak and Datong yak individuals exhibited copy number gain. The pairwise comparison between Datong versus polled, Tianzhu, Gannan, Plateau yak, and Polled versus Tianzhu, Gannan, Plateau yak breeds was highly significant with a $P$-value of 0.0001 ($P < 0.01$, Table 7) and a non significant difference was observed in Tianzhu versus Gannan yak breeds ($P = 0.14$), plateau versus Tianzhu ($P = 0.58$), and Gannan versus plateau ($P = 0.11$) yak breeds. Across all breeds, there was an overall frequency of GPC1 copy number (Figure 4B and Table 6) loss (copy number 0 or 1), normal level (copy number 2), and gain (more than two copies) of 39%, 20.1%, and 40.9%, respectively. Among them, 80% of Tianzhu, 80% of Plateau, 96.7% of Gannan and 44.8% polled yak breeds contained only loss copy number genotype while 59.9% of Datong yak contained only gain copy numbers genotypes (Figure 4B and Table 6).

As shown in Figure 5A, the CNV of the CHRM3 gene in polled, Datong, Tianzhu, Gannan and plateau yak individuals exhibited more copy number gains. No significant differences were observed between the five yak breeds when compared with individual populations (Table 9). Among the populations, 70% of Tianzhu, 66.7% of Gannan, 58.1% of Datong, 66.7% of polled and 60% of plateau yak breeds contained only gain copy numbers genotypes (Figure 5B and Table 8). As shown in Figure 5A, the five types of CHRM3 CNVs showed similar distributions.

| Table 3. Pairwise comparison of CNV of the CHKB gene in five Chinese domestic yak breeds |
|---------------------------------|--------|--------|--------|--------|-----------------|
| Yak breeds                      | Mean diff. | $P$-value | $q$ or $t$ ratio | Significance | 95% CI of diff. |
| Polled vs. Datong               | 0.21    | 0.35    | 3.3     | NS     | −0.034 to 0.46  |
| Polled vs. Tianzhu              | 0.8     | 0.0001  | 6.5     | ***    | 0.32 to 1.3     |
| Polled vs. Gannan               | 1.1     | 0.0001  | 9       | ***    | 0.63 to 1.6     |
| Polled vs. Plateau              | 0.25    | 0.08    | 2.1     | NS     | −0.23 to 0.73   |
| Datong vs. Tianzhu              | 1       | 0.0001  | 8.4     | ***    | 0.54 to 1.5     |
| Datong vs. Gannan               | 1.3     | 0.0001  | 10.9    | ***    | 0.85 to 1.8     |
| Datong vs. Plateau              | 0.46    | 0.13    | 3.9     | NS     | −0.006 to 0.94  |
| Tianzhu vs. Gannan              | 0.31    | 0.22    | 1.9     | NS     | −0.31 to 0.94   |
| Tianzhu vs. Plateau             | −0.55   | 0.039   | 3.4     | NS     | −1.2 to 0.078   |
| Gannan vs. Plateau              | −0.86   | 0.001   | 5.4     | **     | −1.5 to −0.23   |

*** and ** are the variance significant at $P < 0.01$ and $P < 0.05$. NS: non significant.
4. Discussion

Copy number variations contribute significantly to the genetic makeup of a species and have become a valuable resource for investigating population genetics of animal and plant species. CNVs surveying has been completed for most of the domestic animals including yak breeds (Zhang et al., 2016). The considerable genetic differences found between the Black Angus and Holstein cattle in CNVs have been investigated using read depth platforms (Stothard et al., 2011). Similarly, Fadista, Thomsen, Holm, and Bendixen (2010) reported that a high diversity of CNVs was detected in beef and dairy cattle breeds. Furthermore, Liu et al. (2010) and Hou et al. (2011) confirmed that differences in the CNV frequencies across diverse breeds of cattle; suggesting that some cattle CNVs are likely to have arisen independently within breeds and could contribute to observed

| Breeds         | Sample size | Gain     | Loss     | Normal   |
|----------------|-------------|----------|----------|----------|
| Datong yak     | 222         | 36.5% (81)| 31.5% (70)| 32% (71) |
| Polled yak     | 165         | 61.2% (101)| 17.6% (29)| 21.2% (35)|
| Tianzhu Yak    | 30          | 10.0% (3) | 56.7% (17)| 33.3% (10)|
| Plateau Yak    | 30          | 6.7% (2)  | 56.7% (17)| 36.7% (11)|
| Gannan yak     | 30          | 10.0% (3) | 60.0% (18)| 30% (9)   |
| Overall        | 477         | 39.8% (190)| 31.7% (151)| 28.5% (136)|
breeding differences. Similarly, Bickhart et al. (2012) suggested that CNV differences were found between taurine and indicine cattle breeds reflected their breeding history. Hou et al. (2011) also found that differences between CNV variations, in some cattle breeds, is possibly due to genetic bottlenecks, caused through both selection and genetic drift. Interestingly, a comparative analysis revealed potential contributions of CNVs to the process of yak domestication and significantly different copy number were present between domestic and wild yak animals (Zhang et al., 2016).

However, most of the studies focusing on the detection CNVRs in domesticated and wild yak population are based on whole genomic sequencing, while, CNVs-based population genetics of yak breeds diversity have not been adequately investigated.

As shown in Figures 2, 3, and 4A, the CNVs of the CHKB, KLF6 and GPC1 genes had greater frequency copy number losses in the Tianzhu, Gannan and plateau yak populations while Datong and polled yak showed a higher frequency of copy gains. However, the CHRM3 CNV showed a significantly higher frequency of gain in the five yak populations. These results concur with a previous finding in which Upadhyay et al. (2017) observed that a higher number of genic deletions, as well as higher cumulative length of the genome under deletion in the British and Irish cattle breeds, compared to breeds from across European cattle. The results obtained in Upadhyay’s study could be explained through changes in the past effective population size, gene flow, selection process (purifying selection) and genetic drift, all which may contribute to shaping CNV diversity that existed between the different European cattle populations (Upadhyay et al., 2017).

The present results have suggested that the CNVs in the CHKB, KLF6, GPC1 and CHRM3 genes are associated with yak breed formation. Our results showed that Tianzhu yaks had high genetic variability within a population, due to intensive selection program for white hair individuals (Zhang et al., 2008); for the reason that the white hair is easily dyed into different colors and highly valued in markets, the nomadic pastoralists migrated from Qinghai province and started to select and breeding pure white herds about 120 years ago in Tianzhu County (Wiener et al., 2003). In addition, Yue, Liang, Liang, and Li (2016) observed that the Tianzhu white yak possessed relatively high genetic diversity resulting from hybridization with Bos Taurus. Our results correlated positively to previous findings in which the genetically isolated small.

### Table 5. Pair-wise comparison of CNV of the KLF6 gene in five Chinese domestic yak breeds

| Yak breeds          | Mean diff. | P-value | q or t ratio | Significance | 95% CI of diff. |
|---------------------|------------|---------|--------------|--------------|-----------------|
| Datong vs. Polled   | −0.80      | 0.0001  | 10           | ***          | −1.09 to −0.52  |
| Datong vs. Tianzhu  | 0.19       | 0.35    | 1.4          | NS           | −1.36 to 2.75   |
| Datong vs. Gannan   | 0.51       | 0.12    | 3.7          | NS           | −0.032 to 1.05  |
| Datong vs. Plateau  | 0.30       | 0.1     | 2.2          | NS           | −0.24 to 0.85   |
| Polled vs. Tianzhu  | 0.99       | 0.0001  | 7            | ***          | 0.44 to 1.5     |
| Polled vs. Gannan   | 1.3        | 0.0001  | 9.3          | ***          | 0.76 to 1.9     |
| Polled vs. Plateau  | 1.1        | 0.0001  | 7.8          | ***          | 0.55 to 1.7     |
| Tianzhu vs. Gannan  | 0.32       | 0.085   | 1.7          | NS           | −0.4 to 1.0     |
| Tianzhu vs. Plateau | 0.11       | 0.57    | 0.61         | NS           | −0.06 to 0.83   |
| Gannan vs. Plateau  | −0.21      | 0.20    | 1.1          | NS           | −0.93 to 0.51   |

***The variance significant at P < 0.01. NS: non significant.
populations have been found to accumulate an abundance of deletion type CNVs (Upadhyay et al., 2017). Similarly, within Chinese cattle breeds, most individuals have higher frequencies of copy number losses in the MYH3, LEPR and MAPK10 genes (Liu et al., 2016; Shi et al., 2015; Xu et al., 2013). Interestingly, an earlier study of CNVs in the Bos taurus UL16-binding protein (ULBP) gene found that the highest copy number gain and loss were detected in Limousin and Angus cattle breeds, respectively, due to a combination of genetic bottlenecks in some breeds.

Table 6. Distribution of the GPC1 gene CNV types among five yak breeds sampled from across China

| Breeds          | Sample size | CNV type       |
|-----------------|-------------|----------------|
|                  |             | Gain     | Loss     | Normal       |
| Datong yak      | 222         | 59.9% (133)| 15.8% (35)| 24.3% (54)   |
| Polled yak      | 165         | 35.2% (58)| 44.8% (74)| 20% (33)     |
| Tianzhu Yak     | 30          | 13.3% (4) | 80.0% (24)| 6.7% (2)     |
| Plateau Yak     | 30          |             | 80.0% (24)| 20.0% (6)    |
| Gannan yak      | 30          |             | 96.7% (29)| 3.3% (1)     |
| Overall         | 477         | 40.9% (195)| 39% (186)| 20.1% (96)   |
selection, and genetic drift (Liu et al., 2010). Population differentiation in this study was discovered using log₂ ratio resultant to copy number gains, normal and losses (Fadista et al., 2010; Liu et al., 2016; Olshen, Venkatraman, Lucito, & Wigler, 2004). The differences in CNV frequencies in the five yak breeds, may reflect the breed variation between individual populations (Lehnert et al., 2007; Yang et al., 2017) as differences in CNV frequency amongst cattle breeds is related to breed formation and adaptation (Liu et al., 2010). The data in this study was analyzed using log₂ ratio to differentiate the genetic diversity between yak breeds, which appears to be more sensitive than the detection of homozygous deletions (Liu et al., 2016; Olshen et al., 2004).

According to Fadista et al. (2010) findings, the biological and technical detection methods of CNVs are biased towards deletions, and more deletions than duplications have been produced by mechanisms of non-allelic homologous recombination (NAHR) (Turner et al., 2008). The results obtained in this present research, is in agreement with the findings by Fadista et al. (2010), in which log₂ ratio exhibited more powerful detection of loss than gain copy number types. Lee et al. (2013) revealed that significant copy number variations in the OR gene were found in Korean native cattle, Black Angus and Holstein, indicating that there is selection pressure maintaining the integrity of OR genes. In addition, the CNVRs of the OR gene is different in Angus, Red Angus, and Holstein cattle and CNVRs of the sterol carrier protein 2 (SCP2) gene is different in Gelbvieh and Red Angus (Liu et al., 2010). Perry et al. (2007) suggested that natural selection has contributed in shaping AMY1 copy number variation in both Japanese and Yakut populations. In human, high copies of the AMY1 number were reported as compared to other primates, which have been attributed to the dietary shift that occurred early in the hominid evolutionary history (Perry et al., 2007). The human UDP glucuronosyltransferase family two member B17 (UGT2B17) gene varies in copy number from zero to two per individual from Africa, Europe, and East Asia people (Xue et al., 2008). This variation is a result of positive selection and evolutionary change over the last centuries. The deletion in the UGT2B17 CNVs is relatively rare in African and European populations and common in East Asian populations (Xue et al., 2008). The CNVs of the UGT2B17 and AMY1 genes have been hypothesised to have a crucial role in population differentiation (Fernández & Wiley, 2017; Perry et al., 2007; Xue et al., 2008). Xu et al. (2016) indicated that CNV deletions showed large differences in frequency across diverse

table 7. Pair-wise comparison of CNV of the GPC1 gene in five Chinese domestic yak breeds

| Yak breeds       | Mean diff. | P-value | q or t ratio | Significance | 95% CI of diff. |
|------------------|------------|---------|--------------|--------------|-----------------|
| Datong vs. Polled| 1.2        | 0.0001  | 12.4         | ***          | 0.80–1.5        |
| Datong vs. Tianzhu| 2.1       | 0.0001  | 11.6         | ***          | 1.4–2.8         |
| Datong vs. Gannan | 2.7        | 0.0001  | 15.3         | ***          | 2.0–3.4         |
| Datong vs. Plateau| 2.2       | 0.0001  | 12.3         | ***          | 1.5–2.9         |
| Polled vs. Tianzhu | 0.88   | 0.0001  | 4.8          | ***          | 0.17–1.6        |
| Polled vs. Gannan | 1.6        | 0.0001  | 8.6          | ***          | 0.85–2.3        |
| Polled vs. Plateau| 1         | 0.0001  | 5.6          | ***          | 0.31–1.7        |
| Tianzhu vs. Gannan | 0.67      | 0.14    | 2.8          | NS           | −0.25 to 1.6    |
| Tianzhu vs. Plateau| 0.13      | 0.58    | 0.58         | NS           | −0.78 to 1.1    |
| Gannan vs. Plateau| −0.54     | 0.11    | 2.3          | NS           | −1.5 to 0.38    |

*** The variance significant at P < 0.01. NS: non significant.
groups of cattle. Intriguingly, Olson (1999) mentioned that gene loss was a major pattern of molecular evolution and is a possible evolutionary pressure that may explain the geographic pattern (Ko et al., 2012), hence, this combination may play a prominent role in molecular evolution and function (Gamazon & Stranger, 2015). Correspondingly, Zarrei, MacDonald, Merico, and Scherer (2015) found that CNVs caused by deletions and duplications in the genome were the predominant causes of the variability among populations, with duplications and deletions occurring as results of positive and negative selective pressure in the individual populations, respectively.
The distribution of CNVs for the CHKB, KLF6 and GPC1 genes showed a similar pattern across all five yak breeds. Copy number distributions of the CHKB and GPC1 genes were not in agreement to previous analysis of Jincheng, Zhihua, Sujun, and Yuping (2006) using a analysis of microsatellite markers, RAPD, AFLP, chromosomes and blood protein polymorphisms methods showing that Datong and the Tianzhu white yaks clustered together. Copy numbers of the KLF6 and CHRM3 genes, on the other hand, were consistent with previous findings of Jincheng et al. (2006). A high degree of genetic diversity in the Tibetan yak populations in relation to mtDNA COIII gene has been reported (Song et al., 2015). Moreover, the results of the distribution of copy number of the CHKB and GPC1 genes strongly supported the hypothesis proposed by Bao (2016) that the Datong and polled clustere together while Gannan yak clustere into a separate clade, consistent with the known yak breed history. The current results support the work of Liang et al. (2016) who showed that the pairwise distances between Datong and polled yaks were not larger than within population and there was no significant differentiation between the two breeds when the phylogenetic tree was constructed using whole genome-wide association study. However, SNPs studies have confirmed that most polled yak individuals were heterozygous compared to Datong individuals (Liu et al., 2013). A recent study has demonstrated the detection of five different haplotypes within the Datong and polled yak populations, and the haplotype frequency that was differently distributed between the polled and Datong yaks (Liu et al., 2013). However, the level of nucleotide diversity is higher in the polled yaks than in the Datong yaks (Liu et al., 2013). Hence, Liang et al. (2016) suggested that expression variations rather than structural variations in the encoded protein probably contributed to the polled phenotype.

However, unlike previous studies using mitochondrial proteomics (Bao, 2016) and other markers (Jincheng et al., 2006) the breeds from the same region or ecotype did not cluster together, for instance, Datong, polled and plateau yak in the CNVs of the KLF6 and CHRM3 genes (Figures 3 and 5). This indicated that, there was relatively low level of divergence in copy number of the KLF6 and CHRM3 genes within the different yak breeds. On the other hand, Datong, plateau, Tianzhu, and Gannan yak populations were all clustered together, whereas the polled population clustered alone in CNVs of the KLF6. The plateau, Tianzhu, and Gannan yak populations were clustered together

| Yak breeds               | Mean diff. | P-value | q or t ratio | Summary | 95% CI of diff. |
|--------------------------|------------|---------|--------------|---------|-----------------|
| Datong vs. Pollled       | -0.061     | 0.072   | 0.57         | NS      | -0.45 to 0.36   |
| Datong vs. Plateau       | -0.15      | 0.98    | 0.0128       | NS      | -0.94 to 0.63   |
| Datong vs. Gannan        | -0.35      | 0.54    | 1.7          | NS      | -1.30 to 0.45   |
| Datong vs. Tianzhu       | -0.34      | 0.42    | 1.7          | NS      | -1.1 to 0.45    |
| Polled vs. Plateau       | -0.094     | 0.057   | 0.46         | NS      | -0.89 to 0.71   |
| Polled vs. Gannan        | -0.29      | 0.39    | 1.4          | NS      | -1.1 to 0.52    |
| Polled vs. Tianzhu       | -0.28      | 0.54    | 1.3          | NS      | -1.0 to 0.53    |
| Plateau vs. Gannan       | 0.19       | 0.49    | 0.71         | NS      | -0.85 to 1.2    |
| Plateau vs. Tianzhu      | 0.18       | 0.27    | 0.68         | NS      | -0.86 to 1.2    |
| Gannan vs. Tianzhu       | -0.012     | 0.97    | 0.045        | NS      | -1.01 to 1.0    |

NS: non significant.
while Datong and polled yak populations were clustered with other groups in CNVs of GPC1. In relation to CHRM3 gene, we detected no variation. These results imply that the classifications of yak populations are mainly based on geographic distribution, ecosystem, conditions, the cultivation and differentiation of the breeds (Jincheng et al., 2006). Similarly, clustering based on CNVs of the aldehyde oxidase 3 (APOL3) and aldehyde oxidase 1 (AOX1) genes in cattle has been done in line to the evolutionary history of the breeds (Bickhart et al., 2012). Similarly, based on nuclear microsatellite markers, Zhang et al. (2008) revealed that 94.4% of the genetic variation was observed within yak breeds, while only 5.6% of the genetic variation existed between breeds. In addition, studies performed by Wang et al. (2010), support our findings, in which 93.91% total variance within populations, 5.46% among populations, and 0.64% among regions in domestic yaks were observed by use of mitochondrial DNA analysis (Wang et al., 2010). The results obtained by Wang et al. (2010), strongly support the theory that variations due to genetic diversity, genetic distance, and cluster analyses among the Tibetan yak populations are not directly correlated to geographical distribution. This supports the findings of several studies which have shown that the effect of geographical environment on genetic relationships between yak populations has gradually decreased over time (Song et al., 2015; Wang et al., 2008; Zhang et al., 2008). Thus, low sample size per breed does not follow the same pattern of demographic events in the history of yak domestication, as previously observed in cattle population (Upadhyay et al., 2017).

It’s perhaps a surprising observation that the separation and clustering of the cattle using CNVs is not superior to those based on SNPs and suggests that CNV genotyping still has room for improvement (Bickhart et al., 2016). In addition, Upadhyay et al. (2017) speculated that de-novo CNVs, CNV hot-spot regions in the genome and false CNV calls due to variation in the genotyping intensities also affect inference of population stratification. Indeed, Xu et al. (2016) suggested the CNVs have been used to separate cattle individuals into distinct groups, however, the clustering resolution within groups based on CNVs was not better than those based on SNPs. For instance, CNVs cannot distinguish the Holstein breed from the Angus breed in European taurine cattle and certain degrees of mixing within indicine individuals from CNV-based clustering results (Xu et al., 2016). Therefore, due to the accuracy of genotyping of SNPs, CNV based population stratification can be only detected among the large population groups at the continent level (Jakabsson et al., 2008; Redon et al., 2006; Upadhyay et al., 2017; Xu et al., 2016). Similarly, Liu et al. (2010) and Hou et al. (2011) described that more CNV were detected in indicine (17.63) and composite (15.75) than in taurine breeds (10.73), while within the taurine breeds, more CNVs were found in African breeds (11.60) than in European breeds (10.63), suggesting that differences CNVs among cattle breeds may be linked to the reference sample of Hereford cow of European origin, breed divergence, and history. It was intriguing to note that gene ancestries, structures, functions, and genomic distributions, gene duplication, positive selection, and conversion have shaped CNVs (Liu et al., 2010).

In addition, the within-population variability of Tianzhu, Gannan, and plateau yak breeds compared to Datong and polled yak breeds in CHKB, KLF6 and GPC1 copy number may be attributed to the high admixture pattern of their genomes (Qi, Jianlin, Wang, Rege, & Hanotte, 2009). High admixture of yak breeds in the Qinghai-Tibetan Plateau, Mongolian and Russian regions, however, a low admixture of yak breeds in the Himalayan and Pamir Plateau region have been reported (Qi et al., 2009). Such difference of copy number between different populations has also been reported in other species (Bickhart et al., 2012; Liu et al., 2010; Xu et al., 2016). The difference in CNV genes between the yak breeds may reflect gene flow, selection, and genetic drift. The frequencies of cattle introgression ranging from 1.5% Tianzhu white yak to 10.6% Gannan black yak has been shown using mt-DNA (Xuebin, Jianlin, Rege, & Hanotte, 2002). Zhang, Xu, et al. (2014) indicated that genetic variations coat colors of yak are probably due to natural (evolution and domestication) and artificial (interspecific hybridization, i.e. natural and artificial) selection. This may be the cause of some of the observed genetic variability among yak individuals. Recent studies have reported several examples of gene flow and introgression from cattle to yaks (Qi et al., 2009), the introgressed regions are enriched in genes involved in nervous system development and function, and particularly in glutamate
metabolism and neurotransmission and may contribute to the improvement of yak management and breeding (Medugorac et al., 2017). In contrast, Nguyen et al. (2005) suggested that there was no significant differentiation between the Datong and Gannan yak populations using bovine microsatellite markers. As outlined above, the CNVs of yak populations with high diversity within individuals were described previously (Qiu et al., 1986; Zhang, 1989). The northern Tibet and eastern Qinghai Tibetan Plateau areas were areas of domestication and the origin of the yak breeds (Jincheng et al., 2006). These findings are explained by Hanotte et al. (2002) where higher diversity indices result from the retention of greater numbers of alleles from the original domestication event and the least loss of genetic diversity closest to the center of origin of domestication during the migration of domestic animals. Similarly, genome CNV evolution and adaption may be a potent evolutionary force for more loss events (Zhang, Jia, et al., 2014). These results indicated that in the yak population, there are several important genes related to clear phenotypic changes and breed differences have been modified by CNVs and gene modulations. The high degree of phenotypic and genomic differences between breeds is partly due to a long domesticization history and adaptation to different local and climatic conditions throughout China (Qi et al., 2005; Zhang et al., 2008). CNVs are significant effects on genomic sequences due to changing gene structure and dosage, alternating gene regulation and exposing recessive alleles (Henrichsen, Chaignat, & Reymond, 2009; Zhang, Gu, Hurles, & Lupski, 2009). Interestingly Liu et al. (2010) reported that the difference of CNV among the cattle breeds may coincide with multiple independent domestications of cattle in the Fertile Crescent, Southwest Asia, and Africa (Caramelli, 2006; Troy et al., 2001). Wiener et al. (2003) observed that there is frequent exchange of yak breeds and migration of herders in yak habitat areas of China; hence, distribution centers of certain Chinese yak breeds have been changed corresponding to climate change over the past 50 years (Wu, 2016). Our study suggests the need for further investigation using larger yak population sizes to further validate and improve CNV genotyping approaches.

The present results show that the copy number gain events may be due to extra genetic material (artificial insemination, hybridization, and introgression) obtained during breed formation in Datong and Polled yak, since Datong yaks are produced by mating Huanhu yak cows with wild yaks (male) for meat production and Polled yak are produced from breeding Polled yak cows with horned yak bulls (Liu et al., 2013; Wiener et al., 2003); the other three yak breeds belong to the indigenous domestic yaks of China. In agreement with this results, CNV differences among cattle breeds due to altered metabolic and immune requirements, microbial fermentation in the rumen, the herd environment, and the reproductive strategy of cattle through human selection (Liu et al., 2010). In agreement with our results, Stothard et al. (2011) found that strong selection in Holstein cattle breeds in the past 50 years has led to high copy number gains and shaping CNVs. The current results suggested that a very low level of genetic diversity is present in the Chinese domestic yak breeds. Similar results had been reported in the Chinese cattle (Shi et al., 2015), for CNVs of LEPR gene. Interestingly, the CNV distribution of the KLF6 gene in the Datong yak breed, the CNV distribution of the CHKB and GPC1 genes in the polled and Datong yak breeds, and the CNV distribution of the CHRM3 gene in the Datong, polled, Tianzhu, plateau and Gannan yak breeds seems to be more centralized than that for other breeds. Beckmann, Estivill, and Antonarakis (2007) reported that continued characterization of genetic CNVs, particularly between breeds, could be an important step towards deciphering the effects of trait variations.

5. Conclusions
The development of new techniques at molecular levels have help in resolving the population genetics dynamics and more among the closely related organisms (Dufresne, Stift, Vergilino, & Mable, 2014). CNVs is a more recent technique has been successfully applied in various studies for instance, in the analysing their effect on gene expression and their association with disease susceptibility and other complex human traits (Gamazon, Nicolae, & Co, 2011). In this study, we applied the CNVs in determining...
the variation among the yaks in China. The results showed that the CNVs of the CHKB, KLF6, and GPC1 genes presented more copy number losses in Tianzhu, Gannan and Plateau yak populations while Datong and polled yak more copy number gains. However, the CHRM3 CNV was more copy number gains in five yak populations. Therefore, these results support the hypothesis that log_2 ratio is more powerful detection loss than gain copy number types. It is interesting, unlike previous studies, the breeds from the same provincial did not cluster together using CNVs of different genes, due to the different diversity of individuals and the technology used for detection. In the present results shown that the copy number gain events might be due to artificial insemination, hybridization, and introgression of breeds during breed formation. Taken together, these observations suggest that the CNVs of the CHKB, KLF6, GPC1 and CHRM3 genes could exert a significant effect on phenotypic differences in yak populations, highlighting their diversity in terms of CNVs. This study was based on real-time polymerase chain reaction (qPCR) using five different domestic yaks; therefore, it is difficult to ascertain the CNVs of genes in other studies due to the different populations, sizes, platforms, and algorithms applied for surveying.

Acknowledgements
We thank Mr. Yong Feng Zhang for guidance and technical assistance during the laboratory work and Dr. Hong Bo Wang, Ms. Li Xiaoxiao and Mr. Peng Tang for blood sample collection.

Ethics statement
The study was conducted between September 2016 and December 2017 at Key Laboratory of Yak Breeding Engineering of Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, the data obtained from yak breeding cooperatives of Gansu and Qinghai Province, China. All blood sample collection and body measurements were strict accordance of the guide for the Care and Use of Laboratory Animals, Lanzhou Institute of Husbandry Animal and Pharmaceutical Sciences, China. Also, all animals were slaughtered under anesthesia, and all necessary efforts were made to minimize risk of suffering. Thus, we agree to perform the research on the yak and the legal certificate number was SCXK (Gan) 2014-0002.

Competing interests
The authors declare no competing interests.

Author details
Habtamu Abera Goshu 1,2
E-mail: habtamu2016@yahoo.com
ORCID ID: http://orcid.org/0000-0003-2693-0540
Wu Xiao Yun 1
E-mail: wxiaoyunkevin@gmail.com
Min Chu 1
E-mail: pingyanlz@163.com
Bao Pengjia 1
E-mail: chumin@caas.cn
Ping Yan 1
E-mail: habtamu2016@yahoo.com
1 Key Laboratory of Yak Breeding Engineering, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Science, Lanzhou, P.R. China.
2 Oromia Agricultural Research Institute, Bako Agricultural Research Center, Bako, Ethiopia.

Author contributions
Habtamu Abera, data analysis, reviewed different journals and books in the line of this manuscript and Professor Ping Yan and Dr. Bao Pengjia, contributed to its supervisory and guiding. Dr. Wu Xiaoyun involved in drafting the manuscript and laboratory works. Habtamu made all writing and sketch of the manuscript.

Citation information
Cite this article as: Population genetic copy number variation of CHKB, KLF6, GPC1 and CHRM3 genes in Chinese domestic yak (Bos grunniens) breeds, Habtamu Abera Goshu, Wu Xiao Yun, Min Chu, Bao Pengjia & Ping Yan, Cogent Biology (2018), 4: 1471779.

Cover image
Source: Habtamu Abera Goshu (2017), Key Laboratory for Yak Genetics, Breeding & Reproduction Engineering of Gansu Province, Chinese Academy of Agricultural Science.

References
Almal, S. H., & Padth, H. (2012). Implications of gene copy-number variation in health and diseases. Journal of Human Genetic, 57, 6–13. doi:10.1038/jhg.2011.108
Bae, J. S., Cheong, H. S., Kim, L. H., NamGung, S., Park, T. J., Chun, J.-Y., … Shin, H. D. (2010). Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genomics, 11, 232. doi:10.1186/1471-2164-11-232
Bao, P. (2016). Integrative analysis of muscle mitochondria proteome in yak hypoxic adaptation (doctoral dissertation abstract). Chinese Academic of Agricultural Science, Lanzhou, China.
Beckmann, J. S., Estivill, X., & Antonarakis, S. E. (2007). Copy number variants and genetic traits: Closer to the resolution of phenotypic to genotypic variability. Nature Reviews Genetics, 8, 639–646.
Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., … Liu, G. E. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. Genome Research, 22(4), 778–790. doi:10.1101/gr.133967.111
Bickhart, D. M., Xu, L., Hutchison, J. L., Cole, J. B., Null, D. J., Schroeder, S. G., … Liu, G. E. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Research: an International Journal for Rapid Publication of Reports on Genes and Genomes, 23(3), 253–262. doi:10.1093/dnares/dsw013
Caramelli, D. (2006). The origins of domesticated cattle. Human Evolution, 21, 107–122.
Chen, C., Qiao, R., Wei, R., Guo, Y., Ai, H., Ma, J., … Huang, L. (2012). A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. BMC Genomics, 13, 733. doi:10.1186/1471-2164-13-733
Conrad, D. P., Pinto, D., Redon, R., Feuk, L., Gokcenm, O., Zhang, Y., … Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. Nature, 464(7289), 704–712. doi:10.1038/nature08516

Cook, E. H., Jr., & Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature, 455(7215), 919–923.

Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014). Recent progress and challenges in population genetics of polyloid organisms: An overview of current status – Of-the-art molecular and statistical tools. Molecular Ecology, 23(1), 40–69. doi:10.1111/mec.12581

Fadista, J., Thomsen, B., Holm, L.-E., & Bendixen, C. (2010). Of-the-art molecular and statistical tools.

Fernández, C. I., & Wiley, A. S. (2017). N. Linking the starch digestion hypothesis for AMY1 copy number variation in humans. American Journal of Physical Anthropology, 163(4), 645–657. doi:10.1002/ajpa.23237

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Aitkenshuler, D. M., … Lee, C. (2006). Copy number variation: New insights in genome diversity. Genome Research, 16(6), 949–961. doi:10.1101/gr.367720

Gamazon, E. R., Nicolae, D. L., & Co, N. J. (2011). A study of CNVs as traits-assocated polymorphism and as expression quantitative traits loci. Plos Genetics, 7(12), e1002192. doi:10.1371/journal.pgen.1002192

Gamazon, E. R., & Stranger, B. E. (2015). The impact of copy number variation on gene expression. Briefings in Functional Genomics, 14(5), 352–357. doi:10.1093/bfgp/elv017

Girirajan, S., Dennis, M. Y., Baker, C., Malig, M., Coe, B. P., Campbell, C. D., … Eichler, E. E. (2013). Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. American Journal of Human Genetics, 92(2), 221–237. doi:10.1016/j.ajhg.2012.12.016

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., & Rege, J. E. O. (2002). African pastoralism: Genetic imprints of origins and migrations. Science, 296, 336–339. doi:10.1126/science.1069878

Henrichsen, C. N., Chaignat, E., & Reymond, A. (2009). Copy number variants, diseases, and gene expression. Journal of Human Molecular Genetics, 18(R1), R1–R8. doi:10.1093/jhmg/dop011

Hou, Y., Liu, G. E., Bickhart, D. M., Cardone, M. F., Wang, K., Kim, E., … Van Tassell, C. P. (2011). Genomic characteristics of cattle copy number variations. BMC Genomics, 12, 127. doi:10.1186/1471-2164-12-127

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., … Lee, C. (2004). Detection of large-scale variation in the human genome. Nature Genetics, 36(9), 949–951. doi:10.1038/ng1416

Iskow, R. C., Gokcenum, O., & Lee, C. (2012). Exploring the role of copy number variants in human adaptation. Trends in Genetics: TIG, 28(6), 245–257. doi:10.1016/j.tig.2012.03.002

Jakobsen, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, J. M., Fung, H. C., & Singleton, A. B. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature, 451(7181), 998–1003. doi:10.1038/nature06742

Jincheng, Z., Zhihua, C., Sujun, Z., & Yuping, X. (2006). Classification of ecological types of the Chinese yak. Acta Ecologica Sinica, 26, 2068–2072.

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., & McGrath, A.; for the International Sheep Genomics Consortium. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS ONE, 4(3), e4668.

Ko, D. C., Gamazon, E. R., Shukla, K. P., Pfetznner, R. A., Whittington, D., Holden, T. D., … Miller, S. I. (2012). Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. Proceedings of the National Academy of Sciences of the United States of America, 109(35), E2343–E2352. doi:10.1073/pnas.120670109

Lee, K., Nguyen, D. T., Choi, M., Cha, S.-Y., Kim, J.-H., Dadi, H., … Park, C. (2013). Analysis of cattle olfactory subgenome: The first detailed study on the characteristics of the complete olfactory receptor repertoire of a ruminant. BMC Genomics, 14, 596. doi:10.1186/1471-2164-14-596

Lehnert, S. A., Reverter, A., Byrne, K. A., Wang, Y., Nattorss, G. S., Hudson, N. J., & Greenwood, P. L. (2007). Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. BMC Developmental Biology, 7, 95. doi:10.1186/1471-213X-7-95

Li, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellomore, A., … Keele, J. W. (2010). Analysis of copy number variations among diverse cattle breeds. Genome Research, 20(5), 693–703. doi:10.1101/gr.105403.110

Li, M., Liu, B., Huang, Y., Yanga, M., Lana, X., Lei, C., & Chena, H. (2016). Copy number variation of bovine MAPK10 modulates the transcriptional activity and affects growth traits. Livestock Science, 194, 44–50. doi:10.1016/j.livsci.2016.09.014

Liu, W. B., Liu, J., Liang, C. N., Guo, X., Bao, P. J., Chu, M., & Yan, P. (2013). Associations of single nucleotide polymorphisms in candidate genes with the polled trait in Datong domestic yaks. Animal Genetics, 45(1), 138–141. doi:10.1111/age.12081

Medugorac, I., Graf, A., Grohs, C., Rothammer, S., Zagsduren, Y., Glaydy, E., … Capillon, A. (2017). Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nature Genetics, 49(3), 470–475. doi:10.1038/ng.3775

Nguyen, T. T., Genini, S., Menentrey, F., Malek, M., Vogeli, P., Goe, M. R., & Stranzinger, G. (2005). Application of bovine microsatellite markers for genetic diversity analysis of Swiss yak (Poephagus grunniens). Animal Genetics, 36, 484–489.

Novembre, J., & Ramachandran, S. (2011). Perspectives on the human population structure at the cusp of the sequencing era. Annual Review of Genomics and Human Genetics, 12, 245–274. doi:10.1146/annurev-genome-090810-183123

Olshen, A. B., Venkatraman, E., Lucto, R., & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics, 5, 557–572.

Olson, M. V. (1999). When less is more: Gene loss as an engine of evolutionary change. American Journal of Human Genetics, 64(1): 18–23.

Parsch, J., & Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. Nature Reviews Genetics, 14(2), 83–87. doi:10.1038/nrg3176

Perry, G. H, Dominy, N. J, Clow, K. G, Lee, A. S, Fieglher, L, Redon, R, … Stone, A. C. (2007). Diet the evolution of
Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Natural Review Genetics*, 16(3), 172–183. doi:10.1038/nrg3871

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review Genomics of Human*, 10, 451–481. doi:10.1146/annurev. genom.9.081307.164217

Zhang, G. X., Chen, W. S., Xue, M., Wang, Z. G., Chang, H., Han, X., & Wang, D. L. (2008). Analysis of genetic diversity and population structure of Chinese yak breeds (Bos grunniens) using microsatellite markers. *Journal of Genetics and Genomics*, 35, 233–238.

Zhang, L., Jia, S., Yang, M., Xu, Y., Li, C., Sun, J., … Chen, H. (2014). Detection of copy number variations and their effects in Chinese bulls. *BMC Genomics*, 15(1), 480. doi:10.1186/1471-2164-15-480

Zhang, M. Q., Xu, X., & Luo, S. J. (2014). The genetics of brown coat color and white spotting in domestic yaks (Bos grunniens). *Animal Genetics*, 45, 652–659. doi:10.1111/ age.12191

Zhang, Q., Youji, M., Yang, Z., & Xingxu, Z. (2015). Identification of copy number variations in qinchuan cattle using BovineHD genotyping beadchip array. *Molecular Genetics and Genomics: MGG*, 290(1), 319–327. doi:10.1007/s00438-014-0923-4

Zhang, R. (1989). *China Yak*. Lanzhou: Gansu Scientific, and Technology.

Zhang, X., Wang, K., Wang, L., Yang, Y., Ni, Z., Xie, X., … Qiu, Q. (2016). Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics*, 17, 379. doi:10.1186/s12864-016-2702-6