

2021-06-17

Genome Scan for Variable Genes Involved in Environmental Adaptations of Nubian Ibex

Chebii, Vivien

Springer Nature Switzerland AG.

<https://doi.org/10.1007/s00239-021-10015-3>

Provided with love from The Nelson Mandela African Institution of Science and Technology



Genome Scan for Variable Genes Involved in Environmental Adaptations of Nubian Ibex

Vivien J. Chebii^{1,4} · Emmanuel A. Mpolya¹ · Samuel O. Oyola² · Antoinette Kotze^{3,6} · Jean-Baka Domelevo Entfellner⁴ · J. Musembi Mutuku^{4,5}

Received: 2 August 2020 / Accepted: 5 June 2021 / Published online: 17 June 2021
© The Author(s) 2021

Abstract

The Nubian ibex (*Capra nubiana*) is a wild goat species that inhabits the Sahara and Arabian deserts and is adapted to extreme ambient temperatures, intense solar radiation, and scarcity of food and water resources. To investigate desert adaptation, we explored the possible role of copy number variations (CNVs) in the evolution of *Capra* species with a specific focus on the environment of *Capra nubiana*. CNVs are structural genomic variations that have been implicated in phenotypic differences between species and could play a role in species adaptation. CNVs were inferred from *Capra nubiana* sequence data relative to the domestic goat reference genome using read-depth approach. We identified 191 CNVs overlapping with protein-coding genes mainly involved in biological processes such as innate immune response, xenobiotic metabolisms, and energy metabolisms. We found copy number variable genes involved in defense response to viral infections (Cluster of Differentiation 48, UL16 binding protein 3, Natural Killer Group 2D ligand 1-like, and Interferon-induced transmembrane protein 3), possibly suggesting their roles in Nubian ibex adaptations to viral infections. Additionally, we found copy number variable xenobiotic metabolism genes (carboxylesterase 1, Cytochrome P450 2D6, Glutathione S-transferase Mu 4, and UDP Glucuronosyltransferase-2B7), which are probably an adaptation of Nubian ibex to desert diets that are rich in plant secondary metabolites. Collectively, this study's results advance our understanding of CNVs and their possible roles in the adaptation of Nubian ibex to its environment. The copy number variable genes identified in Nubian ibex could be considered as subjects for further functional characterizations.

Keywords Nubian ibex · Copy number variation · Genome adaptations · Desert adaptation

Handling editor: Rosa Fregel.

✉ Vivien J. Chebii
chebiiv@nm-aist.ac.tz; chevivien@gmail.com

¹ School of Life Science and Bioengineering, Nelson Mandela Africa Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania

² International Livestock Research Institute (ILRI), Nairobi, Kenya

³ South African National Biodiversity Institute, Pretoria, South Africa

⁴ Biosciences Eastern and Central Africa - International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

⁵ Central and West African Virus Epidemiology (WAVE), Pôle Scientifique et d'Innovation de Bingerville, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

⁶ Department of Genetics, University of the Free State, Bloemfontein, South Africa

Introduction

The Nubian ibex (*Capra nubiana*) is one of the ten species belonging to the genus *Capra*. The other species are *C. caucasica*, *C. cylindricornis*, *C. falconeri*, *C. pyrenaica*, *C. ibex*, *C. sibirica*, *C. walie*, *C. aegagrus*, and the domestic goat species (*C. hircus*). Nubian ibex is found as small fragmented populations in Egypt, Sudan, Eritrea, Jordan, Oman, Yemen, and Saudi Arabia (Ross et al. 2020). The International Union for Conservation of Nature (IUCN) Red List classifies the Nubian ibex as Vulnerable (Ross et al. 2020). The Nubian ibex is well adapted to hot desert environments characterized by high diurnal temperatures and scarcity of feed and water (Habibi 1994). The Nubian ibex has evolved mechanisms to survive in harsh environments; for instance, it feeds on desert plants rich in secondary metabolites such as alkaloids, suggesting that they have an excellent detoxification system (Hakham and Ritte 1993). Therefore, Nubian

ibex's ability to survive in harsh environments makes it a potential model species for studying environmentally resilient genetic adaptations applicable in other livestock species of economic value.

Copy number variations (CNVs) are genomic structural variants involving duplications or deletions of segments greater than 1000 bp leading to copy number differences among individuals within or between species (Redon et al. 2006). CNVs confer phenotypic effects by changing gene dosage, transcript structure, or regulating genes' expression and functions (Bickhart et al. 2012). Whole gene CNVs affect gene expression such that the higher the number of copies of a gene, the higher its expression levels and vice versa (Butler et al. 2011; Cardoso-Moreira et al. 2016). Exonic CNVs, on the other hand, provide an alternative way of meeting functional demands of a cell through mutually exclusive splicing of the resulting exon duplicates; this allows a single gene to encode functionally diverse proteins (Jones et al. 2012; Kondrashov and Koonin 2001; Letunic et al. 2002). CNVs overlapping with intronic regions of protein-coding genes are likely to affect gene expression since they host regulatory elements such as enhancers, silencers, and non-coding RNAs that modulate transcription (Chorev and Carmel 2012; Rigau et al. 2019).

There is growing evidence of the association between CNVs and a diverse range of phenotypic deviations among species (Bhanuprakash et al. 2018; Iskow et al. 2012); for example, the human genome contains an average of 272 copies of the DUF1220 domain, which is suggested to contribute to the differences in the brain size and cognitive ability between humans and other primates with fewer copies (35 in macaques, 99 in gorillas, and 125 in chimpanzees) (O'Bleness, 2012). Similarly, CNVs of neural development genes (CAST family member 2, Gamma-aminobutyric acid receptor subunit beta-2 and Neuronal PAS domain protein 3) in the domestic goat (Dong et al. 2015), and the behavioral genes (Glutamate Ionotropic Receptor NMDA Type Subunit 2D, Netrin 5 and Neurotrophin-4) in domestic yak (Zhang et al. 2016) may have played a role in the domestication of these species from their respective wild ancestors. A desert-adapted Bactrian camel has two copies each of *NR3C2* and *IRS1* genes known to be involved in water and salt homeostasis, suggesting their role in the desert adaptation of camels compared to their closely related temperate-adapted alpaca (Wu et al. 2014).

Advances in genome sequencing technologies, particularly the whole genome shotgun sequencing platforms such as the Illumina HiSeq 2500, provide approaches such as read-depth, split-read, and paired-end mapping (Alkan et al. 2011) for detecting CNVs from sequence data de novo. Among the sequence-based techniques for calling CNVs, read-depth algorithms can detect exact copy numbers and CNVs in complex genomic regions; hence it is most

commonly used for CNV identifications (Alkan et al. 2011). The read-depth approach assumes that the number of reads from a randomly generated shotgun sequence data set that transverse any genomic loci in a reference genome sequence is proportional to the number of copies of the orthologous loci in the test genome. Therefore, gain of copy number events at a given genomic interval in the test genome is seen as higher than average stacking of sequence reads at the orthologous interval in the reference genome, and conversely, a loss of copy number events manifests as a paucity of reads spanning the interval (Abyzov et al. 2011). The recent completion of the domestic goat reference genome (Bickhart et al. 2017) allows in-depth study of adaptive evolution in *Capra* species through CNV screening. This study's objective was to determine CNVs in the Nubian ibex genome compared with the domesticated goat and their potential role in the adaptations to desert habitats.

Materials and Methods

Whole Genome Sequence Data

Sequence data for three Nubian ibex individuals were downloaded from the National Center for Biotechnology Information database (<https://www.ncbi.nlm.nih.gov/>). The data are comprised of sequence data of Nubian ibex individuals obtained from South Africa (Accession number: SRR12990712), Egypt (Accession number: SRR8437789), and Saudi Arabia (Accession number: SRR8437792). The latest reference genome sequence (ARS1 assembly; GCA_001704415.1) (Bickhart et al. 2017) of *Capra hircus* (domestic goat) was downloaded from Ensembl (ftp://ftp.ensembl.org/pub/release-102/fasta/capra_hircus/dna/).

The sequence data downloaded from the public database were assessed for quality using fastQC (Andrews 2010). Adapter sequences in Saudi Arabia data were trimmed using the ILLUMINACLIP function in Trimmomatic version 0.38 (Bolger et al. 2014).

Mapping and Copy Number Variants (CNV) Calling

Sequence reads for each of the Nubian ibex individuals were aligned to the domestic goat reference genome using Burrow's Wheeler Alignment Maximal Exact Match algorithm (BWA-MEM) version 0.7.15a with parameters set to default (Li and Durbin 2010). Polymerase Chain Reaction (PCR) duplicates were removed using *rmDup* command line of SAMtools (v0.1.8) (Li et al. 2009). The Sequence Alignment Map (SAM) files were converted into Binary Alignment Map (BAM) files using SAMtools 1.4.1 (Li et al. 2009).

Copy number variants in each of the Nubian ibex individual genomes were inferred using a read-depth approach

implemented in CNVnator (version 0.3.3) (Abyzov et al. 2011), relative to the domestic goat reference genome. CNVnator was set to call CNVs using a genomic window size of 100 base pairs from sequence data for Nubian ibexes sampled from Egypt and South Africa. A genomic window size of 200 base pairs was used to infer CNVs from the Nubian ibex sequence data sampled from Saudi Arabia. The genomic window size for each sequence data was selected based on the authors' recommendations (the ratio of average read-depth signal to its standard deviation should be between 4 and 5) (Abyzov et al. 2011). Other parameters were set to default. The *-unique* flag was used to obtain the zero-mapping quality (q0) score of the calls as per the author's recommendation. The zero-mapping quality (q0) score refers to the fraction of reads in an interval that aligns to more than two locations in the genome.

The output from CNVnator was filtered to retain high quality CNV calls using the following criteria: a call with normalized read depth < 0.7 , p value < 0.05 , $q0 > 0.7$ was considered a gain of copy number event in the domestic goat reference genome. Similarly, CNV with normalized read depth > 1.20 , p value < 0.05 , $q0 > 0.7$ was considered a gain of copy number event in the reference genome; with Nubian ibex having more than one copy but fewer than those in the reference. CNV region with normalized read depth > 1.5 , p value < 0.05 , $q0 < 0.2$ was considered as gain of copy number event in Nubian ibex, while CNV with normalized read depth < 0.7 , p value < 0.05 , and $q0 < 0.2$ was considered as loss of copy number event in Nubian ibex. CNV calls shared across the Nubian ibex individuals with $> 50\%$ overlap were retained. Finally, the DNA sequences spanning the candidate reference gain of copy number events were extracted from the domestic goat reference genome and subjected to dot plot analysis (using online dot plotter at NCBI; <https://www.ncbi.nlm.nih.gov/>) to test for tandem repeats or otherwise aligned to the entire genome of the domestic goat to detect segmental duplication using blast (parameters set to e value; $1e-10$, percentage identity $> 80\%$) (Altschul et al. 1990).

To account for variable regions that are not exclusive to Nubian ibex, we downloaded the genome coordinates of CNVs that have been identified in goat populations (Di Gerlando et al. 2020; Guan et al. 2020) and extracted overlapping sites using bedtools intersect (Quinlan and Hall 2010). Copy number variable regions overlapping ($> 10\%$) with domestic goat CNV sites were excluded from further analysis.

Assessment of Read Depth Using Simulated CNV Events

CNVnator has been mostly applied to detect CNVs within species, and the interpretation framework is not clear when applied to CNV detection between species, albeit closely

related ones. For example, in the default framework, a read depth of 0.5 or less is taken to indicate a loss of copy in the test genome; however, a read depth of less than 0.5 can also be due to other genotypes such as a duplication in the reference. We carried out a simulation experiment to determine suitable cut-offs for CNV events and test the sensitivity of the CNVnator in detecting gain of copy events in the reference genome. Briefly, we simulated two duplications sites per chromosome in the domestic goat reference genome located at coordinates where the genome coverage was equal to the average read depth (copy number neutral). One of the sites was simulated to reflect two tandem copies and the other site four copies of fragments ranging in size from 3000 to 176,000 bp. CNVnator was then run as described previously using the simulated domestic goat genome as the reference.

Gene Content and Functional Annotation

Genomic positions of the breakpoints (start and end positions of a CNV) shared across three Nubian ibex individuals as detected by CNVnator were annotated using Variant Effect Predictor (VEP v.95) (McLaren et al. 2016) relative to the domestic goat reference genome. Each position was assigned Sequence Ontology terms, defining different genomic regions: coding, upstream, downstream, non-coding, intergenic, intronic, and untranslated (UTR) sequences regions. CNVs that overlapped with only the introns of a gene (does not overlap with any exon) were classified as intronic CNVs, while those overlapping with some exons were considered exonic CNVs. CNVs that covered entire genes were classified as whole gene CNVs.

Gene ontology (GO) terms assignments for the CNV genes were found by searching the genes in Ensembl Goat Genes v.97 using Biomart (Kinsella et al. 2011), and additional gene functions were sourced from the literature. Gene enrichment analysis was carried out using The Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 (Dennis et al. 2003). Since very few genes in the domestic goat reference genome have been assigned gene ontology terms, goat Ensembl gene IDs were converted to the orthologous human Ensembl gene ID using Biomart (<https://www.ensembl.org/biomart/>). The corresponding human Ensembl gene IDs were used for gene enrichment analysis as described above.

Results

Copy Number Variants Identification

A total of 446 million, 549 million, and 781 million paired-end sequence reads obtained from Nubian ibex individuals

sampled from Egypt, Saudi Arabia, and South Africa, respectively, were downloaded from the public database. Approximately 88% and 97% of the sequence reads for the Saudi Arabia and Egyptian Nubian ibex individuals were mapped to the domestic goat reference genome. Successfully mapped sequence reads yielded 27-fold and 26-fold coverage for Saudi Arabia and Egyptian Nubian ibex. More than 98% of the sequence reads of Nubian ibex obtained from South Africa mapped successfully onto specific sites of the reference genome, ARS1, of the domesticated goat, *C. hircus*, with coverage of 36x. A summary of Nubian ibex sequence data is provided in Online Resource 1.

Simulated CNV Sites Detection

We simulated tandem repeat CNV at 58 sites, two each per autosome, 38–43 were identified by CNVnator across the three Nubian ibex genomes, implying a sensitivity of just over 66–75%. The read depths for the simulated CNVs sites were 0.5 or less, as expected. Crucially, the read depth for the simulated CNV with two tandem copies was approximately 0.5, while for those with four tandem copies it was 0.25. We observed that all simulated CNVs detected had a mean q_0 value ranging from 0.7 to 0.99, indicating that any novel CNV site with a q_0 score greater than 0.7 is a valid CNV in which the number of copies in the reference is more than one. The simulation showed that CNVnator has a poor accuracy of detecting CNV sites' boundaries, with calls being on average 1654 bp away from the actual positions.

CNVs in Nubian ibex and Domestic Goat Genomes

CNVnator detected 13,472, 7724, and 9064 raw CNV loci in the Nubian ibex individuals obtained from South Africa, Egypt, and Saudi Arabia (see Online Resource 2). A total

of 1726, 1544, and 1234 CNV loci detected in Nubian ibex individuals from South Africa, Egypt, and Saudi Arabia were retained after stringent filtering (see Online Resource 2). Twenty-seven CNVs detected previously in domestic goat genomes were further filtered out; refer to the Table in the Online Resources 3. Altogether 367 putative CNV loci were shared across the three individuals; 271 were gain of copy events in Nubian ibex and the domestic goat, while 96 were loss of copy events in Nubian ibex (see Table 1). The final CNV set of 367 loci covered less than 1% (5.6Mbp) of the Nubian ibex genome. The lengths of the CNVs ranged from 1.1 kbp to 214 kbp, with a median of 9.1 kbp (see Fig. 1). The estimated number of copies ranged between 0 and 463 copies.

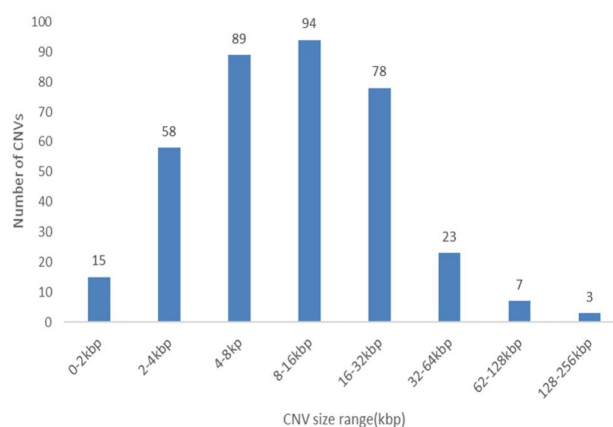


Fig. 1 CNV size distribution across the Nubian ibex genome

Table 1 CNV loci detected in Nubian ibex individuals

CNV type	Nubian ibex sample source			CNV shared across Nubian ibexes SA_E_A (intersect)
	South Africa (SA)	Egypt (E)	Saudi Arabia (A)	
Deletions in Nubian ibex	971	750	682	97
Duplications in Nubian ibex	483	543	424	206
Duplication in reference	251	234	125	62
Duplications in reference genome with more copies in Nubian ibex genome	21	17	3	2
Total CNVs	1726	1544	1234	367

Deletions in Nubian ibex (Read depth < 0.7, p value < 0.05, q_0 < 0.2); test genomes have few copy numbers relative to the reference genome. Duplication in Nubian ibex (Read depth > 1.5, p value < 0.05, q_0 < 0.2); more copies in test genomes relative to the reference. Duplication in reference (Read depth < 0.7, p value < 0.05, q_0 > 0.7); more copies in reference genome relative to the test genomes. Duplication in reference with more copies in Nubian ibex (RD > 1.20, p value < 0.05, q_0 > 0.7); all genomes have many copies; however, the reference has more copy numbers than the test genomes

CNVs Sequence Annotation

Sequence annotation showed that 161 (36%) of the CNVs were in intergenic regions. Other CNVs were in the genic regions, with 62 (14%) being exonic and 60 (13%) were intronic CNVs (see Fig. 2). CNV sequence annotation data is provided in Online Resource 4.

Gene Content of the CNV Loci

The CNV events overlapped with 191 protein-coding genes, 9 lincRNAs, 7 snRNAs, 3 rRNAs, 4 pseudogenes, and 1 snoRNA. The CNVs spanning genes were outside the coding sequence regions (introns, upstream, downstream and untranslated gene regions) or within the coding sequence region (exons or entire genes). A total of 63 and 33 protein-coding genes overlapped with exonic and upstream region CNVs. Five whole genes overlapped with gain of copy number variable regions in Nubian ibex genome. Additionally, 60 intronic CNVs were reported, and 23 downstream CNVs (refer to Online resource 5). A summary of the number of protein-coding CNVs is provided in Table 2.

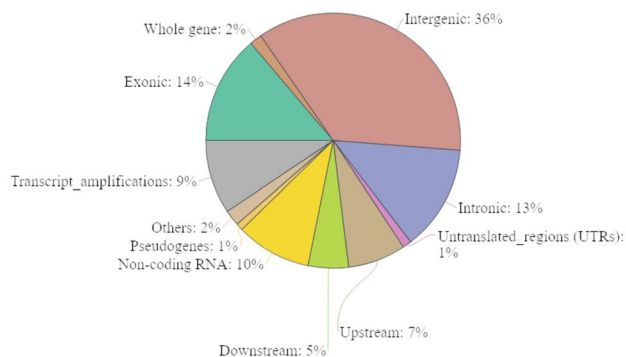


Fig. 2 The distribution of 367 CNVs across the Nubian ibex genome (Intergenic, introns, exons, upstream/downstream, and untranslated regions). The majority of the CNVs were found in the intergenic regions of the genome

Of the protein-coding genes, 126 were in the gain of copy number variable regions, while 47 were in loss of copy number regions in Nubian ibex. Illustrations of gain and loss of copy events spanning protein-coding genes are provided in Online Resource 6. Eighteen protein-coding genes detected were in the gain of copy number variable regions in the domestic goat reference genome. Exonic gain of copy events in the domestic goat reference genome overlapped with six protein-coding genes, while intronic CNV overlapped with eight genes. Additionally, three protein-coding genes overlapped with upstream CNVs in the domestic goat reference genome. A dot plot of CNV event in the domestic goat reference genome with read depth < 0.7 and $(q0 > 0.7)$ depicting tandem duplication is illustrated in Fig. 3.

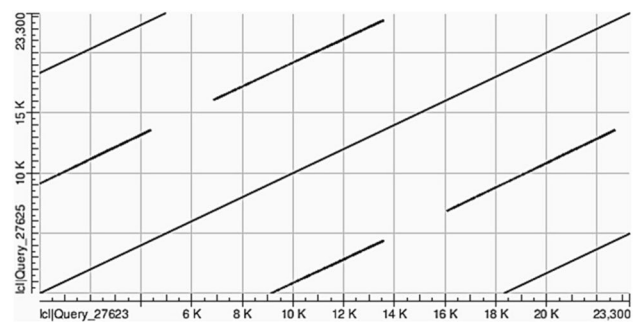


Fig. 3 A dot plot of tandem duplication at chr1:67,890,500–67,913,800 in the domestic goat reference genome detected by CNVnator. The DNA sequence corresponding to the copy number variable region was extracted from the domestic goat reference genome using bedtools *-getfasta* tools (Quinlan and Hall 2010); the dot plot was generated from the sequence from alignment against itself in NCBI BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The CNV region overlaps with the first and the second exons of the Myosin Light Chain Kinase (*MYLK*) gene. The read depth of the CNV loci is 0.21, indicating that the domestic goat has four copies of the first two exons of *MYLK* gene, while the Nubian ibex has one copy

Table 2 CNV loci overlapping with protein-coding genes detected in Nubian ibex and domestic goat genomes

CNV type	CNV genes in coding sequence regions		CNV genes outside coding sequence region			
	Exonic	Whole gene	Introns	Upstream	Downstream	UTRs
Deletions in Nubian ibex	3	2	30	8	4	0
Duplications in Nubian ibex	54	5	22	22	19	4
Duplication in reference	6	0	8	3	0	1
Total CNV genes	63	7	60	33	23	5

Deletions in Nubian ibex; test genomes have few copy numbers relative to the reference genome. Duplication in Nubian ibex; more copies in test genomes relative to the reference. Duplication in reference; more copies in reference genome relative to the test genomes

Functions of the CNV-Associated Protein-Coding Genes

Gene ontology assignments showed that the CNV-associated protein-coding genes are involved in diverse biological processes such as complement activation, inflammatory response, proteolysis, negative regulation of endopeptidase activity, drug metabolism, and positive regulation apoptotic cell clearance, among several other roles. No gene ontology (GO) terms were significantly enriched (See Online Resource 5).

Gene functional analysis showed that the CNV-associated protein-coding genes are involved in diverse biological processes including innate immune response (Cluster of Differentiation 54 (CD54)), Interferon Beta 1 (IFNB1)), Intercellular adhesion molecule-1 (ICAM-1), bovine major histocompatibility complex (BoLA), Interferon-induced transmembrane protein 3 (IFITM3) and Complement Factor H Related 4 (CFHR4)). Other copy number variable immune genes reported include: UL16 binding protein 3 (ULBP3), Natural Killer Group 2D ligand 4-like (NKG2D ligand 4-like), NKG2D ligand 1-like (NKG2D ligand 1-like), Bactericidal/permeability-increasing fold containing family A, member 1 (BPIFA1), Cluster of Differentiation 48 (CD48), complement C3 (C3), complement C4B (Chido blood group) (C4B), Cluster of Differentiation 54 (CD54), adhesion G protein-coupled receptor E3 (ADGRE3) and serpin family A member 3 (SERPINA3) (See Table 3).

Similarly, xenobiotic and drug metabolic process genes (carboxylesterase 1 (CES1), cytochrome P450 family 2 subfamily B member 6 (CYP2B6), Cytochrome P450 2D6 (CYP2D6), cytochrome P450 family 2 subfamily D

member 14 (CYP2D14), Glutathione S-transferase Mu 4 (GSTM4), UDP-glucuronosyltransferase 2B31 (UGT2B31) UDP-Glucuronosyltransferase-2B7 (UGT2B7), phytanoyl-CoA hydroxylase (PHYH), and Multidrug resistance protein 4 (MRP4) were reported to be copy number variable in Nubian ibex. Furthermore, lipid and energy metabolism genes such as low-density lipoprotein receptor-related protein 11 (LRP11-like), Acyl-CoA thioesterase 13 (ACOT13), Mitochondrial Encoded ATP Synthase Membrane Subunit 6 (MT-ATP6), and Fatty acid desaturase 2 (FADS2) were shown to be copy number variable in Nubian ibex genome. Other CNV-associated genes reported are involved in reproductive functions such as MAN2B2, MAN2B2-Like, CSH2, MYADM, ADAM18 in Nubian ibex and DDX25 in domestic goat. A summary of genes overlapping with CNVs in the exons and upstream gene regions is provided in Table 3 and Online Resource 7.

Discussion

Copy Number Variants Calling

CNVs were inferred from sequence data of three Nubian ibexes relative to the domestic goat reference genome using read-depth approach (Abyzov et al. 2011). Read depth is a well-established approach whose sensitivity and power to detect CNVs have been confirmed experimentally (Paudel et al. 2015; Pezer et al. 2015; Wang et al. 2019). For example, assessment of read-depth using droplet digital PCR (ddPCR) showed that the CNVs detected using ddPCR strongly correlated with those identified using a read-depth

Table 3 Protein-coding genes overlapping with CNVs within the exons and upstream gene regions

Biological process	Gain of copy number genes in Nubian ibex (RD > 1.5, $p < 0.05$, $q_0 < 0.2$)	Gain of copy number genes in the domestic goat (RD < 0.7, $p < 0.05$ > 0.7)
Detoxification/xenobiotic metabolism	GSTM4, CYP2D14, CYP2B6, CYP2D6, CYP2C31, UGT2B31, MRP4Like, MRP4, PHYH, CES1, UGT2B7	–
Innate immunity	GSTM4, GIMAP7, CD48, C3, NKG2D ligand 4-like, NKG2D ligand 1-like, ULBP3, BPIFA1, MYADM, SERPIN A3-6, SERPIN B6, C4A, C4B, IFNB1, CD54, CFHR4, UBD, SIGLEC5, KIR2DL1, TNFRSF10B-like, CD163L1, ADGRE3, TRAV22, IGLV5-45, IGLV1-40, IFITM3, BoLA, Antigen WC1.1-like, IFI27L2	ADGRE2
Cell growth, recycling and metabolism	GIMAP7, LRP11L, PHYH, ZYMND8, MYADM, NPC1, RMC1, POLB, PIK3R4, C19orf12, UBD	TBC1D22B
Lipid and energy metabolism	LRP11L, MRP4, MRP4L, PCSK2, FADS2L, ACOT13, CFHR4, C19orf12, NPC1, LRP11, MT-ATP6	–
Fertility/reproduction	MAN2B2, MAN2B2-Like, CSH2, MYADM, ADAM18	DDX25, CELSR1
Other	CYP21A2, KRTAP1-1, GPR157, STK19, PINLYP, C19orf12, OR2H1D, C6orf62, PRDM9, SLC2A5	KRTAP3-1, MYLK, KRTAP1-1, PRKCA, CRYM-AS1, TBC1D22B

approach with a low false discovery rate of 8.7% (Pezer et al. 2015). Based on previous experimental validation of the read-depth method, we did not confirm the CNVs identified in this study; hence, the results should be considered in light of this limitation. However, our simulation experiment carried out by introducing ‘artificial CNVs’ to known regions in the reference genome showed that CNVnator could detect approximately 71% of the simulated CNVs, thus supporting that most of the CNVs detected in our study are potential true positives. Our simulation experiment showed that CNVnator has poor accuracy in detecting CNV site boundaries; this limitation could lead to overestimation or underestimation of CNVs. In total, 1234, 1544, and 1726 CNV loci were detected in the three analyzed Nubian ibexes, a number which is comparable to CNVs discovered in other livestock species such as yak (Zhang et al. 2016). Altogether, 367 CNVs were detected across the Nubian ibex genomes. Twenty-seven (27) CNVs identified across the Nubian ibexes reported previously in domestic goat genomes (Di Gerlando et al. 2020; Guan et al. 2020) were discarded since they reflect polymorphisms in/across goat species. Although only three Nubian ibexes were investigated, the samples were derived from unrelated individuals representing different geographical populations (South Africa, Egypt and Saudi Arabia). Therefore, the identified overlapping CNVs potentially represent Nubian ibex specific CNVs.

Copy Number Variable Protein-Coding Genes

Several protein-coding genes overlapping with copy number variable regions were reported in the three Nubian ibexes and domestic goat genome, representing a valuable resource for future studies on the relation between CNV genes and phenotypic variations. Twenty GO terms were associated with the protein-coding CNV genes; however, none were significant in terms of enrichment. Nevertheless, the gene ontology assignment indicated that the CNV-associated protein-coding genes are involved in diverse biological processes such as cell growth, recycling and metabolism, and energy metabolism. Consistent with other CNV studies in livestock species (Bickhart et al. 2012; Zhang et al. 2016), we found clusters of drug metabolism and immune-related genes in copy number variable regions in the Nubian ibex genome.

The Possible Roles of CNV Genes in Nubian ibex Adaptations

We found clusters of CNV-associated protein-coding genes involved in immune response and drug metabolisms that deserved more attention because of their known functions. Genes involved in defense response to bacterial or viral infections such as BPIFA2, CD48, ULBP3, NKG2D ligand

1-like, and NKG2D ligand 4-like were reported to have more copies in the genomes of Nubian ibex relative to the domestic goat reference. BPIFA2 is a glycoprotein expressed in the airway epithelium and submucosal glands of the upper airways, where it offers protection against bacterial and viral infections (Akram et al. 2018; Liu et al. 2013). Previous studies have shown that BPIFA2 has a defensive role in *K. pneumoniae* infection virus (Zhou et al. 2008), suggesting that BPIFA2 might have defensive roles against other viruses. Gain of copy number variations of CD48, a member of the signaling lymphocyte activation molecular family, was observed in the investigated Nubian ibexes. CD48 is involved in diverse immune responses ranging from T cell activation, granulocyte activity, allergic inflammation to natural killer function and antimicrobial immunity (McArdel et al. 2016). Although CD48 plays diverse immune response roles, studies have shown that it is a target of immune evasion by viruses (McArdel et al. 2016). The mucin-like protein m153 in murine cytomegalovirus (CMV) has been shown to reduce expression of CD48 on macrophages, limiting NK-cell-mediated control of viral infection (Zarama et al. 2014). We hypothesize that the gain of copy number variation in CD48 might be an evolutionary adaptation that enables it to produce diverse functional proteins to confer the needed immunosurveillance. Our results showed that NKG2D ligands system genes (ULBP3, NKG2D ligand 1-like, and NKG2D ligand 4-like) were under gain of copy events in Nubian ibex. ULBP3 and NKG2D ligand 4-like expressions are induced by stressors, such as viral infections, heat shock, tissue damage, tumorigenesis, and DNA damage (Lanier 2015). Once expressed, NKG2D ligands bind to NKG2D receptors, which mounts cell-mediated cytotoxicity and cytokine production, thus eliminating stressed cells (Zingoni et al. 2018). Nubian ibex is exposed to viral diseases such as the Peste des petits ruminants virus (Clarke et al. 2017; Wensman et al. 2018) and Malignant catarrhal fever virus (Okeson et al. 2007) in its environment. The exonic gain of copy numbers immune genes (BPIFA1, CD48, ULBP3, NKG2D ligand 1-like, and NKG2D ligand 4-like) could be an evolutionary mechanism that enables Nubian ibex to encode functionally diverse proteins through alternate splicing in response to viral stressors.

Other immune response genes reported include C3, C4A, and C4B that play essential roles in the activation of the classical and lectin pathways of the complement system that lead to neutralization of invading microbes and clearance of immune complexes (Miyagawa et al. 2008; Yang et al. 2007). Deficiency of the C3 gene is associated with susceptibility to systemic lupus erythematosus (SLE) (Miyagawa et al. 2008). Similarly, many copy numbers of C4 have been shown to alleviate susceptibility to systemic lupus erythematosus (SLE) in humans (Jüptner et al. 2018). Other recent studies have shown that increased copy numbers of C4A

offer protection against age-related macular degeneration (AMD) (Grassmann et al. 2016). Further studies are needed to ascertain the possible role of complement component genes (C3, C4A, and C4B) in Nubian ibex adaptations; however, it is known to play critical roles in immunosurveillance. Clusters of immune response genes may reflect the different immune response strategies between the Nubian ibex and the domestic goat in response to pathogens in their environments.

Xenobiotic metabolism genes including CYP2D14, CYP2B6, CYP2D6, CYP2C31, UGT2B31, UGT2B7, GSTM4, CES1, MPRP4, and MRP4Like were shown to be under the gain of copy number variations in Nubian ibex. Drug-metabolizing genes play crucial roles in eliminating plant secondary metabolites that could otherwise be toxic to livestock species (Maréchal et al. 2008). The cytochromes P450 genes (CYP2D14, CYP2B6, CYP2D6, and CYP2C31) and Carboxylesterase (CES1) are oxidizing enzymes that mediate biotransformation of xenobiotic compounds (phase I reactions) (Maréchal et al. 2008; Wang et al. 2018). The UDP-glucuronosyltransferases (UGT2B7 and UGT2B1) and glutathione S-transferases (GSTM4), on the other hand, are conjugative enzymes that play a vital role in glucuronidation (phase II) (Iyanagi 2007). The conjugated compounds with glucuronate or glutathione are transported out of the cell by the ATP-binding cassette transporters (ABC transporters) such as MRP4Like and MRP4 (Russel et al. 2008). Desert plants are rich in secondary metabolites such as alkaloids, flavonoids, oxalates, and tannins (Jacobson et al. 2009; Robertson et al. 2018; Zahedi et al. 2019). Nubian ibex have evolved in deserts, and it has been observed that they consume alkaloid-producing plants when food is scarce (Habibi 1994; Hakham and Ritte 1993). Gain of copy numbers of xenobiotic metabolisms genes could be one of the evolutionary mechanisms which enable the Nubian ibex to cope with the toxic secondary metabolites in their diet.

Conclusion

In this study, we compared the genome of the endangered desert-adapted Caprine species (Nubian ibex) with that of the domesticated goat (*C. hircus*), which is of substantial economic importance globally. We sought to gain insights from the copy number variants at the genome-wide level and their possible role in adaptations of the Nubian ibex to the desert habitat. A total of 367 CNV loci shared across three analyzed Nubian ibex individuals were identified as potential CNV sites, many of which overlapped known protein-coding genes. From the analysis of the involved genes, we conclude that potentially a significant driver for the difference between domestic goat and the Nubian ibex is the response to viral disease burdens and plant secondary metabolites in their

diet as indicated by a preponderance of genes involved in response to viral infections and xenobiotic metabolisms among the CNV loci. This study is exploratory and provides a basis for further evolutionary studies in Nubian ibex genome.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00239-021-10015-3>.

Acknowledgements We thank Morris Agaba for conceiving the experiment, obtaining the research funding, and for the supervision. The authors would like to thank Joyce Njuguna and John Juma for Bioinformatics support. The bioinformatics analysis was carried out using the high-performance clusters (HPC) courtesy of the Biosciences eastern and central Africa - International Livestock Research Institute (BecA - ILRI).

Author Contributions VC carried out the research and wrote the manuscript. JMM, EM, JDE, AK, and SO provided input on analysis and result interpretations. All authors participated in useful discussion, revised, and approved the final manuscript.

Funding The research was funded by Swedish International Development Cooperation Agency (SIDA) through grants to Biosciences eastern and central Africa—International Livestock Research Institute (BecA—ILRI Hub) (Grant number: UF2011/55504/UD/UP). Vivien Chebii's graduate fellowship was funded by the Deutscher Akademischer Austausch Dienst (DAAD) and was supplemented by BecA-ILRI Hub through Africa Biosciences Challenge Fund (ABCF) program. The ABCF program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill and Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID); and the Swedish International Development Cooperation Agency (SIDA).

Data Availability The sequence data (FASTQ) files used in this study were obtained from public databases under Accession (SRR12990712, SRR8437789 and SRR8437792). Other CNV data generated in this study are provided at Figshare (https://figshare.com/articles/dataset/Copy_number_variants_in_Nubian_ibex/13633943).

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abyzov A, Urban AE, Snyder M, Gerstein M (2011) CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21(6):974–984
- Akram KM, Moyo NA, Leeming GH, Bingle L, Jasim S, Hussain S, Schorlemmer A, Kipar A, Digard P, Tripp RA, Shohet RV, Bingle CD, Stewart JP (2018) An innate defense peptide BPIFA1/SPLUNC1 restricts influenza A virus infection. *Mucosal Immunol* 11(1):71–81. <https://doi.org/10.1038/mi.2017.45>
- Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12(5):363–376
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge
- Bhanuprakash V, Chhotaray S, Pruthviraj DR, Rawat C, Karthikeyan A, Panigrahi M (2018) Copy number variation in livestock: a mini review. *Vet World* 11(4):535–541. <https://doi.org/10.14202/vetworld.2018.535-541>
- Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF (2012) Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res* 22(4):778–790
- Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisà A, Ponce de León FA et al (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet* 49(4):643–650. <https://doi.org/10.1038/ng.3802>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
- Butler MW, Hackett NR, Salit J, Strulovici-Barel Y, Omberg L, Mezey J, Crystal RG (2011) Glutathione S-transferase copy number variation alters lung gene expression. *Eur Respir J* 38(1):15–28
- Cardoso-Moreira M, Arguello JR, Gottipati S, Harshman LG, Grenier JK, Clark AG (2016) Evidence for the fixation of gene duplications by positive selection in *Drosophila*. *Genome Res* 26(6):787–798
- Chorev M, Carmel L (2012) The function of introns. *Front Genet* 3:55–55. <https://doi.org/10.3389/fgene.2012.00055>
- Clarke B, Mahapatra M, Friedgut O, Bumarov V, Parida S (2017) Persistence of Lineage IV Peste-des-petits ruminants virus within Israel since 1993: an evolutionary perspective. *PLoS ONE* 12(5):e0177028
- Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 4(9):1–11
- Di Gerlando R, Mastrangelo S, Moscarelli A, Tolone M, Suter AM, Portolano B, Sardina MT (2020) Genomic structural diversity in local goats: analysis of copy-number variations. *Animals*. <https://doi.org/10.3390/ani10061040>
- Dong Y, Zhang X, Xie M, Arefnezhad B, Wang Z, Wang W, Feng S, Huang G, Guan R, Shen W (2015) Reference genome of wild goat (*capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* 16(1):1–11
- Grassmann F, Cantsilieris S, Schulz-Kuhnt A-S, White SJ, Richardson AJ, Hewitt AW, Vote BJ, Schmied D, Guymer RH, Weber BHF, Baird PN (2016) Multiallelic copy number variation in the complement component 4A (C4A) gene is associated with late-stage age-related macular degeneration (AMD). *J Neuroinflammation* 13(1):81–81. <https://doi.org/10.1186/s12974-016-0548-0>
- Guan D, Martínez A, Castelló A, Landi V, Luigi-Sierra MG, Fernández-Álvarez J, Cabrera B, Delgado JV, Such X, Jordana J, Amills M (2020) A genome-wide analysis of copy number variation in Murciano-Granadina goats. *Genet Sel Evol* 52(1):44. <https://doi.org/10.1186/s12711-020-00564-4>
- Habibi K (1994) The desert ibex: Life history, ecology and behaviour of the Nubian ibex in Saudi Arabia. National Commission for Wildlife Conservation and Development (NCWCD).
- Hakham E, Ritte U (1993) Foraging pressure of the Nubian ibex *Capra ibex nubiana* and its effect on the indigenous vegetation of the En Gedi Nature Reserve, Israel. *Biol Conserv* 63(1):9–21
- Iskrow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. *Trends Genet* 28(6):245–257
- Iyanagi T (2007) Molecular mechanism of Phase I and Phase II drug-metabolizing enzymes: implications for detoxification. In: International review of cytology, vol 260. Academic Press, pp. 35–112. [https://doi.org/10.1016/S0074-7696\(06\)60002-8](https://doi.org/10.1016/S0074-7696(06)60002-8)
- Jacobson ER, Berry KH, Stacy B, Huzella LM, Kalasinsky VF, Fleetwood ML, Mense MG (2009) Oxalosis in wild desert tortoises, *Gopherus agassizii*. *J Wildl Dis* 45(4):982–988. <https://doi.org/10.7589/0090-3558-45.4.982>
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK et al (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61. <https://doi.org/10.1038/nature10944>
- Jüptner M, Flachsbarth F, Caliebe A, Lieb W, Schreiber S, Zeuner R, Franke A, Schröder JO (2018) Low copy numbers of complement C4 and homozygous deficiency of C4A may predispose to severe disease and earlier disease onset in patients with systemic lupus erythematosus. *Lupus* 27(4):600–609. <https://doi.org/10.1177/0961203317735187>
- Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A (2011) Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* 2011:bar030
- Kondrashov FA, Koonin EV (2001) Origin of alternative splicing by tandem exon duplication. *Hum Mol Genet* 10(23):2661–2669
- Lanier LL (2015) NKG2D receptor and its ligands in host defense. *Cancer Immunol Res* 3(6):575–582
- Letunic I, Copley RR, Bork P (2002) Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet* 11(13):1561–1567. <https://doi.org/10.1093/hmg/11.13.1561>
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079
- Liu Y, Bartlett JA, Di ME, Bomberger JM, Chan YR, Gakhar L, Mallampalli RK, McCray PB Jr, Di YP (2013) SPLUNC1/BPIFA1 contributes to pulmonary host defense against *Klebsiella pneumoniae* respiratory infection. *Am J Pathol* 182(5):1519–1531. <https://doi.org/10.1016/j.ajpath.2013.01.050>
- Maréchal J-D, Kemp CA, Roberts GCK, Paine MJ, Wolf CR, Sutcliffe MJ (2008) Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. *Br J Pharmacol* 153(Suppl 1):S82–S89. <https://doi.org/10.1038/sj.bjp.0707570>
- McArdel SL, Terhorst C, Sharpe AH (2016) Roles of CD48 in regulating immunity and tolerance. *Clin Immunol* 164:10–20. <https://doi.org/10.1016/j.clim.2016.01.008>

- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F (2016) The ensembl variant effect predictor. *Genome Biol* 17(1):122. <https://doi.org/10.1186/s13059-016-0974-4>
- Miyagawa H, Yamai M, Sakaguchi D, Kiyohara C, Tsukamoto H, Kimoto Y, Nakamura T, Lee J-H, Tsai C-Y, Chiang B-L, Shimoda T, Harada M, Tahira T, Hayashi K, Horiuchi T (2008) Association of polymorphisms in complement component C3 gene with susceptibility to systemic lupus erythematosus. *Rheumatology* 47(2):158–164. <https://doi.org/10.1093/rheumatology/kem321>
- O'Bleness MS, Dickens CM, Dumas LJ, Kehrer-Sawatzki H, Wyckoff GJ, Sikela JM (2012) Evolutionary history and genome organization of DUF1220 protein domains. *G3* 2(9):977–986
- Okeson DM, Garner MM, Taus NS, Li H, Coke RL (2007) Ibex-associated malignant catarrhal fever in a bongo antelope (*Tragelaphus euryceros*). *J Zoo Wildl Med* 38(3):460–464
- Paudel Y, Madsen O, Megens H-J, Frantz LA, Bosse M, Crooijmans RP, Groenen MA (2015) Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* 16(1):330
- Pezer Ž, Harr B, Teschke M, Babiker H, Tautz D (2015) Divergence patterns of genic copy number variation in natural populations of the house mouse (*Mus musculus domesticus*) reveal three conserved genes with major population-specific expansions. *Genome Res* 25(8):1114–1124
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shaper MH, Carson AR, Chen W (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444–454
- Rigau M, Juan D, Valencia A, Rico D (2019) Intronic CNVs and gene expression variation in human populations. *PLoS Genet* 15(1):e1007902–e1007902. <https://doi.org/10.1371/journal.pgen.1007902>
- Robertson LP, Hall CR, Forster, PI, Carroll AR (2018) Alkaloid diversity in the leaves of Australian *Flindersia* (Rutaceae) species driven by adaptation to aridity. *Phytochemistry* 152:71–81. <https://doi.org/10.1016/j.phytochem.2018.04.011>
- Ross S, Elalqamy H, Al Said T, Saltz D (2020) Capra nubiana. The IUCN Red List of Threatened Species 2020: E. T3796A22143385. <https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T3796A22143385.en>
- Russel FGM, Koenderink JB, Masereeuw R (2008) Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules. *Trends Pharmacol Sci* 29(4):200–207. <https://doi.org/10.1016/j.tips.2008.01.006>
- Wang D, Zou L, Jin Q, Hou J, Ge G, Yang L (2018) Human carboxylesterases: a comprehensive review. *Acta Pharmaceut Sin B* 8(5):699–712. <https://doi.org/10.1016/j.apsb.2018.05.005>
- Wang H, Chai Z, Hu D, Ji Q, Xin J, Zhang C, Zhong J (2019) A global analysis of CNVs in diverse yak populations using whole-genome resequencing. *BMC Genomics* 20(1):61. <https://doi.org/10.1186/s12864-019-5451-5>
- Wensman JJ, Abubakar M, Shabbir MZ, Rossiter P (2018) Peste des petits ruminants in wild ungulates. *Trop Anim Health Prod* 50(8):1815–1819
- Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, Zhang L, Abutarboush MH, Xing Y, Xie Z (2014) Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun* 5(1):1–10
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride KL, Higgins GC, Rennebohm RM, Rice RR, Hackshaw KV, Roubey RAS, Grossman JM, Tsao BP et al (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80(6):1037–1054. <https://doi.org/10.1086/518257>
- Zahedi SM, Karimi M, Venditti A (2019) Plants adapted to arid areas: specialized metabolites. *Nat Prod Res* 1–18. <https://doi.org/10.1080/14786419.2019.1689500>
- Zarama A, Pérez-Carmona N, Farré D, Tomic A, Borst EM, Messerle M, Jonjic S, Engel P, Angulo A (2014) Cytomegalovirus m154 hinders CD48 cell-surface expression and promotes viral escape from host natural killer cell control. *PLoS Pathog* 10(3):e1004000. <https://doi.org/10.1371/journal.ppat.1004000>
- Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, Shao X, Han J, Wan D, Qiu Q (2016) Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics* 17(1):1–12
- Zhou H-D, Li X-L, Li G-Y, Zhou M, Liu H-Y, Yang Y-X, Deng T, Ma J, Sheng S-R (2008) Effect of SPLUNC1 protein on the *Pseudomonas aeruginosa* and Epstein-Barr virus. *Mol Cell Biochem* 309(1):191–197. <https://doi.org/10.1007/s11010-007-9659-3>
- Zingoni A, Molfetta R, Fionda C, Soriani A, Paolini R, Cipitelli M, Cerboni C, Santoni A (2018) NKG2D and its ligands: “one for all, all for one.” *Front Immunol* 9:476. <https://doi.org/10.3389/fimmu.2018.00476>