

ARTICLE

Received 6 Feb 2014 | Accepted 9 Sep 2014 | Published 21 Oct 2014

DOI: 10.1038/ncomms6188

# Camelid genomes reveal evolution and adaptation to desert environments

Huiguang Wu<sup>1,\*</sup>, Xuanmin Guang<sup>2,\*</sup>, Mohamed B. Al-Fageeh<sup>3,\*</sup>, Junwei Cao<sup>1,\*</sup>, Shengkai Pan<sup>2,\*</sup>, Huanmin Zhou<sup>1,\*</sup>, Li Zhang<sup>1</sup>, Mohammed H. Abutarboush<sup>3</sup>, Yanping Xing<sup>1</sup>, Zhiyuan Xie<sup>2</sup>, Ali S. Alsharqeti<sup>3</sup>, Yanru Zhang<sup>1</sup>, Qiulin Yao<sup>2</sup>, Badr M. Al-Shomrani<sup>3</sup>, Dong Zhang<sup>1</sup>, Jiang Li<sup>2</sup>, Manee M. Manee<sup>3</sup>, Zili Yang<sup>1</sup>, Linfeng Yang<sup>2</sup>, Yiyi Liu<sup>1</sup>, Jilin Zhang<sup>2</sup>, Musaad A. Altammami<sup>3</sup>, Shenyuan Wang<sup>1</sup>, Lili Yu<sup>2</sup>, Wenbin Zhang<sup>4</sup>, Sanyang Liu<sup>2</sup>, La Ba<sup>4</sup>, Chunxia Liu<sup>1</sup>, Xukui Yang<sup>2</sup>, Fanhua Meng<sup>1</sup>, Shaowei Wang<sup>2</sup>, Lu Li<sup>1</sup>, Erli Li<sup>2</sup>, Xueqiong Li<sup>1</sup>, Kaifeng Wu<sup>1</sup>, Shu Zhang<sup>2</sup>, Junyi Wang<sup>2,\*</sup>, Ye Yin<sup>2</sup>, Huanming Yang<sup>2</sup>, Abdulaziz M. Al-Swailem<sup>3</sup> & Jun Wang<sup>2,5,6</sup>

Bactrian camel (*Camelus bactrianus*), dromedary (*Camelus dromedarius*) and alpaca (*Vicugna pacos*) are economically important livestock. Although the Bactrian camel and dromedary are large, typically arid-desert-adapted mammals, alpacas are adapted to plateaus. Here we present high-quality genome sequences of these three species. Our analysis reveals the demographic history of these species since the Tortonian Stage of the Miocene and uncovers a striking correlation between large fluctuations in population size and geological time boundaries. Comparative genomic analysis reveals complex features related to desert adaptations, including fat and water metabolism, stress responses to heat, aridity, intense ultraviolet radiation and choking dust. Transcriptomic analysis of Bactrian camels further reveals unique osmoregulation, osmoprotection and compensatory mechanisms for water reservation underpinned by high blood glucose levels. We hypothesize that these physiological mechanisms represent kidney evolutionary adaptations to the desert environment. This study advances our understanding of camelid evolution and the adaptation of camels to arid-desert environments.

<sup>1</sup> College of Life Science, Inner Mongolia Agricultural University, No. 306 Zhaowuda Road, Hohhot 010018, China. <sup>2</sup> BGI-Shenzhen, Shenzhen 518083, China. <sup>3</sup> National Centre for Biotechnology, King Abdulaziz City for Science and Technology (KACST), P.O. Box 6086, Riyadh 11442, Saudi Arabia. <sup>4</sup> Institute of Bactrian Camels, Alxa League 750306, Inner Mongolia, China. <sup>5</sup> Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark. <sup>6</sup> King Abdulaziz University, Jeddah 21589, Saudi Arabia. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.M.AI.-S. (email: alswailem@kacst.edu.sa) or to J.W. (email: wangj@genomics.cn) or to H.Z. (email: huanminzhou@263.net).

The evolution and adaptation of desert animals to their environment at the genome level remains largely unknown. Camels (*Camelus*, Camelini) contain two extant domesticated species: the two-humped Bactrian camel and single-humped dromedary. These species live in the extreme desert environments of Africa and Asia, and their adaptations to arid conditions include a tolerance of temperatures exceeding 40 °C and water losses greater than 25% of total body weight<sup>1</sup> (in non-desert mammals, losses of greater than 15% are fatal<sup>2</sup>). In contrast, their nearest relatives (Lamini), which have no hump and consist of two wild species (guanaco and vicuña) and two domesticated species (llama and alpaca), live in the high altitudes of South America and do not exhibit similar adaptations to hot desert environments (Supplementary Fig. 1). The Bactrian camel, dromedary, llama and alpaca are important not only as beasts of burden for transport but also as sources of meat, milk and wool. In addition, both the dromedary and Bactrian camel represent ideal models for studying desert adaptation among mammals.

Genomic research into the evolutionary history of camelid species and the unusual adaptations of camels to arid-desert environments has been limited, although recent studies have focused on the divergence time between the ancestors of the Bactrian camel and cattle, genomic single-nucleotide polymorphism (SNP) rates, population parameters, artificial domestication, insulin resistance, salt tolerance and other factors<sup>3,4</sup>.

In this study we perform high-quality whole-genome sequencing of the Bactrian camel, dromedary and alpaca, and transcriptome sequencing of the Bactrian camel. Our results reveal the divergence times and demographic history of these species. These genomic and transcriptomic analyses provide evidence of camel adaptation to desert environments. Our analysis also provides numerous insights into the unusual physiological features and genetic architecture that enable camels to survive in desert environments.

## Results

**Genome sequencing and assembly.** We sequenced the genomes of a female Bactrian camel (79.3-fold coverage), a male dromedary (65.0-fold coverage) and a female alpaca (72.5-fold coverage) using the Illumina HiSeq2000 platform. The present estimated genome size of the Bactrian camel (2.45 Gb) is comparable with that of a previous report (2.38 Gb) based on K-mer analysis<sup>3</sup>. The assembled genome sizes for the three individuals were 2.01, 2.01 and 2.05 Gb, respectively (Supplementary Tables 1–10 and Supplementary Figs 2 and 3). The present assembled genome size for the Bactrian camel is identical to the previously reported size<sup>3</sup>. The contig N50 and scaffold N50 lengths (Table 1) were 24.9 kb and 8.7 Mb for the Bactrian camel, 54.1 kb and 4.1 Mb for the dromedary, and 66.3 kb and 5.1 Mb for the alpaca, respectively. Compared with the genome of the wild Bactrian camel<sup>3</sup>, the present genomes of these three camelids have shorter contig N50 lengths but larger scaffold N50 lengths. Mapping of libraries with an insert size of 2 kb to the scaffold indicated that each of the genome sequences was of high quality (Supplementary Fig. 4 and Supplementary Methods), and the Bactrian camel transcriptome also demonstrated a high-quality

genome assembly for the present and wild Bactrian camel<sup>3</sup> (Supplementary Tables 11 and 12). The camelid genomes shared high synteny with the human and cattle reference genomes (coverage rate >83%) and a relatively low rate of genomic rearrangement within the Camelidae (Supplementary Tables 13 and 14, and Supplementary Methods). The synteny between the Bactrian camel and cattle genomes observed in the present study is greater than that reported previously<sup>3</sup>. Our study supports the notion that divergent evolution in Camelidae occurred through single gene mutations or minor chromosomal rearrangements<sup>5</sup>. We estimated the segmental duplication of these three individuals: the total length of segmental duplication in both the Bactrian camel and dromedary was 26 Mb, smaller than that of the alpaca (36 Mb) (Supplementary Table 15). The segmental duplication among these three organisms is lower than that reported in cattle (94.4 Mb)<sup>6</sup>.

**Genome annotation.** Using a combination of homologous sequence searches and *ab initio* gene predictions, we annotated 20,251, 20,714 and 20,864 genes in the Bactrian camel, dromedary and alpaca genomes, respectively (Supplementary Fig. 5, and Supplementary Tables 16 and 17). We used the CEGMA method<sup>7</sup>, which includes 458 core eukaryotic genes, to assess the completeness of the genomes and annotation. The vast majority of these core genes were aligned to the camelid genomes (99.12% for Bactrian camel, 98.47% for dromedary and 99.12% for alpaca), and the majority were present in our predicted gene sets (97.82% for Bactrian camel, 96.73% for dromedary and 93.87% for alpaca), supporting the completeness of the assembled genomes and gene sets identification (Supplementary Tables 18–20). Comparative analyses of the three camelid gene sets revealed a high gene sequence similarity (>90%) but different non-synonymous/synonymous (*Ka/Ks*) distributions (Supplementary Figs 6 and 7). Functional analyses of the gene sets indicated that >91% of the genes were functionally annotated in each genome (Supplementary Tables 21–23).

The repeat sequence content of the camelid genomes (30.4% in Bactrian camel, 32.1% in alpaca and 28.4% in dromedary) was 10% lower than that of cattle (42.5%) and humans (46.1%) due to the small number of short interspersed nucleotide elements in the camelid genomes (Supplementary Tables 24–27). The repeat sequence content of the Bactrian camel genome was similar to that reported previously<sup>3</sup>. Annotation of non-coding RNA genes revealed similar copy numbers for each genome (Bactrian camel = 1,942; dromedary = 2,209; alpaca = 2,328; Supplementary Tables 28–30). We identified 12,539 homologous gene families that are shared by 4 species in the order Cetartiodactyla (Bactrian camel, dromedary, alpaca and cattle): 156, 153 and 296 gene families were specific to the Bactrian camel, dromedary and alpaca, respectively (Fig. 1).

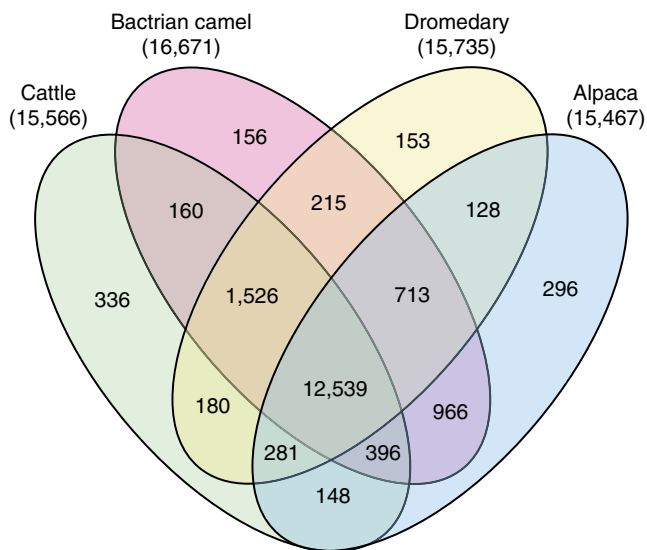
**Evolutionary analysis and phylogeny.** A phylogenetic tree was constructed including the camelids (Bactrian camel, dromedary and alpaca) and seven other species (cattle, horse, dog, panda, human, mouse and opossum). The tree was generated using PhyML<sup>8</sup> based on fourfold degenerate codon sites extracted from

**Table 1 | Genomes statistics for three camelids.**

Species	Genome coverage (X)	Genome size (Gb)	Contig N50 (kb)	Scaffold N50 (Mb)	GC content (%)	Repeat content (%)
Bactrian camel	79.3	2.01	24.9	8.7	41.3	30.4
Dromedary	65.0	2.01	54.1	4.1	41.2	28.4
Alpaca	72.5	2.05	66.3	5.1	41.4	32.1

7,398 single-copy orthologous genes identified by TreeFam<sup>9</sup> (Supplementary Table 31, and Supplementary Figs 8 and 9). The estimated divergence time between camelids and cattle is 42.7 million years ago (Mya) (Fig. 2 and Supplementary Fig. 10). This result is consistent with the time (45.9 Mya) at which palaeontological evidence indicates that the Camelidae family first appeared in North America<sup>10</sup> but is in contrast with a previous estimate of the time of divergence of the cattle and Bactrian camel lineages based on 332 orthologues (55–60 Mya)<sup>3</sup>. The estimated divergence time of the ancestors of alpaca and the two camels (16.3 Mya) is consistent with palaeontological

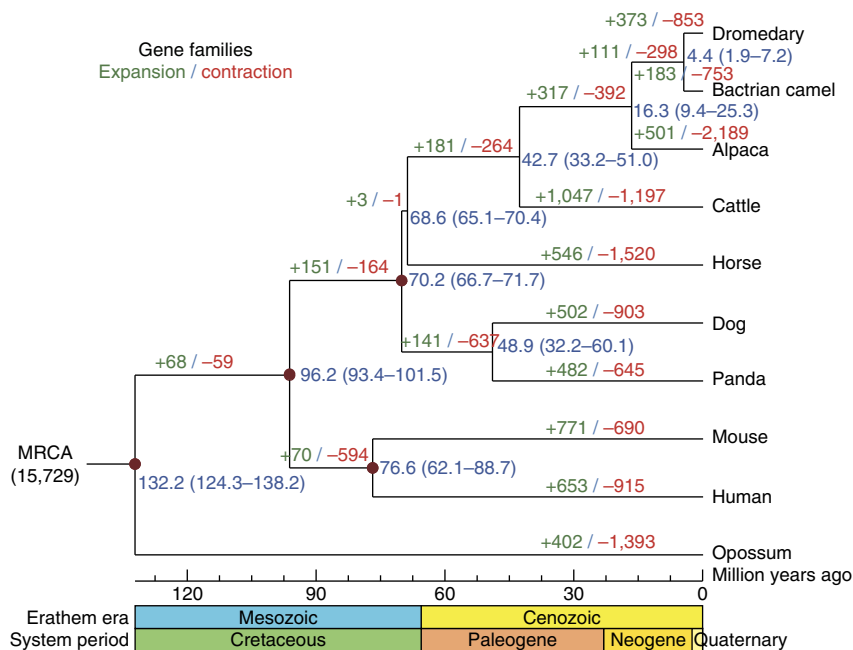
findings, indicating that the division between Camelini and Lamini occurred in North America ~17 Mya (ref. 10). The divergence time between the Bactrian camel and dromedary is ~4.4 Mya, implying they probably diverged after their common ancestor migrated from North America to Eurasia via the Bering Isthmus during the Late Miocene (7.246–4.9 Mya)<sup>10,11</sup>. We analysed the branch-specific *Ka/Ks* substitution ratios ( $\omega$ ) for these ten mammals using the method of Kosiol *et al.*<sup>12</sup>; the Bactrian camel and dromedary had higher branch  $\omega$ -values (Supplementary Fig. 11, Supplementary Table 32 and Supplementary Methods). This accelerated evolution in camels raises the possibility of camel-specific evolution to adapt to a desert environment.



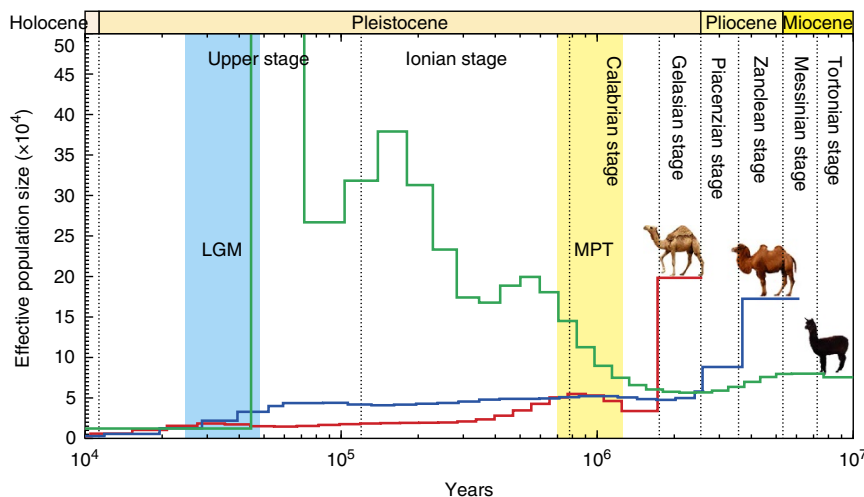
**Figure 1 | Unique and homologous gene families.** The number of unique and shared gene families is shown in each of the diagram components and the total number of gene families for each animal is given in parentheses.

**Heterozygote rates and demographic history.** SNPs were identified using SOAPsnp<sup>13</sup>. The estimated heterozygote rates of the Bactrian camel, dromedary and alpaca genomes were  $1.16 \times 10^{-3}$ ,  $0.74 \times 10^{-3}$  and  $2.66 \times 10^{-3}$ , respectively (Supplementary Tables 33–35). The heterozygote rate of the Bactrian camel estimated here is comparable to that reported previously ( $1.0 \times 10^{-3}$  and  $1.29 \times 10^{-3}$ )<sup>3,4</sup>. The genomic SNP distributions among these mammals are different (Supplementary Fig. 12).

The demographic history of these camelids was constructed based on SNP data by applying the pair-wise sequentially Markovian coalescent (PSMC) model<sup>14</sup> (Fig. 3). The results of our analysis indicated that the ancestor of the Bactrian camel had stable population sizes following two declines that occurred 3.69 and 2.61 Mya. Two declines in population size occurring 1.72 and 0.77 Mya were calculated for the ancestor of the dromedary. These estimated declines in the population size of the ancestors of both species are consistent with transitions between geological ages, including the Zanclean and Piacenzian (3.60 Mya), the Piacenzian and Gelasian (2.59 Mya), the Gelasian and Calabrian (1.81 Mya), and the Calabrian and Ionian (0.78 Mya)<sup>15</sup>, suggesting a probable correlation. Moreover, the expansion of



**Figure 2 | Phylogenomic analysis.** The number of expanded (green) and contracted (red) gene families is shown at each branch. The red points on the four internal nodes indicate the fossil calibration times that were used in the analysis. The estimated divergence time (Mya) of each evolutionary lineage is shown in blue. The blue numbers in brackets are confidence intervals. MRCA, most recent common ancestor.



**Figure 3 | Demographic history.** The blue, red and green line represent the estimated population size of Bactrian camel, dromedary and alpaca, respectively. The geological time boundaries<sup>15</sup> of each unit from the Miocene Epoch to the Holocene Epoch are marked using broken lines. The middle Pleistocene transition (MPT) is highlighted in orange, while the Last Glacial Maximum (LGM) of South America is highlighted in blue.

the ancestral population of the dromedary occurred between 1.25 and 0.77 Mya, coinciding with the middle Pleistocene transition from 1.25 to 0.70 Mya, a period of fundamental changes in the Earth's climatic cyclicality<sup>16</sup> that had a profound effect on the distribution and evolution of biota<sup>17</sup>. This time interval also coincides with the Galerian Mammal Age (1.2 to 0.60 Mya), which was characterized by a renewal of fauna that, in some cases, gave rise to new species that were adapted to arid, cold climates<sup>18</sup>; more importantly, however, this time interval also coincides with the maximal diversity of the family Camelidae, which occurred in the early Galerian<sup>19</sup>. This correlation supports the adaptation of the ancestor of the dromedary to environmental changes and an expansion of its population during the middle Pleistocene transition. The most recent decrease in the population of the Bactrian camel ancestor occurred ~60 thousand years ago (Kya), which corresponds to the dispersal of modern humans out of Africa to Eurasia<sup>20</sup>, home of the Bactrian camel. Therefore, human activities may have impacted the recent ancestral population of the Bactrian camel.

The effective population size of the ancestor of the alpaca gradually decreased between ~5.37 Mya, which is closer to the time boundary of the Messinian and Zanclean stage (5.33 Mya)<sup>15</sup>, and 2.09 Mya, which is in the Uquian age (3 to 1.2 Mya), during which the ancestor of the alpaca migrated to South America through the Panamanian land bridge in the Great American Biotic Interchange<sup>21</sup>. This suggests that the migration may have contributed to the reduction in the population size of the ancestor of the alpaca. Its population size then expanded during the Pleistocene, followed by three periods of major bottlenecks before 501, 139 and 44 Kya. The population underwent a major expansion ~72 Kya, reaching a size of  $\sim 113 \times 10^4$  individuals. The most recent bottleneck (44 Kya) corresponds to the Last Glacial Maximum (48–25 Kya), which was advanced in South America<sup>22</sup>, and resulted in a dramatic reduction in population size to  $\sim 1.2 \times 10^4$  individuals. This implies that the cold conditions in South America at that time may have resulted in the constriction of the population size of the ancestor of the alpaca towards the end of the Pleistocene.

**Gene evolution.** We next investigated camelid genes that underlie adaptation to the environment. We adopted CAFE<sup>23</sup> to identify gene families that have undergone significant expansion and contraction during evolution (Fig. 2 and Supplementary

Methods) and identified 373 expanded and 853 contracted gene families in the dromedary genome, 183 expanded and 753 contracted gene families in the Bactrian camel genome and 501 expanded and 2,189 contracted gene families in the alpaca genome. Many of the expanded gene families in these three camelids are significantly enriched in the cellular process, cell part, olfactory receptor activity, iron and immune-related Gene Ontology (GO) categories (Supplementary Figs 13–15 and Supplementary Tables 36–38). We identified 287 positively selected genes (PSGs) in the Bactrian camel (Supplementary Data 1), 324 PSGs in the dromedary (Supplementary Data 2) and 151 PSGs that were common to both genomes, indicating similar selective pressures. An assessment of unique amino acid residue changes in orthologous genes that are present in 23 species identified 350 and 343 altered genes in the Bactrian camel and dromedary, respectively. Several overrepresented categories of genes with unique amino acid residue changes in camels were related to catalytic activity, small molecule binding and ATP binding (Supplementary Figs 16 and 17, and Supplementary Tables 39 and 40). Based on an analysis of syntenic blocks, 190 gained genes were identified in the Bactrian camel and 126 in the dromedary. These gained genes are significantly enriched in the olfaction and immune-related categories (Supplementary Tables 41 and 42, and Supplementary Methods).

**Energy and fat metabolism.** As energy is important for camels living in food-scarce deserts, the selection of genes involved in energy-related processes was analysed. The genome-wide features of adaptation were identified by GO categories with lineage-specific accelerated evolution (Supplementary Data 3–14). In contrast to cattle, the common rapidly evolving GO categories of the three camelids included cellular response to insulin stimulus (GO:0032869,  $P < 0.001$ ) and the insulin receptor signalling pathway (GO:0008286,  $P < 0.001$ ) (Supplementary Data 4, 8 and 14). In addition, we identified a number of categories associated with energy, glucose and fat metabolism that evolved more rapidly in these camelids than in cattle. Some of the energy-related GO categories identified as evolving more rapidly in the Bactrian camel than in cattle are consistent with those reported previously<sup>3</sup>. Moreover, 13 genes involved in mitochondrial function,  $\beta$ -oxidation and cholesterol synthesis and transport had amino acid residue changes that were unique to the Bactrian camel and dromedary. Several genes (*ACC2*, *DGKZ* and *GDPD4*)



involved in fat metabolism underwent expansion in the Bactrian camel genome, while the expanded gene families of the dromedary were enriched in the mitochondrion (GO:0005739,  $P = 2.30 \times 10^{-5}$ ) category (Supplementary Table 37).

The different number of humps in these three camelids may reflect their distinct fat metabolism abilities. Functional categories associated with ATP (GO:0006200, GO:0016887, GO:0042626,  $P < 0.01$ ), mitochondria (GO:0005739, GO:0005759,  $P < 0.01$ ), lipid transport (GO:0006869,  $P_{\text{Bactrian camel}} = 5.33 \times 10^{-5}$ ,  $P_{\text{dromedary}} = 0.00016$ ) and response to insulin stimulus (GO:0032868,  $P_{\text{Bactrian camel}} = 0.0005$ ,  $P_{\text{dromedary}} = 1.33 \times 10^{-5}$ ) evolved rapidly in both camel species compared with the alpaca (Supplementary Table 43). Categories associated with lipid metabolism evolved more rapidly in the Bactrian camel than in the dromedary, for example, lipid catabolic process (GO:0016042,  $P = 0.0015$ ) and fat cell differentiation (GO:0045444,  $P = 2.54 \times 10^{-9}$ ) (Supplementary Table 44). These genes may enhance the energy storage and production capacity of a camel in the desert and may also reflect a difference in fat metabolism and in turn be related to the number of humps.

**Stress response.** To investigate adaptations to arid and hot environments, we further analysed genes involved in stress responses. Compared with cattle, categories associated with DNA damage and repair (GO:0006974, GO:0003684, GO:0006302,  $P < 0.01$ ), apoptosis (GO:0006917, GO:0043066,  $P < 0.01$ ), protein stabilization (GO:0050821,  $P_{\text{Bactrian camel}} = 0.00021$ ,  $P_{\text{dromedary}} = 3.44 \times 10^{-19}$ ) and immune responses (GO:0006955, GO:0051607,  $P < 0.01$ ) exhibited accelerated evolution in both camel species (Supplementary Data 8 and 14). Compared with the alpaca, significant functional categories were identified for T-cell co-stimulation (GO:0031295,  $P_{\text{Bactrian camel}} = 8.67 \times 10^{-32}$ ,  $P_{\text{dromedary}} = 9.33 \times 10^{-9}$ ), oxidation–reduction processes (GO:0055114,  $P_{\text{Bactrian camel}} = 4.88 \times 10^{-15}$ ,  $P_{\text{dromedary}} = 5.22 \times 10^{-21}$ ) and oxidoreductase activity (GO:0016491,  $P_{\text{Bactrian camel}} = 2.27 \times 10^{-10}$ ,  $P_{\text{dromedary}} = 7.23 \times 10^{-7}$ ), all of which exhibited accelerated evolution in both camels (Supplementary Data 6 and 12). Three genes (*ERP44*, *NFE2L2* and *MGST2*) were correlated with oxidative stress responses and featured unique amino acid residue changes in both camel genomes. The expanded gene families of the dromedary were enriched in cytochrome *c* oxidase activity (GO:0004129,  $P = 5.80 \times 10^{-10}$ ) and monooxygenase activity (GO:0004497,  $P = 1.32 \times 10^{-5}$ ) (Supplementary Table 37). These results provide evidence of selection in camels to adapt to the harsh arid conditions of the desert environment.

**Adaptation of the respiratory system.** Another challenge of the desert environment is airborne dust, which can lead to respiratory diseases such as asthma. Thirteen PSGs in both camels, including *FOXP3*, *CX3CR1*, *CYSLTR2* and *SEMA4A*, were related to respiratory diseases in humans. We also found that the lung development GO category (GO:0030324,  $P_{\text{Bactrian camel}} = 3.26 \times 10^{-5}$ ,  $P_{\text{dromedary}} = 1.18 \times 10^{-19}$ ) (Supplementary Data 6 and 12) evolved rapidly in the dromedary and Bactrian camel compared with the alpaca. Selection of these genes provides further evidence of the adaptation of camels to endure the challenges of the desert environment.

**Adaptation of the visual system.** Solar radiation is another aspect of the desert environment. Long-term exposure to ultraviolet radiation can lead to a number of ophthalmic conditions. We examined genes that might accustom the eyes of camels to the extreme solar irradiation in the desert and identified positive selection at the *OPN1SW*, *CX3CR1* and *CNTFR* genes, which are

related to photoreception and visual protection, in both camels. The results also indicated that visual perception (GO:0007601,  $P_{\text{Bactrian camel}} = 0.0018$ ,  $P_{\text{dromedary}} = 2.49 \times 10^{-14}$ ) evolved rapidly in both camels compared with the alpaca (Supplementary Data 6 and 12). These results suggest a genetic basis for the ability of camels to endure prolonged exposure to ultraviolet light without damage to the visual system.

**Salt metabolism.** We then focused on the salt metabolism of camels by considering the main effect of salt on water balance. In contrast to a previous report on salt tolerance<sup>3</sup>, our results indicated that the category of sodium ion transport (GO:0006814,  $P_{\text{Bactrian camel}} = 0.0014$ ,  $P_{\text{dromedary}} = 0.00012$ ) evolved more rapidly in both camels than in cattle (Supplementary Data 8 and 14). The category associated with the voltage-gated potassium channel complex (GO:0008076,  $P_{\text{Bactrian camel}} = 8.77 \times 10^{-8}$ ,  $P_{\text{dromedary}} = 2.68 \times 10^{-10}$ ) evolved rapidly in both camels compared with the alpaca (Supplementary Data 6 and 12). Notably, the genome of the Bactrian camel contains two copies of the *NR3C2* and *IRS1* genes, both of which play critical roles in sodium reabsorption and water balance in the kidney<sup>24–26</sup>, whereas other mammals possess only a single copy of each gene. This difference suggests that camels may metabolize and transport salt more efficiently than alpaca and cattle, and these pathways are important for water reabsorption.

**Differentially expressed genes and enrichment analysis.** To obtain greater insight into the characteristics of adaptation to the arid desert, we sequenced the renal cortical and medullary transcriptomes of a group of Bactrian camels after 24 days of water-restricted (WR) conditions and those of a control group (CG) (Supplementary Table 45, and Supplementary Data 15 and 16). We selected significantly upregulated or downregulated genes in these tissues (Supplementary Figs 18–21 and Supplementary Methods) and then analysed the enriched GO categories of these genes (Supplementary Figs 22–25, Supplementary Data 17–20 and Supplementary Methods). An overrepresentation of categories associated with metal ion binding (GO:0046872,  $P = 1.53 \times 10^{-23}$ ) and the regulation of body fluid levels (GO:0050878,  $P = 1.37 \times 10^{-6}$ ) was detected in the set of upregulated renal cortical genes (Supplementary Data 17). The GO categories associated with the glucose metabolic process (GO:0006006,  $P = 4.11 \times 10^{-6}$ ), gluconeogenesis (GO:0006094,  $P = 0.0026$ ), mitochondrion (GO:0005739,  $P = 2.13 \times 10^{-5}$ ), the generation of precursor metabolites and energy (GO:0006091,  $P = 0.0077$ ), response to nutrient levels (GO:0031667,  $P = 0.0064$ ) and response to stress (GO:0006950,  $P = 0.0094$ ) were enriched in the set of upregulated renal medullary genes (Supplementary Data 19).

**Sodium reabsorption.** Genes encoding the  $\text{Na}^+/\text{K}^+$ -ATPase and the epithelial  $\text{Na}^+$  channel (ENaC), which reabsorb sodium in the kidney, were upregulated in the renal cortex and medulla under WR conditions (Supplementary Tables 46 and 47). The flexible transcription of the subunits of ENaC in different tissues and under different conditions suggests that the camel regulates the  $\text{Na}^+$  reabsorbing activity of ENaC to cope with different physiological water requirements. These findings indicate that the regulation of sodium reabsorption may be essential for the survival of camels in a water-scarce environment.

**Water reservation.** The camel is renowned for its adaptation to prolonged water restriction. We therefore investigated the mechanism of water reservation by analysing the transcription of the aquaporin family genes, which are selective water channels with important functions in water reabsorption and metabolism.

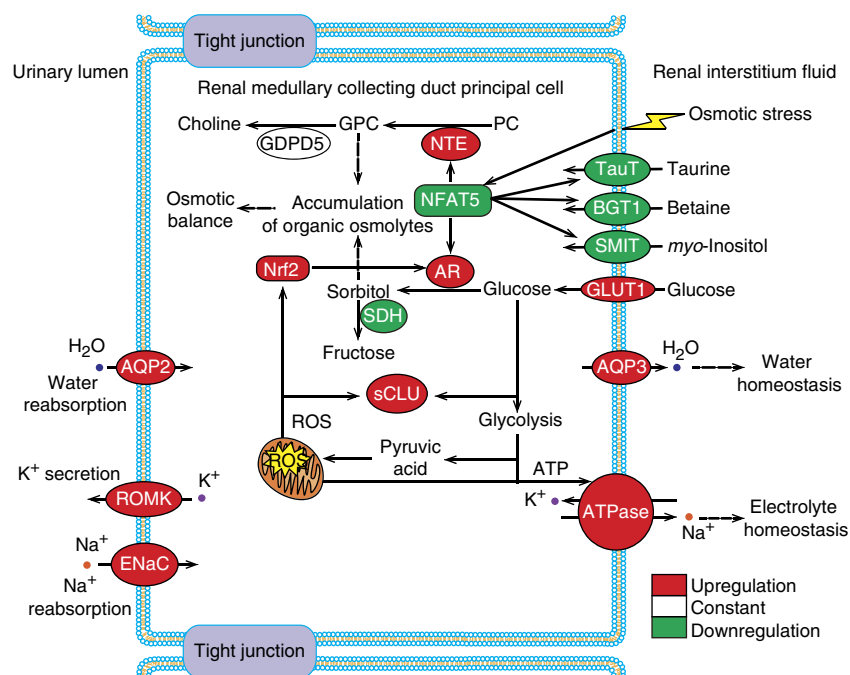
*AQP1*, *AQP2* and *AQP3* were the top three differently expressed genes in the renal cortex and medulla under WR conditions (Supplementary Tables 48 and 49, and Supplementary Fig. 26). These genes may allow camels to reabsorb water more efficiently in a water-scarce environment. However, we did not detect *AQP4* mRNA in the Bactrian camel kidney, consistent with its lack of expression in the desert rodent *Dipodomys merriami*<sup>27</sup> but in contrast to its abundant expression in the human kidney<sup>28</sup>. Interestingly, a unique amino acid residue change (R261C) was observed in *AQP4* in the Bactrian camel genome (Supplementary Fig. 27). These findings may suggest a unique strategy for water reabsorption and metabolism in the camel kidney.

**Osmoregulation.** As hypertonicity is the foundation of water balance and reabsorption in the kidney, the expression of genes that are involved in osmoregulation in the renal medulla were analysed. Nuclear factor of activated T-cells 5 (*NFAT5*), the only known tonicity-regulated transcription factor in mammals<sup>29</sup>, was expressed at 3.66% of the control level under WR conditions (Supplementary Table 50). Accordingly, sodium/*myo*-inositol cotransporter (*SMIT*), sodium- and chloride-dependent taurine transporter (*TauT*) and sodium- and chloride-dependent betaine transporter (*BGT1*) exhibited reduced expression under WR conditions. These three transporters transactivated by *NFAT5* transport compatible organic osmolytes into renal medullary cells (RMCs) in response to hypertonicity<sup>30</sup> (Fig. 4). The downregulation of *NFAT5* and its target genes during hypertonic stress has not been observed in other mammals<sup>29,31</sup>, including desert animals such as the Spinifex hopping mouse (*Notomys alexis*)<sup>32</sup>. Our findings indicate that camels may rely on other osmoregulatory strategies to protect against hypertonic stress during long-term water restriction.

**Organic osmolytes.** The accumulation of organic osmolytes helps RMCs to balance the osmotic pressure between the intracellular

and extracellular environments<sup>30</sup>. The downregulation of *TauT*, *BGT1* and *SMIT* implies that the transport of taurine, betaine and *myo*-inositol into cells is decreased. Remarkably, we observed the transcriptional upregulation of aldose reductase (*AR*) and the downregulation of sorbitol dehydrogenase (*SDH*) in the sorbitol pathway; we also observed the transcriptional upregulation of neuropathy target esterase (*NTE*) and the stable transcription of glycerophosphodiester phosphodiesterase domain-containing protein 5 (*GDPD5*) in the glycerophosphocholine (GPC) pathway (Fig. 4 and Supplementary Table 50). The expression patterns of these genes suggest that in the camel, sorbitol and GPC may accumulate under WR conditions and that osmolytes may be mainly produced by the RMCs themselves. Sorbitol can serve as a source of energy<sup>33</sup> and help balance the osmolality of high extracellular NaCl<sup>34</sup>; the energetic cost of the accumulation of GPC in response to high NaCl or urea in the renal medulla<sup>30</sup> may be less than that of transporting betaine into cells against a high concentration gradient<sup>30</sup>. Thus, these variations in the expression of osmolyte-related genes indicate that two osmolytes rather than five are principally used in response to hypertonicity as part of a low-energy-consumption model for camel survival in the food-scarce desert.

Importantly, we observed that the expression levels of *GLUT1* (glucose transporter 1) and genes involved in glycolysis were profoundly increased in the renal medulla under WR conditions (Supplementary Table 51). Together with a previous report that the expression level of *GLUT1* is induced by osmotic and metabolic stress<sup>35</sup>, our results suggest that the enhancement of glucose intake not only ensures a sufficient glucose concentration for the synthesis of sorbitol but also supplies the energy required for the upregulated Na/K-ATPase to maintain the internal ion gradient for adapted hypertonicity (Fig. 4). Collectively, our observations suggest that the characteristic high blood glucose (6–8 mmol l<sup>-1</sup>)<sup>36,37</sup> of camels may be an adaptive evolutionary strategy for osmoregulation and water reabsorption of RMCs during antidiuresis.



**Figure 4 | Schematic model of renal medullary osmosis and water reservation in the camel under water restriction.** The shading of boxes indicates the upregulation (red), constant expression (white) or downregulation (green) of genes in the renal medulla of Bactrian camels during WR. The dashed lines indicate the ultimate functions or effects of the gene expression and the related product's activities.

**Osmoprotection.** Given the potential for hyperosmotic damage to cells<sup>30</sup>, we analysed the expression of genes related to cell protection and found that the expression levels of 25 genes encoding antioxidants and related enzymes (Supplementary Table 52) were higher in the renal medulla under WR conditions. Genes encoding antioxidative transcription factors, including Nrf2, heat shock factor-1, activator protein-1 complex, p53, nuclear factor- $\kappa$ B and signal transducer and activator of transcription 4 also exhibited elevated expression in the WR renal medulla. In addition, we identified 14 heat shock genes, which contribute to the elimination of misfolded proteins under hyperosmolality<sup>30</sup>, that were upregulated in the WR renal medulla (Supplementary Table 52). Gene *clusterin*, a cytoprotective chaperone, was dramatically increased by ~8.9-fold and had the highest transcription level in the WR renal medulla (reads per kilobase per million mapped reads = 27,069). Previous studies have shown that clusterin is induced by glucose<sup>38</sup> and associated with diverse pathological states, including diabetes<sup>39</sup> and kidney injury<sup>40</sup>. The identification of *clusterin* as a PSG in the dromedary suggests that this gene may play a major role in the cytoprotection of the camel renal medulla during water restriction and that the high level of blood glucose in camels may serve a function during osmoprotection. Overall, the upregulation of osmoprotective genes indicates that camels have a sophisticated osmoprotective capability under WR conditions.

## Discussion

Using deep sequencing and *de novo* assembly, we generated high-quality genome sequences of the Bactrian camel, dromedary and alpaca. Comparative genomic analyses provided new insights into the evolutionary history of these camelids and the adaptations of camels to the harsh desert environment. Transcriptome analyses revealed a network of genes that are involved in water reabsorption and glucose function in osmoregulation and water reservation, which suggests that the hyperglycemia that is characteristic of camels may be an adaptation to life in the water-scarce desert. The identification of key genes involved in the adaptation to the desert environment may have applications in breeding programmes and may provide some perspective for disease-resistance research. Future studies on camel genomes and transcriptomes may contribute to a detailed understanding of these important physiological mechanisms with relevance to human medical conditions (for example, the links between sodium metabolism and hypertension, hyperglycemia and diabetes, fat metabolism and obesity, and dust and respiratory diseases). Finally, in light of ongoing global warming and the increasing incidence of droughts, these camelid genomes are valuable resources for studying biological adaptations to environmental changes.

## Methods

**Genome sequencing and assembly.** We used a whole-genome shotgun strategy and next-generation sequencing technologies on an Illumina HiSeq 2000 platform to sequence the genomes of a female Bactrian camel, a male dromedary and a female alpaca from blood DNA. A series of paired-end sequencing libraries with insert sizes of 170, 500 and 800 bp, and 2, 5 and 10 kb was constructed and sequenced for the Bactrian camel, dromedary and alpaca, respectively. An additional library with insert sizes of 20 kb was constructed for the Bactrian camel. The genomes of these three organisms were assembled using SOAPdenovo (version 1.05)<sup>41</sup>.

**Gene prediction.** We used both homology-based and *de novo* methods to predict genes in the three genomes. For the homology-based prediction, gene sets from other species were mapped onto the genomes using TblastN<sup>42</sup> ( $E$ -value  $\leq 1e-5$ ). Homologous genome sequences were then aligned to the matching proteins that were identified as having synteny using Genewise<sup>43</sup>, to define gene models. For *de novo* prediction, Augustus<sup>44</sup> and Genscan<sup>45</sup> were used to predict coding genes. Finally, the homology-based and *de novo*-derived gene sets were merged using

GLEAN (<http://sourceforge.net/projects/glean-gene/>). For the Bactrian camel, LASTZ<sup>46</sup> was used to predict genes. Genes with coding lengths < 150 bp were filtered out to reduce the number of false positives.

**Gene set evaluation and function annotation.** The CEGMA<sup>7</sup> pipeline, which contains 458 core eukaryotic genes that are highly conserved across a wide range of eukaryotes, was used to assess the completeness of the genomes and gene sets. The functions of genes were aligned based on the best match derived from the alignments to proteins annotated in the SwissProt, TrEMBL<sup>47</sup>, InterProScan<sup>48</sup> and Kyoto Encyclopedia of Genes and Genomes protein databases using Blastp.

**Repeat annotation and non-coding RNA annotation.** Tandem repeats were predicted using Tandem Repeats Finder<sup>49</sup> (version 4.04). Transposable elements (TEs) were identified in the genomes using a combination of homology-based and *de novo* approaches. For the homology-based approach, we first identified known TEs using RepeatMasker programme<sup>50</sup> against the Repbase<sup>51</sup> (version 16.10) database. We then used RepeatProteinMask<sup>50</sup> to identify TEs by aligning the genome sequence to the TE protein database. For the *de novo* method, we constructed a repeat library generated by RepeatModeler<sup>52</sup>, and RepeatMasker was used to identify the repeat sequences. The transfer RNA genes were predicted based on tRNAscan-SE<sup>53</sup>, and the ribosomal RNA fragments were identified by aligning human ribosomal RNA sequences to the camelid genomes using BLASTN. The micro RNA and small nuclear RNA genes were predicted by INFERNAL<sup>54</sup> software against the Rfam database<sup>55</sup> (Release 9.1) using the family-specific 'gathering' cutoff of Rfam.

**Gene family cluster and phylogenetic tree.** DNA and protein sequences for seven other mammals (human, mouse, cattle, horse, dog, panda and opossum) were downloaded from the Ensembl database (release 60). We used Treefam<sup>56</sup> methodology to define gene families. We constructed a phylogenetic tree based on 7,398 single-copy orthologous genes from these 10 species. Fourfold degenerate sites were extracted from each single family and concatenated to one supergene for each species. PHYML<sup>57</sup> was used to build phylogenetic trees under GTR + gamma for fourfold degenerate sites.

**Divergence time.** The coding sequences of 7,398 single-copy orthologous genes were used to estimate divergence times based on the phylogenetic tree. The PAML MCMCTREE<sup>58</sup> performs Bayesian estimation of species divergence times using soft fossil constraints<sup>59</sup>. The fossil calibration times were derived from a website (<http://www.fossilrecord.net>).

**Reconstruction of demographic history.** We used a PSMC model<sup>14</sup> with the generation time and mutation rates to reconstruct the most recent common ancestor distribution. We aligned short reads to the reference genome using BWA (version 0.5.9)<sup>60</sup>. The consensus sequences were obtained using SAMtools (version 0.1.17)<sup>61</sup> and divided into non-overlapping 100 bp bins. The resulting binned sequences were used as the input for the PSMC estimation.

**Positive selection.** We used PAML<sup>58</sup> to detect genes under positive selection. The branch-site model was used to identify positive genes for the two camels. We compared ModelA1, in which sites may evolve neutrally and under purifying selection, with ModelA, which allows sites to also be under positive selection.  $P$ -values were computed using the  $\chi^2$ -statistic adjusted by the false discovery rate method to allow for multiple testing.

**Rapidly evolving GO.** The approach based on  $Ka/Ks$  was used to identify GO categories that were significantly enriched or depleted in the object species. PRANK<sup>62</sup> was used to perform multiple sequence alignment of the single copies, and the lower-quality alignments were filtered using Gblocks<sup>63</sup>. The  $Ka$  and  $Ks$  values were then estimated using PAML. For a given GO category with more than 20 orthologues, the probability of observing an equal or higher number of non-synonymous substitutions was calculated assuming a binomial distribution. To determine whether a subset of the categories are evolving under significantly high (low) constraints, we repeated this procedure 10,000 times on the same data set after randomly permuting the GO annotations to test whether the probability was less than a threshold value. Finally, we selected the GO categories with  $P$ -values  $\leq 0.01$  for further analysis.

**Transcriptome sequencing and analysis.** All RNA samples were isolated from the renal cortex and medulla of eight female Bactrian camels. Four were randomly chosen as the CG, and the others were assigned to the WR. The Bactrian camels were fed hay with no water content *ad libitum*. The Bactrian camels in the WR group were not given water for 24 days, while the CG Bactrian camels drank water freely. Total RNA derived from the same treatment and same tissue were equally mixed to produce a single pool. After total RNA extraction, libraries were generated and sequenced using an Illumina HiSeq 2000 platform.



Gene expression levels were measured by reads per kilobase per million mapped reads<sup>64</sup>. Differentially expressed genes (DEGs) were detected according to a method<sup>65</sup>. All DEGs were mapped to GO terms in the database (<http://www.geneontology.org/>) and the number of genes for every term were calculated to identify GO terms that were significantly enriched in the input list of DEGs.

## References

- Schmidt-Nielsen, K. The physiology of the camel. *Sci. Am.* **201**, 140–151 (1959).
- Ingram, D. L. & Mount, L. E. *Man and Animals in Hot Environments* (Springer-Verlag, 1975).
- Jirimutu *et al.* Genome sequences of wild and domestic bactrian camels. *Nat. Commun.* **3**, 1202 (2012).
- Burger, P. A. & Palmieri, N. Estimating the population mutation rate from a de novo assembled Bactrian camel genome and cross-species comparison with dromedary ESTs. *J. Hered.* doi: 10.1093/jhered/est005 (2013).
- Taylor, K. M., Hungerford, D. A., Snyder, R. L. & Ulmer, Jr F. A. Uniformity of karyotypes in the Camelidae. *Cytogenet. Genome Res.* **7**, 8–15 (1968).
- Elsik, C. G. *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528 (2009).
- Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
- Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
- Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572–D580 (2006).
- Honey, J. G., Harrison, J. A., Prothero, D. R. & Stevens, M. S. in *Evolution of Tertiary Mammals of North America: Volume 1, Terrestrial Carnivores, Ungulates, and Ungulate like Mammals* 439–461 (Cambridge University Press, 1998).
- Alroy, J. *Fossilworks*. <<http://fossilworks.org>> (2013).
- Kosiol, C. *et al.* Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**, e1000144 (2008).
- Li, R. *et al.* SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19**, 1124–1132 (2009).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Gradstein, F. M., Ogg, J. G., Schmitz, M. D. & Ogg, G. M. *The Geologic Time Scale 2012 2-Volume Set* 1st edn (Elsevier, 2012).
- Clark, P. U. *et al.* The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO<sub>2</sub>. *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
- Head, M. J., Pillans, B. & Farquhar, S. A. The Early-Middle Pleistocene transition: characterization and proposed guide for the defining boundary. *Episodes* **31**, 255–259 (2008).
- Rook, L. & Martínez-Navarro, B. Villafranchian: the long story of a Plio-Pleistocene European large mammal biochronologic unit. *Quat. Intl* **219**, 134–144 (2010).
- Vislobokova, I. A. Main stages in evolution of Artiodactyla communities from the Pliocene-Early Middle Pleistocene of northern Eurasia: Part 2. *Paleontol. J.* **42**, 414–424 (2008).
- Mellars, P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc. Natl Acad. Sci. USA* **103**, 9381–9386 (2006).
- Marshall, L. G., Webb, S. D., Sepkoski, Jr J. J. & Raup, D. M. Mammalian evolution and the great American interchange. *Science* **215**, 1351–1357 (1982).
- Rutter, N., Coronato, A., Helms, K., Rabassa, J. & Zárata, M. *Glaciations in North and South America from the Miocene to the Last Glacial Maximum: Comparisons, Linkages and Uncertainties* (Springer, 2012).
- De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
- Arriza, J. L. *et al.* Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science* **237**, 268–275 (1987).
- Sun, X. J. *et al.* Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* **352**, 73–77 (1991).
- Wang, J. *et al.* SGK integrates insulin and mineralocorticoid regulation of epithelial sodium transport. *Am. J. Physiol. Renal Physiol.* **280**, F303–F313 (2001).
- Huang, Y. *et al.* Absence of aquaporin-4 water channels from kidneys of the desert rodent *Dipodomys merriami merriami*. *Am. J. Physiol. Renal Physiol.* **280**, F794–F802 (2001).
- Mobasher, A. *et al.* Distribution of the AQP4 water channel in normal human tissues: protein and tissue microarrays reveal expression in several new anatomical locations, including the prostate gland and seminal vesicles. *Channels* **1**, 29–38 (2007).
- Cheung, C. Y. & Ko, B. C. NFAT5 in cellular adaptation to hypertonic stress—regulations and functional significance. *J. Mol. Signal.* **8**, 5 (2013).
- Burg, M. B., Ferraris, J. D. & Dmitrieva, N. I. Cellular response to hyperosmotic stresses. *Physiol. Rev.* **87**, 1441–1474 (2007).
- Woo, S. K., Dahl, S. C., Handler, J. S. & Kwon, H. M. Bidirectional regulation of tonicity-responsive enhancer binding protein in response to changes in tonicity. *Am. J. Physiol. Renal Physiol.* **278**, F1006–F1012 (2000).
- Bartolo, R. C. & Donald, J. A. The effect of water deprivation on the tonicity responsive enhancer binding protein (TonEBP) and TonEBP-regulated genes in the kidney of the Spinifex hopping mouse, *Notomys alexis*. *J. Exp. Biol.* **211**, 852–859 (2008).
- Yancey, P. H. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* **208**, 2819–2830 (2005).
- Gallazzini, M. & Burg, M. B. What's new about osmotic regulation of glycerophosphocholine. *Physiology (Bethesda)* **24**, 245–249 (2009).
- Barnes, K. *et al.* Activation of GLUT1 by metabolic and osmotic stress: potential involvement of AMP-activated protein kinase (AMPK). *J. Cell Sci.* **115**, 2433–2442 (2002).
- Elmahdi, B., Sallmann, H. P., Fuhrmann, H., von Engelhardt, W. & Kaske, M. Comparative aspects of glucose tolerance in camels, sheep, and ponies. *Comp. Biochem. Physiol. A Physiol.* **118**, 147–151 (1997).
- Kaske, M., Elmahdi, B., von Engelhardt, W. & Sallmann, H. P. Insulin responsiveness of sheep, ponies, miniature pigs and camels: results of hyperinsulinemic clamps using porcine insulin. *J. Comp. Physiol. B* **171**, 549–556 (2001).
- Kim, G. *et al.* SREBP-1c regulates glucose-stimulated hepatic clusterin expression. *Biochem. Biophys. Res. Commun.* **408**, 720–725 (2011).
- Trougakos, I. P. *et al.* Serum levels of the senescence biomarker clusterin/apolipoprotein J increase significantly in diabetes type II and during development of coronary heart disease or at myocardial infarction. *Exp. Gerontol.* **37**, 1175–1187 (2002).
- Dieterle, F. *et al.* Urinary clusterin, cystatin C, beta2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat. Biotechnol.* **28**, 463–469 (2010).
- Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
- Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
- Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD dissertation (The Pennsylvania State University, 2007).
- Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
- Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
- Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-3.0*. <<http://www.repeatmasker.org>> (1996–2010).
- Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
- Smit, A., Hubley, R. & Green, P. *RepeatModeler Open-1.0* (2008–2010).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
- Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
- Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
- Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572–D580 (2006).
- Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
- Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226 (2006).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).



62. Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102**, 10557–10562 (2005).
63. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
64. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
65. Chen, S. *et al.* De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS ONE* **5**, e15633 (2010).

### Acknowledgements

We thank Dr Huijue Jia, Ph.D., Laurie Goodman, Dr David Wragg (School of Life Sciences, University of Nottingham) and Professor Olivier Hanotte (School of Life Sciences, University of Nottingham) for their assistance with editing and revising this manuscript. In addition, we thank Mr Fei Huang, Dr Yinfeng Yang, Dr Wenlong Wang, Dr Duger, Associate Professor Xiumei Wang, Professor MUSAAD A. Al-Dubaib and Dr Abdullah Alhawas from Qassim University for their assistance with sample collection, animal husbandry and data analysis. We thank Daniel Allain (Centre de Recherche de Toulouse, France) for providing photos of llama, guanaco and vicuña. We thank Tao Han (College of Material Science and Art Design, Inner Mongolia Agricultural University) for assisting us in production of featured image and thumbnail image. This work was supported by KACST.

### Author contributions

The camelid genome project was initiated by IMAU and KACST. Research collaboration and co-analysis were performed by IMAU, BGI, KACST and other participating authors'

institutes within the camelid genome consortium. H.Z., J.W. and A.A. supervised the project. H.Z. coordinated the project, designed the protocols and experiments, and edited and revised the manuscript. H.Z., M.A. and M.A. provided samples. Y.Y. provided advice and relevant information. J.C. conducted experiments. L.Z., Y.L., S.W., W.Z., L.B., C.L., F.M., L.L., X.L. and K.W. participated in the experiments. Y.P., Y.Z., D.Z. and Z.Y. provided advice and participated in the experiments. Z.X., L.Y. and J.Z. assembled the three genomes. Q.Y. and L.Y. annotated the genomes. H.W., X.G., M.A.F., S.P., H.Z., S.L., X.Y., S.W., E.L., S.Z., J.W., M.M.M., B.A.S. and A.A. analysed evolution, demographic history and adaptation. H.W., X.G., M.A.F., S.P., H.Z., J.L., S.Z. and J.W. analysed the transcriptomes. H.W., X.G., S.P., H.Z., J.W., H.Y. and A.A.S. wrote, edited and revised the manuscript.

### Additional information

**Accession codes:** The Camelid Whole Genome Shotgun Project has been deposited in DDBJ/EMBL/GenBank under accession codes JARL000000000 (Bactrian camel), JDVD000000000 (dromedary) and JEMW000000000 (alpaca). Raw DNA sequencing reads have been submitted to the NCBI Sequence Read Archive database under accession codes SRA179184 (Bactrian camel), SRA179175 (dromedary) and SRA179183 (alpaca).

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Wu, H. *et al.* Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* **5**:5188 doi: 10.1038/ncomms6188 (2014).

## Erratum: Camelid genomes reveal evolution and adaptation to desert environments

Huiguang Wu, Xuanmin Guang, Mohamed B. Al-Fageeh, Junwei Cao, Shengkai Pan, Huanmin Zhou, Li Zhang, Mohammed H. Abutarboush, Yanping Xing, Zhiyuan Xie, Ali S. Alsharqeti, Yanru Zhang, Qiulin Yao, Badr M. Al-Shomrani, Dong Zhang, Jiang Li, Manee M. Manee, Zili Yang, Linfeng Yang, Yiyi Liu, Jilin Zhang, Musaad A. Altammami, Shenyuan Wang, Lili Yu, Wenbin Zhang, Sanyang Liu, La Ba, Chunxia Liu, Xukui Yang, Fanhua Meng, Shaowei Wang, Lu Li, Erli Li, Xueqiong Li, Kaifeng Wu, Shu Zhang, Junyi Wang, Ye Yin, Huanming Yang, Abdulaziz M. Al-Swailem & Jun Wang

*Nature Communications* 5:5188; doi: 10.1038/ncomms6188 (2014); Published 21 Oct 2014; Updated 28 Jan 2015

In this Article Junyi Wang is incorrectly listed as one of several equally contributing authors. The correct list of equally contributing authors is: Huiguang Wu, Xuanmin Guang, Mohamed B. Al-Fageeh, Junwei Cao, Shengkai Pan and Huanmin Zhou.