

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/99329/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Pan, Shengkai, Zhang, Tongzuo, Rong, Zhengqin, Hu, Li, Gu, Zhongru, Wu, Qi, Dong, Shanshan, Liu, Qiong, Lin, Zhenzhen, Deutchova, Lucia, Li, Xinhai, Dixon, Andrew, Bruford, Michael W. ORCID: <https://orcid.org/0000-0001-6357-6080> and Zhan, Xiangjiang 2017. Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai-Tibetan Plateau in a predatory bird. *Molecular Ecology* 26 (11) , pp. 2993-3010. 10.1111/mec.14090 file

Publishers page: <http://dx.doi.org/10.1111/mec.14090>  
<<http://dx.doi.org/10.1111/mec.14090>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.


This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai–Tibetan Plateau in a predatory bird

SHENGKAI PAN,\*†‡<sup>1</sup> TONGZUO ZHANG,§<sup>1</sup> ZHENGQIN RONG,¶ LI HU,\*‡\*\* ZHONGRU GU,\*‡ QI WU,\* SHANSHAN DONG,†† QIONG LIU,\*‡‡ ZHENZHEN LIN,\* LUCIA DEUTSCHOVA,§§ XINHAI LI,\* ANDREW DIXON,†¶¶\*\*\* MICHAEL W. BRUFORD††††<sup>2</sup> and XIANGJIANG ZHAN\*†<sup>2</sup> 

\*Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beichen West Road, Beijing 100101, China, †Institute of Zoology Joint Laboratory for Biocomplexity Research, Cardiff University, Beichen West Road, Beijing 100101, China, ‡University of Chinese Academy of Sciences, Yuquan Road, Beijing 100049, China, §Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China, ¶Beijing HonorTech Co. Ltd., Beijing 102200, China, \*\*BGI-Shenzhen, Shenzhen 518083, China, ††Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China, ‡‡State Key Laboratory of Earth Surface Processes and Resource Ecology & MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China, §§Raptor Protection of Slovakia, Kuklovska 5, SK-841 04, Bratislava 4, Slovakia, ¶¶International Wildlife Consultants Ltd., PO Box 19, Carmarthen SA33 5YL, UK, \*\*\*Environment Agency-Abu Dhabi, PO Box 45553, Al Mamoura Building (A), Muroor Road, Abu Dhabi, United Arab Emirates, †††Organisms and Environment Division, Cardiff School of Bioscience, Cardiff University, Cardiff CF10 3AX, UK

## Abstract

Low oxygen and temperature pose key physiological challenges for endotherms living on the Qinghai–Tibetan Plateau (QTP). Molecular adaptations to high-altitude living have been detected in the genomes of Tibetans, their domesticated animals and a few wild species, but the contribution of transcriptional variation to altitudinal adaptation remains to be determined. Here we studied a top QTP predator, the saker falcon, and analysed how the transcriptome has become modified to cope with the stresses of hypoxia and hypothermia. Using a hierarchical design to study saker populations inhabiting grassland, steppe/desert and highland across Eurasia, we found that the QTP population is already distinct despite having colonized the Plateau <2000 years ago. Selection signals are limited at the cDNA level, but of only seventeen genes identified, three function in hypoxia and four in immune response. Our results show a significant role for RNA transcription: 50% of upregulated transcription factors were related to hypoxia responses, differentiated modules were significantly enriched for oxygen transport, and importantly, divergent *EPAS1* functional variants with a refined co-expression network were identified. Conservative gene expression and relaxed immune gene variation may further reflect adaptation to hypothermia. Our results exemplify synergistic responses between DNA polymorphism and RNA expression diversity in coping with common stresses, underpinning the successful rapid colonization of a top predator onto the QTP. Importantly, molecular mechanisms underpinning

Correspondence: Xiangjiang Zhan, Fax: +86 (0)1064807099;

E-mail: zhanxj@ioz.ac.cn

<sup>1</sup>These two authors contributed equally to the paper.

<sup>2</sup>These two authors are joint senior authors.

**highland adaptation involve relatively few genes, but are nonetheless more complex than previously thought and involve fine-tuned transcriptional responses and genomic adaptation.**

*Keywords:* adaptation, colonization, population transcriptome, Qinghai–Tibetan Plateau, RNA sequencing, saker falcon

*Received 5 August 2016; revision received 10 February 2017; accepted 28 February 2017*

## Introduction

Known as the ‘roof of the world’ and the ‘third pole’, the Qinghai–Tibetan Plateau covers a vast area of c. 2.5 million km<sup>2</sup> featuring extreme altitude (>4000 m on average). These highlands, characterized by low oxygen levels and temperatures, impose unique biological constraints on species that occupy the landscape. Unusual morphological, physiological and behavioural adaptations have been documented for numerous taxa, but understanding the genetic basis of these adaptations and especially their evolutionary dynamics has proved challenging.

Over the past 7 years, genomics has shed new light on the issue of hypoxia adaptation. The hypoxia-inducible factor pathway, including the *EPAS1* gene, has been identified as a major candidate for altitudinal adaptation in humans and other QTP mammals (e.g. Yi *et al.* 2010), and it is thought that they may mechanistically contribute to the constrained erythropoietic response in Tibetans (Cheviron & Brumfield 2011). The same gene has also been suggested to contribute to QTP adaptation in some mammals (e.g. wolves *Canis lupus chanco*, Zhang *et al.* 2014; Tibetan pigs *Sus scrofa*, Ai *et al.* 2014; Tibetan mastiffs *Canis familiaris*, Wang *et al.* 2014a,b; Fig. S1, Supporting information), but not in birds (ground tits *Parus humilis*, Qu *et al.* 2013; Tibetan chickens *Gallus gallus*, Wang *et al.* 2015), reptiles (toad-headed sand lizards *Phrynocephalus erythrurus*, Yang *et al.* 2015) or other mammals (Tibetan antelopes *Pantholops hodgsonii*, Ge *et al.* 2013; yaks *Bos grunniens*, Qiu *et al.* 2012). Recently, several interspecific comparative genomic studies have identified candidate genes involved in temperature adaptation (e.g. hypometabolism) in QTP specialists such as sand lizards (Yang *et al.* 2015) and ground tits (Qu *et al.* 2013), and suggested that strong selection on lipid metabolism may have developed as an optimal energy conservation strategy to cope with the challenge for chronic cold stress. In addition, previous physiological studies have indicated that long-term exposure to hypoxia, possibly together with strong ultraviolet radiation, suppresses immune response in humans living in high-altitude environments (Facco *et al.* 2005; Sica *et al.* 2011). This phenomenon, however, has rarely been studied in genomic

adaptation to the QTP except that a clear gene family contraction was observed in the immune-related genes in the ground tit genome, including the major histocompatibility complex (MHC) protein complex (Qu *et al.* 2013). Nonetheless, these studies have nearly all focused on genomic sequence changes, and until now the contribution of gene expression to QTP adaptation is largely unknown.

As top predators, it is well appreciated that raptors are key regulators in the plateau ecosystem (Sekercioglu 2006), but to date, little is known about the underlying basis for their adaptation to high-altitude life on the QTP. Moreover, unlike endemic and lower trophic level species, which often show distinct adaptive traits between the QTP and their lowland counterparts (Ke & Lu 2009), most raptor traits appear similar to those in the lowlands of Eurasia, and interestingly, in contrast to other avian groups, there are no QTP endemic birds of prey. Raptors therefore provide an interesting example for comparison among populations of the same species that potentially overlap or exchange migrants due to their high vagility, yet where resident populations must have also adapted to the QTP environment (Favre *et al.* 2015). The saker falcon (*Falco cherrug*) provides an ideal model to address this question. The saker has a vast breeding range in Eurasia and is a key predator, inhabiting variable landscapes including agricultural land, steppe, desert and QTP grasslands. On the QTP, sakers may perform an important ecosystem service by regulating its main prey, the plateau pika (*Ochotona curzoniae*), which is regarded as an agricultural pest that can reach high population densities in high alpine grazing lands (Lai & Smith 2003; Dixon *et al.* 2015). Fossil data suggest that sakers evolved very recently, diverging from their hierofalcon ancestor only around 30 thousand years (kyr) ago (Nittinger *et al.* 2005), making them one of the youngest raptor species described. Their current continental range, therefore, implies rapid colonization and local adaptation, but the process by which this has occurred has yet to be determined.

Here we report a population-level transcriptome study of QTP, steppe and lowland saker populations using NGS RNA sequencing to elucidate QTP adaptation and population history. We first generated ~3.6-Gb sequence data for each of 30 saker blood samples

(Table S1, Supporting information) and analysed cDNA single nucleotide polymorphisms (SNPs), which were used to examine population genetic structure, to reconstruct demographic history and to explore putative selection processes related to QTP adaptation. In order to assess the relative importance of adaptation ascribed to genomic or regulatory differences between the populations, we then quantitatively assessed expression levels for each gene within and among locations. We hypothesized that QTP sakers (i.e. Qinghai QH) would show both genomic adaptations and gene expression responses to extreme plateau environments in comparison with other eastern populations in Mongolia (MN) and Kazakhstan (KZ). We predicted that those eastern population samples would show common, regional signatures of selection and gene expression compared with the western lowland populations of Moldova (MD) and Slovakia (SK). We further predicted there would be independence and convergence in DNA adaptations and RNA expression variations to high-altitude life in separate highland populations (i.e. QH population vs. upland populations comprising QH, MN and KZ).

## Materials and methods

### *Sampling and RNA extraction*

Thirty blood samples were collected using a hierarchical geographic and altitudinal design from saker nestlings across Eurasia (Fig. 1a) with 10 individuals from the QTP (QH, highland plateau 4000 m+ altitude), four from Mongolia (steppe 1400 m+), five from eastern Kazakhstan (steppe 400 m+), totalling 19 individuals from eastern/central Eurasia; alongside four collected from Moldova and seven from Slovakia totalling eleven samples from lowland (75–150 m) western Eurasia (Table 1). Except for Moldova where the samples came from two nests, we sampled only one saker from each nest. To control the effect of age on gene expression, we only sampled middle to late stage nestlings aged 3–5 weeks old, prior to fledging. Approximately 0.5 mL fresh blood was taken from each saker falcon and immediately put in QIAGEN RNAProtect Animal Blood Tubes (100 µL each) and stored at –20 °C.

RNA was extracted using the RNeasy Protect Animal Blood Kit (Qiagen). The quality and quantity of RNA extracts were measured using Agilent 2100 (Agilent). We obtained more than 10 µg RNA from all but one samples (5.2 µg for QH13-2). All RNA extracts were then used for mRNA purification, cDNA synthesis and library construction as documented previously (Zhan *et al.* 2013). Briefly, mRNA was purified using Dynal Oligo (dT) beads (Invitrogen) and then fragmented into ~200-nt size fragments using RNA Fragmentation

Reagents (Ambion). cDNA was synthesized using random primers (Invitrogen), Superscript II (Invitrogen), RNase H and DNA polymerase I, followed by end repair with Klenow polymerase, T4 DNA polymerase and T4 polynucleotide kinase (for blunt ends of DNA fragments; Invitrogen). A single adenosine moiety was added to the cDNA using Klenow exonuclease and dATP. Illumina adapters (containing primer sites for sequencing and flow cell surface annealing) were ligated onto the cDNA repaired ends, and electrophoresis was used to separate fragments from unligated adapters by selecting cDNA fragments between 180 and 220 bp in size. cDNA libraries were amplified by 15 cycles of PCR with Phusion polymerase.

### *Data generation*

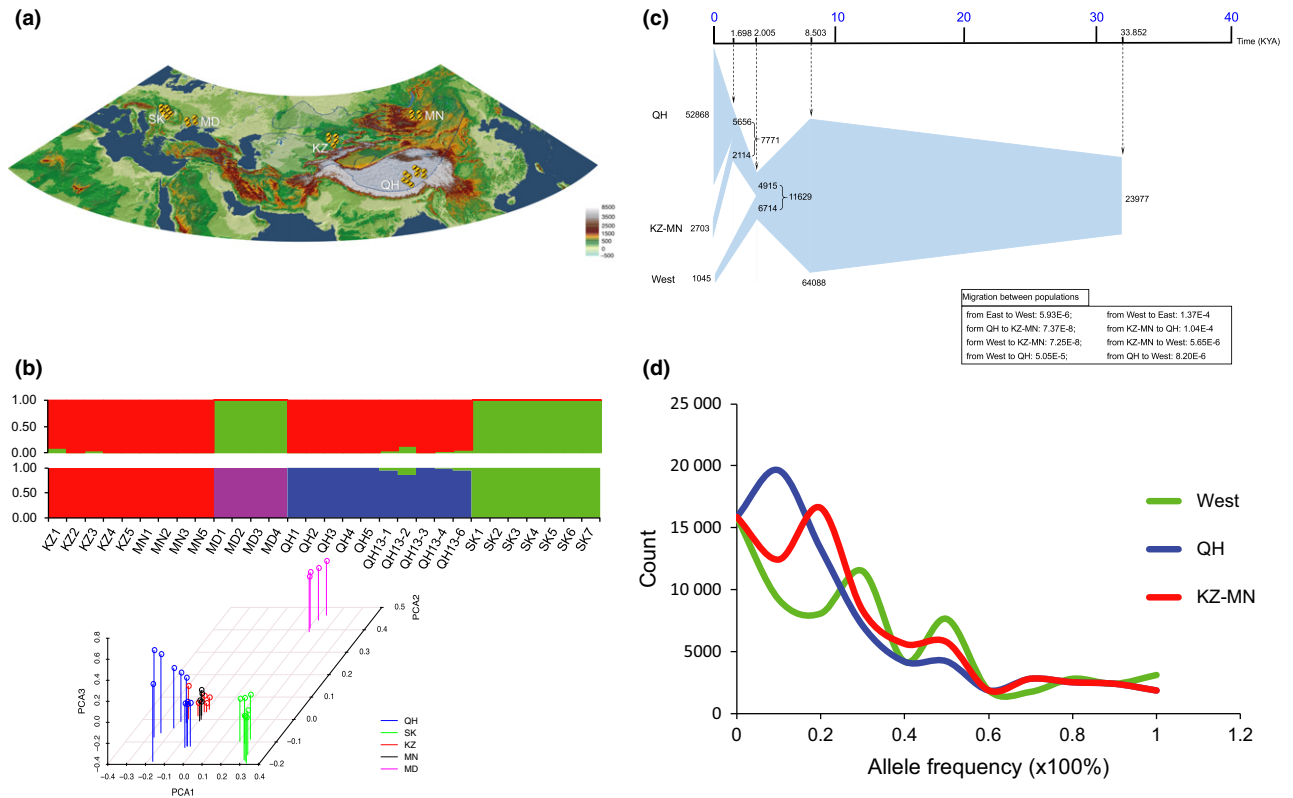
The cDNA libraries were sequenced using a HiSeq2000 (Illumina) platform at BGI-Shenzhen. Raw data were filtered by removing reads with adaptors and reads featuring unknown nucleotides more frequent than 10% and also those with low quality (where PHRED values were <10 for more than 50% of the bases). With this method, ~3.6-Gb clean data were generated for each individual (Table S1, Supporting information). The gene coverage of each blood transcriptome was assessed by mapping the RNA-seq reads to the saker reference genome (Zhan *et al.* 2013) using SOAP2.21 (Li *et al.* 2009a) with *-m* 0, *-x* 10000, *-s* 40, *-l* 32, *-v* 3, *-r* 2.

### *SNP detection*

SNPs were identified from these mapped reads as follows. First, SOAPSNP v 1.04 (Li *et al.* 2009b) was used to calculate the likelihood of genotypes of each individual at each site. Second, likelihood files of all saker individuals were integrated into one file. Specifically, each site was estimated by maximum likelihood, followed by filtering using the parameters: quality ≥20, copy number ≤1.5, sequencing depth (-MinDepth 30, -MaxDepth 1 000 000) and rank-sum test (*P* < 0.05). Third, only those polymorphic SNP sites that were covered by reads in all individuals were retained for the final analysis. Alleles were assigned to each individual depending on genotypes of the final SNPs and the likelihood values of each individual.

### *Reconstructing population structure*

Genetic structure was inferred using the program FRAPPE (Tang *et al.* 2005), implementing an expectation-maximization (EM) algorithm. We predefined the number of genetic clusters *K* as from 2 to 5. The maximum iteration of EM was set to 10 000. To examine the potential



**Fig. 1** Population structure and demographic history of saker falcons across Eurasia. (a) 30 sakers sampled from the five main geographic populations at different altitudes. QH: Qinghai-Tibetan Plateau (4000 m +), MN: Mongolia (~1400 m), KZ: Kazakhstan (500 m +), MD: Moldova (75 m), SK: Slovakia (100–150 m). (b) The *frappe* plots ( $K = 2$  and 4) and PCA results based on 377 530 SNPs identified from studied sakers. (c)  $\partial a\partial i$  results ( $\ln L = -9843.695$ ) showing the demographic history of saker falcons. Inset: migration rate from West to East is two orders of magnitude higher than that from East to West. And the migration from KZ-MN to QH is four orders of magnitudes higher than that from QH to KZ-MN. (d) SNP allele frequencies in each of three genetic clusters. The West population comprises MD and SK sakers.

effect of linkage disequilibrium (LD), we used *Haploview* (Barrett *et al.* 2005) to perform LD analysis based on the SNP data set obtained. LD decay was determined by squared correlations of allele frequencies ( $r^2$ ) against distances between polymorphic sites. We filtered the SNPs with  $r^2 > 0.2$  and reran *FRAPPE* using the remaining SNPs. In addition, we conducted principal components analysis (PCA) on the autosomal biallelic SNPs using *EIGENSOFT3.0* (Patterson *et al.* 2006). Eigenvectors from the covariance matrix were generated with the R function *EIGEN*, and significance levels were determined using a Tracy-Widom test.

### Simulating demographic history

To explore saker colonization history, we used two approaches that analyse the site frequency spectrum of SNPs called from the RNA-seq data:  $\partial a\partial i$  (Gutenkunst *et al.* 2009) and *FASTSIMCOAL2* (Excoffier *et al.* 2013). The two methods infer the demographic history using different statistics. Briefly,  $\partial a\partial i$  computes a composite

Poisson likelihood where the expected SFS is obtained by a diffusion approximation, while *FASTSIMCOAL2* is based on a composite multinomial likelihood where the expected SFS is obtained using coalescent simulations. We only considered those SNPs from intergenic and intronic regions of autosomal sequences ( $N = 57\ 146$ ) to better ensure their neutrality.

For the  $\partial a\partial i$  (Gutenkunst *et al.* 2009) simulations, the peregrine falcon genome sequence (Zhan *et al.* 2013) was used to define ancestral alleles, and a statistical procedure was performed to correct ascertainment bias of the ancestral state (Hwang & Green 2004). For model constructions, to determine whether the first population divergence event resulted from the ancestral population contraction or expansion, we simulated the two scenarios with the same data set as in Fig. S2 (Supporting information), and chose the population contraction one because of its larger  $\ln L$  value (see Results). The ancestral effective population size ( $N_a$ ) was estimated by the Theta value calculated with the mutation rate ( $1.65 \text{ E-}9$ ) previously published (Zhan *et al.* 2013). The population

**Table 1** Saker falcons studied in the present study

| Individual | Source         | Altitude | Collection year |
|------------|----------------|----------|-----------------|
| QH1        | Qinghai, China | 4099     | 2011            |
| QH2        | Qinghai, China | 4200     | 2011            |
| QH3        | Qinghai, China | 4152     | 2011            |
| QH4        | Qinghai, China | 4122     | 2011            |
| QH5        | Qinghai, China | 4152     | 2011            |
| QH13-1     | Qinghai, China | 4140     | 2013            |
| QH13-2     | Qinghai, China | 4087     | 2013            |
| QH13-3     | Qinghai, China | 4483     | 2013            |
| QH13-4     | Qinghai, China | 4486     | 2013            |
| QH13-6     | Qinghai, China | 4311     | 2013            |
| MN1        | Mongolia       | 1350     | 2011            |
| MN2        | Mongolia       | 1350     | 2011            |
| MN3        | Mongolia       | 1350     | 2011            |
| MN5        | Mongolia       | 1350     | 2011            |
| KZ1        | Kazakhstan     | 500–560  | 2011            |
| KZ2        | Kazakhstan     | 500–560  | 2011            |
| KZ3        | Kazakhstan     | 500–560  | 2011            |
| KZ4        | Kazakhstan     | 500–560  | 2011            |
| KZ5        | Kazakhstan     | 500–560  | 2011            |
| MD1        | Moldova        | 75       | 2012            |
| MD2        | Moldova        | 75       | 2012            |
| MD3        | Moldova        | 75       | 2012            |
| MD4        | Moldova        | 75       | 2012            |
| SK1        | Slovakia       | 128      | 2013            |
| SK2        | Slovakia       | 128      | 2013            |
| SK3        | Slovakia       | 154      | 2013            |
| SK4        | Slovakia       | 131      | 2013            |
| SK5        | Slovakia       | 139      | 2013            |
| SK6        | Slovakia       | 126      | 2013            |
| SK7        | Slovakia       | 117      | 2013            |

Altitudes of QH and SK sakers are GPS measured and the remaining are derived from the sampling areas.

size and chronological divergence time were derived from parameters scaled by Na. Finally, nonparametric bootstrapping was performed 50 times to determine the variance of each parameter using the 50 new data sets with equal numbers of loci sampled with replacement from the original data set ( $N = 57\,146$ ).

For FASTSIMCOAL2 (Excoffier *et al.* 2013), we used the population genetic structure detected (see Results), and simulated three different scenarios using the program with the same parameters (-n 100 000 -N 100 000 -M 0.001 -l 100 -L 100 -q -multiSFS -C 10 -c 8): 1) West sakers (MD+SK) diverged first, followed by QH and KM; 2) QH diverged first, followed by West and KM; and 3) KZ-MN diverged first, followed by West and QH using 100 000 replicates. The results were then processed with arlequin or arlsumstat (Excoffier & Lischer 2010), and the model (West sakers diverged first) with the largest lnL value was retained for final parameter estimation. Additionally, 1000 bootstrap replicates were conducted to determine the variance of each parameter for 30 000

loci randomly sampled from the original data set ( $N = 57\,146$ ).

To determine which population of sakers was most likely to be ancestral, we measured the relative genetic distances of saker populations to their closest genome-enabled species, the peregrine falcon (*Falco peregrinus*). We analysed all saker SNPs shared with the peregrine, and constructed a phylogenetic tree using an allele sharing matrix. The sequenced peregrine reads (Zhan *et al.* 2013) were mapped to the saker reference genome (Zhan *et al.* 2013) using SOAP2.21 with four mismatches allowed at most in one read. Genotypes of the peregrine were then used as an out-group at corresponding positions. A neighbour-joining rooted tree was generated by TREEBEST (<http://treesof.t.sourceforge.net/treebest.shtml>) using the p-distance model. In addition, we calculated the SNP frequency for each saker population using the equation  $\mu = n/2N$  where  $\mu$  is the allele frequency of one site in a population,  $n$  is the number of certain allele and  $N$  is the number of individuals in the population. For the purpose of comparison, we also performed the same analysis with the human SNP data (50 individuals for Han Chinese and 40 for Tibetan) published by Yi *et al.* (2010).

#### Detecting positive selection signatures

Following our hierarchical design based on altitude, habitat and finally the population structure detected above, we defined two population pairs to analyse for selection signals: QH vs. KZ-MN and East vs. West. We then used an integrative method (François *et al.* 2016) to detect the selection signals using cDNA SNPs identified from the RNA-seq data, which integrated the results obtained from three selection tests: the FLK test based on comparing different patterns of allele frequencies among populations to the values expected under a scenario of neutral evolution (Bonhomme *et al.* 2010), latent factor mixed models (LFMMs) for which the environment is used as a fixed effect and latent factors used to infer environmental associations (Frichot & François 2015) and PCADAPT v3.03 analysis based on Bayesian factor model (Duforet-Frebourg *et al.* 2016). We conducted the FLK test using HAPFLK (version 1.3.0) (Fariello *et al.* 2013) with parameters (-nfit = 100). The significance values of each locus tested by FLK were then used to calculate the adjusted z-scores using the genomic inflation factor. Then LFMMs were performed as population differentiation tests to identify positively selected loci using the LEA package (Frichot & François 2015) with parameters  $K = 2$ , rep = 10. We coded 0 or 1 as the environmental variables for each of the populations compared. The adjusted LFMM z-scores were obtained using the same method as above. PCADAPT was conducted using R package (Luu *et al.* 2017) with

parameters  $K = 1$ ,  $\text{min.maf} = 0.05$ . The adjusted z-scores for Bayesian factor model were calculated. To summarize these results, we combined the three types of adjusted z-scores obtained above, and calculated the calibrated  $P$  values as suggested by Frichot & François (2015). The candidate loci were ultimately determined using Benjamini–Hochberg FDR (false discovery rate) control. The level of FDR was set to 0.01.

For the *EPAS1* and *MHC class II B* genes, we used the derived allele frequency (DAF) test (Sabeti *et al.* 2007) to localize the signal of selection to populations. We first

used the primers (Appendix S1, Supporting information) to amplify corresponding fragments containing selected SNPs (Table 2) in a gyrfalcon (*Falco rusticolus*) sample, which was kindly provided by Abu Dhabi Falcon Hospital. Then we inferred ancestral alleles using the gyrfalcon, because it is more closely related to the saker than the peregrine (Nittinger *et al.* 2005) and it should be more appropriate for the definition of ancestral alleles of the two targeted genes. In the end, populations with high frequencies of derived alleles were assumed to be under positive selection.

| Scaffold | Position | Gene                    | Expression difference | Gene region | Amino acid change |
|----------|----------|-------------------------|-----------------------|-------------|-------------------|
| 3981     | 3209888  | <i>EPAS1</i>            | *                     | Exon        | Val -> Ala        |
| 3881     | 41491    | <i>MHC class II B</i>   | *                     | Exon        | Ala -> Val        |
| 3881     | 41307    | <i>MHC class II B</i>   |                       | Exon        | Ser -> Gly        |
| 3881     | 41346    | <i>MHC class II B</i>   |                       | Exon        | Leu ->Val         |
| 1111     | 3902053  | <i>LATS1</i>            |                       | Exon        | Synonymous        |
| 1791     | 383246   | <i>SDHA</i>             |                       | Exon        | Synonymous        |
| 201      | 17779580 | <i>DHX30</i>            |                       | Exon        | Synonymous        |
| 3711     | 1220149  | <i>Hemoglobin alpha</i> |                       | Intron      |                   |
| 1611     | 1662466  | <i>ZFYVE16</i>          |                       | Intron      |                   |
| 1671     | 2042639  | <i>UBA52</i>            |                       | Intron      |                   |
| 2681     | 4524397  | <i>CCNT2</i>            |                       | Intron      |                   |
| 3451     | 4802289  | <i>TESK2</i>            |                       | Intron      |                   |
| 3711     | 1158505  | <i>MPG</i>              |                       | Intron      |                   |
| 3711     | 1161071  | <i>MPG</i>              |                       | Intron      |                   |
| 3791     | 1749906  | <i>Unknown</i>          |                       | Intron      |                   |
| 4521     | 1468029  | <i>CCPG1</i>            |                       | Intron      |                   |
| 4621     | 644434   | <i>LDB1</i>             |                       | Intron      |                   |
| 721      | 2073571  | <i>RPS12</i>            |                       | Intron      |                   |
| 6731     | 48611    | <i>CHC_T00008058001</i> |                       | UTR_3       |                   |
| 6731     | 48761    | <i>CHC_T00008058001</i> |                       | UTR_3       |                   |
| 1013     | 3861190  | <i>CD46</i>             |                       | UTR_5       |                   |
| 1013     | 3861191  | <i>CD46</i>             |                       | UTR_5       |                   |
| 1831     | 2165687  | –                       |                       | Intergenic  |                   |
| 1881     | 1048807  | –                       |                       | Intergenic  |                   |
| 1931     | 30346    | –                       |                       | Intergenic  |                   |
| 2391     | 957278   | –                       |                       | Intergenic  |                   |
| 2411     | 4703301  | –                       |                       | Intergenic  |                   |
| 2541     | 6687499  | –                       |                       | Intergenic  |                   |
| 2541     | 6684111  | –                       |                       | Intergenic  |                   |
| 2541     | 6685202  | –                       |                       | Intergenic  |                   |
| 2541     | 6681080  | –                       |                       | Intergenic  |                   |
| 2801     | 3030515  | –                       |                       | Intergenic  |                   |
| 372      | 2458091  | –                       |                       | Intergenic  |                   |
| 3881     | 38102    | –                       |                       | Intergenic  |                   |
| 3891     | 1368100  | –                       |                       | Intergenic  |                   |
| 4341     | 939335   | –                       |                       | Intergenic  |                   |
| 4811     | 887976   | –                       |                       | Intergenic  |                   |

**Table 2** SNP candidates under positive selection between QH and KZ-MN saker populations<sup>†</sup>

\* $q < 0.05$  in the tests of significant difference in gene expression (i.e. *EPAS1* and *MHC class II B*).

<sup>†</sup>Gene coordinates and annotations were derived from the reference saker genome (Zhan *et al.* 2013).

### Expression analysis

Cleaned sequence reads for each sample were aligned to the reference genome (Zhan *et al.* 2013), incorporating known gene annotations (using improved version HISAT2 v.2.0.1) with the parameters  $-N 0, -I 100, -X 600, -fr, -tmo$  (Kim *et al.* 2015). STRINGTIE v.1.2.0 (Pertea *et al.* 2015) was used with default parameters to reconstruct transcript annotations across the genome merged with reference gene models (Zhan *et al.* 2013). The expression of each transcript was quantified as transcripts per million (TPM) using RSEM (Li & Dewey 2011), preferred over RPKM and FPKM measures for quantification because it is independent of the mean expressed transcript length and is thus more comparable across samples and species. Using PCA, we first compared expression profiles for each saker within each of the three populations determined in this study (see Results) and found that the transcript expression distribution was independent of either chick age estimates, sex or sampling date (Fig. S3, Supporting information). We then compared the total number of expressed transcripts between QH and KZ-MN, and between West and East, using a bootstrap strategy ( $N = 100$ ). For each bootstrap, we randomly selected 100 M reads to count the number of expressed transcripts. Based on these results, we obtained a distribution of expressed transcript numbers for each population. A *t*-test was adopted to test the difference in the mean number of expressed transcripts for the two pairs of population comparisons.

We used four methods to identify differentially expressed transcripts (DETs) in the population comparisons (QH vs. KZ-MN, West vs. East): EDGER-ROBUST (Zhou *et al.* 2014) and DESEQ 2 (Anders & Huber 2010) based on exact tests, a nonparametric test in SAMSEQ (Li & Tibshirani 2013) and an empirical Bayesian analysis using LIMMA (Smyth 2004). A transcript was accepted as being a DET only when it was supported by more than two different underlying statistical models. We also used the R package MINOTAUR (Verity *et al.* 2017) to integrate the multivariate results from the expression difference analyses. Mahalanobis distance was calculated based on the *q*-values generated from EDGER-ROBUST, DESEQ 2, SAMSEQ and LIMMA analyses.

### Gene interaction and cluster analysis for DETs between saker populations

To examine the difference in gene expression of transcription factor (TF) between saker populations, we performed a TF analysis using Fisher's exact test implemented in the Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood City, CA, USA). In order to study differences in gene regulation in different

populations, we applied the chicken gene interaction relationship in the STRING database to reconstruct orthologous gene interaction networks in the saker and screened significantly co-expressed gene transcript–gene transcript pairs using the Fisher correlation coefficient with the cut-off of  $P < 0.01$  in each population. Bootstrapping procedure ( $N = 1000$ ) was used to estimate the correlation coefficient distribution in both populations. The probability density distribution was plotted by a two-dimensional Gaussian kernel function estimation programmed using the PYTHON 2.7.5 packages, SCIPY 0.15.1 and MATPLOTLIB 1.4.3.

Expressed genes were hierarchically clustered to represent the expression patterns using Ward's method with Euclidean distances and correlation coefficients as measurements of similarities for genes and for samples. Analyses were conducted in the PYTHON 2.7.5 packages, NUMPY 1.9.2, SCIPY 0.15.1 and MATPLOTLIB 1.4.3.

### Weighted gene co-expression network analysis and intramodular connectivity estimation

A gene co-expression network was used to depict difference in transcriptome complexity between saker populations. We applied weighted gene co-expression network analysis (WGCNA) (Zhang & Horvath 2005) to group transcripts with correlated expression levels into co-expression modules. Subfunctional modules are averaged linkage hierarchical clusters defined by the topological overlap dissimilarity (eqn 1), with the aim to identify submodules with potential biological functions. Once a dendrogram was obtained from the hierarchical clustering, we chose the dynamic hybrid branch cutting method (Langfelder *et al.* 2007) with 'minClusterSize' at 30 to attain a clustering solution. The identified modules correspond to branches of the dendrogram.

For the focal gene transcripts (i.e. *EPAS1* and *MHC Class II B*), we used intramodular connectivity (eqn 4) to describe their importance in co-expressed submodules for different populations. Jackknife method was used to estimating confidence intervals (95%) of intramodular connectivity. Equations:

$$\text{DistTOM}_{ij} = 1 - \text{TOM}_{ij} \quad (1)$$

where *i* and *j* are two different gene transcript nodes in a net, and TOM (topological overlap similarity) is described as:

$$\text{TOM}_{ij} = \frac{\sum_u a_{iu}a_{uj} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}} \quad (2)$$

where the power adjacency *a* for two transcripts is described as:



$$a_{ij} = |\text{cor}(x_i, x_j)|^\beta \quad (3)$$

In this study, we use the scale free topology criterion described by Zhang & Horvath (2005) to determine beta. The *cor* operator is a Spearman correlation coefficient between the expression levels of the two transcripts of gene *i* and *j*. The transcript connectivity (*k*) in the weighted network is described as the sum of connection strengths to another node:

$$k_i = \sum_j a_{ij} \quad (4)$$

#### Association analysis between MHC Class II B gene and environmental variables

Data from 19 bioclimatic variables were obtained from WORLDCLIM (Hijmans *et al.* 2005). The variables were derived from monthly temperature and rainfall values in order to generate more biologically meaningful variables and they were used to proxy annual trends, seasonality and extreme environmental factors (Table S2, Supporting information). In addition, we examined the effect of elevation. Before the association analysis, we used the random forest model (Breiman 2001) to rank each environmental variable based on its contribution to the gene expression of each individual (Fig. S4, Supporting information), and the top 10 variables were retained for the following analysis. We then used correlation analysis and multiple linear regression to establish the correlation between any two variables and the contribution of a few relatively independent variables (i.e. absolute correlation coefficients <0.4) to the gene expression differences among saker individuals.

## Results

#### RNA-seq and SNP calling

We have successfully constructed 30 saker cDNA libraries, and RNA-seq produced a total of 107.56 Gb of clean data (Table S1, Supporting information). SNP calling with strict filtering resulted in the identification of 377 530 SNPs (Table S3, Supporting information). Compared with the established gene models for the saker (Zhan *et al.* 2013), ~67% of genes were expressed in blood. Annotation showed that most SNPs were located in introns (161 555) or intergenic regions (177 641). We also found 1470 novel transcripts (data not shown).

#### Population structure

Based on the 377 530 SNP data set and using PCA and FRAPPE analysis (Fig. 1a, b; Fig. S5, Supporting information), we could infer three distinct population clusters:

QH, KZ-MN (Kazakhstan and Mongolia) and West (Slovakia and Moldova). For the PCA, the first eigenvector separated the western populations (MD, SK) from the east (QH, MN, KZ) ( $P = 2.44 \text{ E-}10$ ), whereas the second eigenvector separated MD from SK ( $P = 8.95 \text{ E-}12$ ). The third eigenvector distinguished QH from MN and KZ ( $P = 0.0001$ ): further eigenvectors detected no meaningful geographic signal (Fig. S6, Supporting information). The population structure inferred using the whole SNP data set was identical to that using the filtered data set (LD  $r^2 > 0.2$  excluded; Fig. S7, Supporting information), indicating little influence of LD in the FRAPPE analysis.

#### Inferred demographic history

$\partial a \partial i$  and FASTSIMCOAL2 produced similar estimates of saker demographic history for population dynamics, divergence time and migration (Fig. 1c; Fig. S8, Table S4, Supporting information). We report the  $\partial a \partial i$  results because it has been found to be the most stable across simulation runs (Excoffier *et al.* 2013). We also provide the FASTSIMCOAL2 results in Supporting information. Our simulations allowed the evolutionary history of saker falcons to be estimated back to 34 kyr, regardless of the method adopted (Fig. 1c; Fig. S8, Supporting information). For more than 20 000 years after its origin, the saker population can be inferred to have been expanding, reaching its peak around 8.5 kyr and beginning to decline thereafter. We simulated the scenarios to determine whether there was a population decline associating with the first population divergence, and selected the model taking account of population decline because of its larger  $\ln L$  value ( $-3299.02$  vs.  $-27027.4$ ; Fig. S2, Supporting information). West and East saker populations were inferred to have diverged first (~2 kyr), followed by another divergence between QH and MN+KZ in East population (~1.7 kyr). The dendrogram analysis based on an allele sharing matrix showed that the westernmost saker population, SK, shared most alleles with the out-group peregrine (Fig. S9, Supporting information).

Since diverging, our analysis allowed us to infer that West saker populations have been declining. In contrast, the eastern populations are inferred to have undergone an expansion (Fig. 1c), and the QH population was inferred to have undergone an almost 10-fold increase. Figure 1d illustrates that QH had the most low-frequency SNP alleles and West had the least. The patterns were the same between coding and intron regions (Fig. S10, Supporting information). However, we did not find this pattern when comparing the Han and Tibetan Chinese (Fig. S10, Supporting information). Migration analysis inferred the eastwards gene flow was 2–4 orders of magnitude higher than westwards

(Fig. 1c):  $1.04E-4$  from KZ-MN to QH but  $7.37E-8$  from QH to KZ-MN, and  $1.37E-4$  from West to East but  $5.93E-6$  from East to West.

#### Positively selected genes in high-altitude sakers

Between QH and KZ-MN, 37 SNPs were identified to be under directional selection from a total of 67 837 loci (Table 2). These SNPs were located to 17 genes and 15 intergenic regions in the saker genome with three functioning in hypoxia response (*EPAS1*, *Hemoglobin alpha* and *LDB1*) and four related to immunity (*MHC*, *DHX30*, *UBA52* and *CD46*).

The SNP identified as being a selection target in *EPAS1* was exonic (exon 4) and nonsynonymous (from *Val* to *Ala*). Except for one bird with a heterozygous *CT* genotype, all QH sakers were homozygous *CC* for the selected SNP in contrast with the KZ-MN sakers that were either *TT* or *CT* (Table S5, Supporting information). The *DAF* value was 0.95 and 0.28 in QH and KZ-MN, respectively. Three selected SNPs were found in the exons of *MHC class II B* gene, which also cause nonsynonymous substitutions (Table 2). As shown in Table S5 (Supporting information), the selected SNPs in KZ-MN were fixed at sites 41491 and 41346, and nearly fixed at Site 41307. The *DAF* values for the three SNPs in KZ-MN were 1, 1 and 0.94, respectively. Different from the above two genes, the positively selected site found in *Hemoglobin A* was intronic (Table 2).

In the comparison of West (SK+MD) and East (QH+KZ-MN) saker populations, we identified 267 positively selected cDNA SNPs involving 58 genes (Table S6, Supporting information). The genes under selection were found to cluster in two regions of the saker reference genome (Fig. S11, Supporting information). GO analysis further showed that the significantly enriched categories for these 58 selected genes were oxygen transport, the haemoglobin complex, oxygen and heme binding ( $P < 0.05$ , Fisher's exact test) (Table S7, Supporting information).

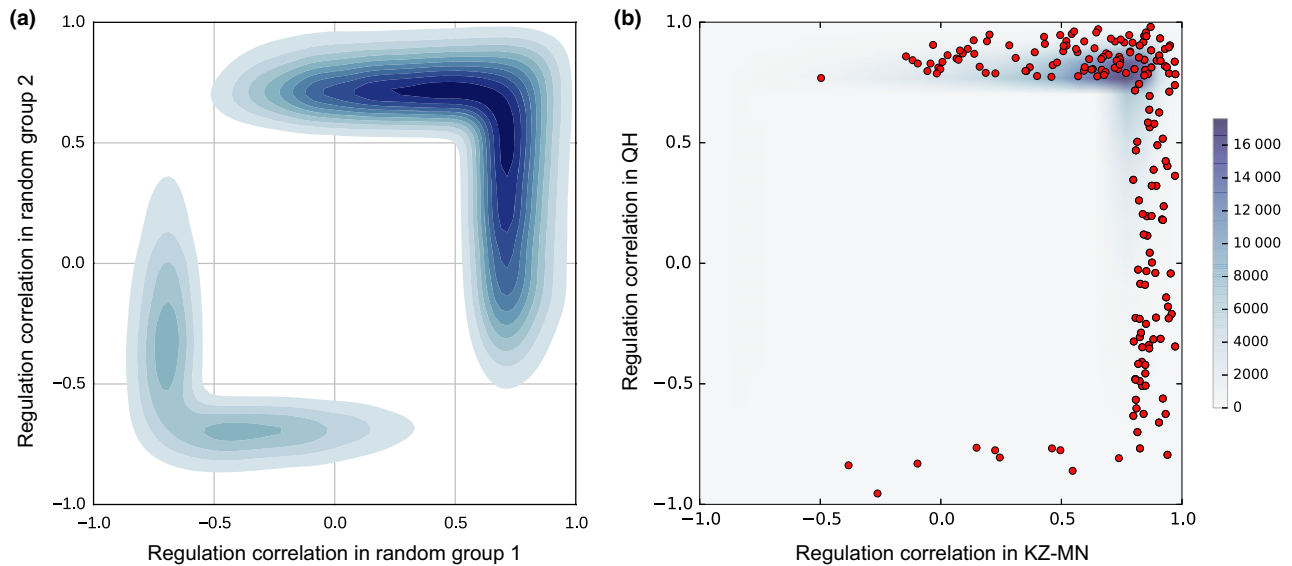
#### Transcriptomic expression differences

We found that fewer transcripts were expressed in the blood of sakers at higher altitude than those at low altitude [QH (16 825 transcripts of 10 778 genes) vs. KZ-MN (19 060 transcripts and 11 900 genes); East (16 608 transcripts and 10 690 genes) vs. West (17 868 transcripts and 11 424 genes)], and the comparisons were significantly different ( $P = 0.046$  for the former and  $P < 2.2 E-16$  for the latter). Moreover, fewer differentially expressed transcripts were found to be upregulated in upland and highland sakers (478/1139 DETs for QH, 305/625 DETs for East) than at lower altitude.

The two integration methods generated different estimates of the number of DETs. Using the criteria that a DET is recognized only if supported by at least two tests from *EDGE-ROBUST*, *DESEQ 2*, *SAMSEQ* and *LIMMA*, we could identify 1139 differentially expressed transcripts in the population comparison of QH and KZ-MN, involving 906 genes (Fig. S12, Supporting information). In contrast, the integrative distance approach implemented in *MINOTAUR* resulted in the identification of 2677 DETs. Among which, 1105 overlapped with the first DET data set, accounting for 98%. However, for the remaining outliers ( $N = 1572$ ), we found that 80% were not supported by any of the four tests, possibly suggesting the high false-negative rate associated with the second integration method. Therefore, we chose to use the data set of 1139 DETs for the following analyses.

Transcription factor analysis revealed significant enrichment for the targets of 19 TFs with 12 upregulated in the QH populations (Table S8, Supporting information). The WGCNA analysis grouped the transcripts and identified 12 and 32 co-expression modules in the QH and KZ-MN saker populations, respectively. Among the QH modules, the module that had the largest number of DETs ( $N = 98$ ) is significantly enriched for genes involved in oxygen transport (GO: 0015671).

We identified three more *EPAS1* transcripts based on the reannotation of RNA-seq data of 19 sakers from QH and KZ-MN which were different from those predicted from the saker reference genome (Zhan *et al.* 2013). The expression of one *EPAS1* transcript (Saker\_CCG011490.1) was found to be significantly higher in QH than in KZ-MN ( $q = 0.011$ , *DESEQ* test;  $q = 0.010$ , *EDGE-ROBUST* test;  $q = 0.004$ , *SAMSEQ* test). With random grouping, the regulation correlation of all the co-expressed gene transcripts demonstrated a distribution with two big clusters (i.e. top right and bottom left corners; Fig. 2a). When we grouped the transcripts into the two actual populations, QH and KZ-MN, the contour image showed a different distribution with a single significant cluster at the top right corner (Fig. 2b vs. Fig. 2a). In particular, the co-expression similarity of *EPAS1* was distinct between the two populations in that negative correlations accounted for half of the co-expressed transcript pairs in QH, but positive correlations were found in most *EPAS1* co-expressed pairs in KZ-MN (Fig. 2b). We measured the intramodular connectivity of each *EPAS1* transcript and found that among the three transcripts passing our quality control, the connectivity of two *EPAS1* transcripts in QH is significantly higher than KZ-MN (Fig. 3a). Overall, five *EPAS1* transcripts was found to co-express with the number of 148 transcripts from 90 genes. Each of these transcripts was found to regulate a mean of 37 transcripts for QH population and 20 for the KZ-MN



**Fig. 2** Regulation correlation coefficients ( $r$ ) of co-expressed gene transcripts in the saker population QH and KZ-MN. (a) The simulated distribution of  $r$  when the sakers are randomly divided into two populations. It demonstrates two large clusters. (b) The actual distribution of  $r$  values obtained from the RNA-seq data of two studied saker falcon populations. The contour image shows a single significant cluster at top right. The  $r$  values of *EPAS1* are highlighted (red dots) in that negative correlations accounted for half of the co-expressed transcript pairs in QH, but positive correlations were found in most *EPAS1* co-expressed pairs in KZ-MN. Blue bar denotes the count of significantly co-expressed transcript.

population cluster. Fig. 3b further shows that different *EPAS1* transcripts could have opposite transcriptional effects on the same transcript.

DET analysis of RNA profiles showed the *MHC class II B* transcript (i.e. Saker\_CCG011292.1) was significantly more highly expressed in KZ-MN than QH ( $q = 0.010$ , *SAMseq*;  $q = 0.030$ , *DESeq*) (Table 2). The co-expression network analysis showed that the intramodular connectivity of *MHC class II B* in KZ-MN was nearly twofold of that in QH (Fig. 3a). Of the twenty environmental variables related to temperature, precipitation and elevation (Table S3, Supporting information), *MHC* expression was best explained by mean annual temperature (31.9%,  $P = 0.018$ ; Table S9, Supporting information).

Between West and East saker populations, 87 transcripts were found to be significantly differently expressed (Fig. S13, Supporting information). GO annotations of these DETs demonstrated that those upregulated in the East population were significantly enriched for oxygen transport (GO: 0015671) and heme binding (GO: 0020037) (Table S10, Supporting information).

#### Differences in the responses to altitudinal stress

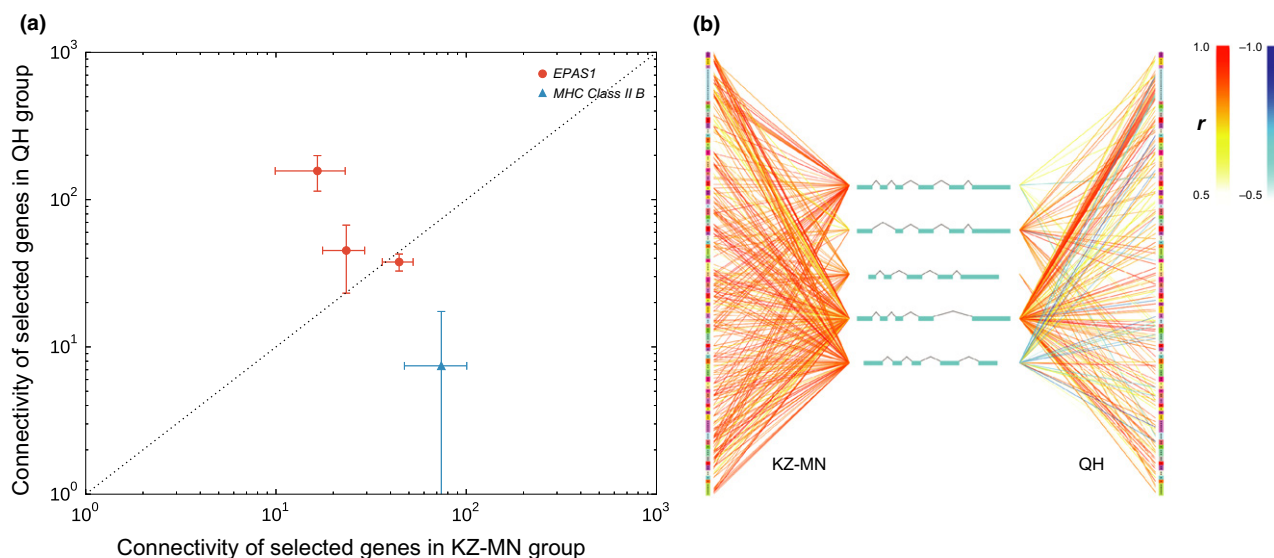
When comparing the two hierarchical levels of population comparison (QH vs. KZ-MN and East vs. West), we found that only one positively selected gene (*HBA*) (Table 2 vs. Table S6, Supporting information) and a

small proportion of DETs were shared between the comparisons (7.5% in QH vs. KZ-MN and 13.7% in East vs. West). We functionally enriched these DETs based on GO and KEGG analysis and found that few categories overlapped between the two comparisons. The shared ones were significantly enriched in Ribosome (GO: 0003735, KO: 03010), Translation (GO: 0006412), Protein transport (GO: 0015031) or GTP activities (GO: 0005525, 0003924).

## Discussion

### Rapid colonization of the Qinghai–Tibetan Plateau

Although saker falcons have a wide breeding range, spanning from central Europe to eastern Asia (Fig. 1a), mitochondrial DNA and microsatellites markers have not detected population genetic structure (Nittinger *et al.* 2005, 2007). However, we recently found genetic differentiation between the Qinghai–Tibetan Plateau and all other regions using 36 exon-located genomic SNPs (Zhan *et al.* 2015). Our present study, based on 377 530 SNPs, has detected clear population structure for saker falcons in Eurasia with three distinct population clusters being identified, comprising central Europe (Slovakia and Moldova), central Asia (Kazakhstan and Mongolia) and the Qinghai–Tibetan Plateau (Fig. 1a, 1b). It is noted that while the Moldova sakers were genetically distinct from Slovakia (Fig. 1), the small



**Fig 3** The intramodular connectivity of focal genes and the regulation pattern of the *EPAS1* transcripts. (a) The intramodular connectivity of *EPAS1* (Saker\_CCG011490) and *MHC Class II B* (Saker\_CCG011292) with 95% jackknife CI shown besides the dots. Noted that only three *EPAS1* transcripts passed our control of transcript expression (at least one bird has TPM  $\geq 2$ ) and intrapopulation missing values ( $\leq 60\%$ ). (b) Transcripts where expression levels were negatively correlated with *EPAS1* in QH ( $r < -0.5$ ) but not in KZ-MN ( $r > 0.5$ ). Schematic representation of the five transcripts in *EPAS1* is shown in the middle. The two vertical bars represent the 148 co-expressed transcripts of *EPAS1* with each colour block symbolizing one target transcript. The red and blue gradients denote positive and negative correlation coefficients, respectively.

sample size in Moldova (two nests) renders this result especially prone to sampling error, and we used West sakers (SK + MD) as a whole for most analyses in this study.

Using cDNA SNPs identified from RNA-seq data, we reconstructed the demographic history of the saker falcon (Fig. 1c; Fig. S8, Supporting information). The origin of the saker was dated back to 34 kyr. This estimate is somewhat earlier than the oldest saker fossils found so far (19 kyr) (Simmons & Nadel 1998) and is very close to the lower bound of the earliest hierofalcon fossil (Nittinger *et al.* 2005), but supports evidence for a very recent origin for the species and the notion of a recent rapid radiation of the saker and hierofalcon lineage (Nittinger *et al.* 2005). Since its origin, the saker population was inferred to have been expanding for more than 20 000 years, which agrees with previous PSMC-based estimates based on the reference genome (Zhan *et al.* 2013). During this time, it seems highly likely that western sakers expanded by dispersing eastwards because the westernmost population, SK, was inferred to be the oldest population (Fig. S9, Supporting information). Interestingly, this population expansion contrasts with the inferred demographic history of many European biota, which underwent a population contraction at this time during the last glaciation (Hewitt 2000), but the cold and arid conditions during this period coincided with the opening up of the steppe and

arid grassland ecosystems to which the saker is adapted (Ferguson-Lees & Christie 2001). The first population decline was inferred to occur around 6000 BC. During that time, a warmer geological period was experienced with higher temperatures across Eurasia (2–3 °C) (Kerwin *et al.* 1999) and subsequent loss of suitable habitat and the onset of agriculture (Werger *et al.* 2012) may have precipitated this population decline and triggered the split of West (MD+SK) and East sakers (KZ+MN+QH). After the split, the eastern populations are inferred to have undergone a moderate expansion (Fig. 1c). In contrast, West saker populations have been declining, in agreement both with recent population studies in Europe and with well-documented anthropogenic contraction of suitable habitat during this period (Dixon 2007).

Our study suggests a rapid colonization history for saker falcons onto the Qinghai–Tibetan Plateau. The almost unidirectional gene flow was inferred from the west into east (West to East:  $1.37\text{E-}4$ ; East to East:  $5.93\text{E-}6$ ) and from KZ-MN to QH (KZ-MN to QH:  $1.04\text{E-}4$ ; QH to KZ-MN:  $7.37\text{E-}8$ ), which coincides with saker migration routes observed using satellite telemetry (from Slovakia to eastern countries, Nemček *et al.* 2014; from Mongolia to QTP, Dixon *et al.* 2017). Most importantly, the separation time of QH from its neighbouring population (i.e. KZ-MN) was inferred to be the most recent among the studied saker populations, just  $\sim 1.7$

kyr or 260 generations ago (generation time = 6.6 year; Zhan *et al.* 2013). Since separation, the QH population was estimated to have undergone an almost 10-fold increase. These results, taken together, suggest a recent colonization and expansion of QTP sakers. Supporting evidence also comes from the comparison of the number of low-frequency SNP alleles: QH had the largest number, KZ-MN intermediate and West population the fewest (Fig. 1d), as would be expected for an expanding population on the Plateau and a contracting population in Europe.

#### *Low-oxygen adaptation in QTP sakers*

Given that saker falcons colonized the Qinghai–Tibetan Plateau (Fig. 1) in a short period of time, we were interested to understand how sakers have adapted to the low-oxygen environment of the plateau. Our selection analysis based on 68K SNPs between QH and KZ-MN identified 17 genes containing significantly selection signals (Table 2). Of these, three genes (i.e. *EPAS1*, *Hemoglobin*, *LDB1*) are associated with hypoxia responses found in previous studies (e.g. Brandt & Koury 2009; Yi *et al.* 2010). In contrast to the limited selection signals detected from the cDNA sequences, we uncovered a series of expression responses of QH sakers to hypoxia. Twelve transcription factors were upregulated in QH, and half of these were found to be related to hypoxia response or hematopoiesis (Table S8, Supporting information). The number of co-expression modules in QH ( $N = 12$ ) was different to that of the KZ-MN population ( $N = 32$ ); however, the largest module in QH was significantly enriched for genes involved in oxygen transport, suggesting that plateau sakers have enhanced their efficiency in utilizing oxygen.

Interestingly, this module also contains the *EPAS1* gene. *EPAS1* is a well-known candidate master gene for determining hypoxia responses at high altitude, with strong signatures of natural selection reported for Tibetan humans and grey wolves (Fig. S1, Supporting information). However, no functional variants of *EPAS1* have been identified to date, possibly because almost all selected SNPs have been found in noncoding regions, either within introns or downstream (Fig. S1, Supporting information), or because little research has been conducted into study of functional elements (RNA or protein). Intriguingly, two recent studies examining *EPAS1* mRNA expression levels found no significant difference in Han Chinese before and during mountain climbing (Chen *et al.* 2012), and lower *EPAS1* expression in Tibetan migrants compared with lowland British (Petousi *et al.* 2014). In contrast, our work demonstrates a refined role for *EPAS1* in the gene adaptation and

expression plasticity of saker falcons to the low-oxygen environment of the QTP.

We found evidence for positive selection on *EPAS1*, similar to other QTP animals (e.g. Tibetan pigs, wolves, mastiffs, humans, but Tibetan chicken and ground tits). The selection signature was only found between QH and KZ-MN, not between East (QH+KZ-MN) and West (SK+MD), suggesting the presence of a unique adaptation in QH sakers. This result supports the past assertion that ~2500 m represents a general threshold biological adaptation to altitude as only QH sakers live higher than this altitude (Beall 2007). However, in contrast to humans and wolves where positive selection occurred in introns, the selection detected in the sakers was exonic causing a nonsynonymous amino acid change. All but one QH sakers are homozygous CC for the selected SNP, whereas the KZ-MN sakers lack the CC genotype. This substitution causes an amino acid change in exon 4 and suggests strong and recent local selection in QH sakers, supported by the derived allele frequency analysis with  $DAF = 0.95$  in QH in contrast to 0.28 in KZ-MN, and by the dominant presence of homozygous TT in the ancestral lowland populations (SK and MD).

Our RNA-seq data reveal an important role of *EPAS1* gene expression in QTP hypoxia responses. First, we found a significant higher expression of *EPAS1* transcripts in QH than in KZ-MN, the first observation of this nature for a wild plateau species. Second, at the transcriptome level, gene interactions for *EPAS1* in QH are different from its neighbouring populations in that negative correlations were found in half of the co-expressed transcript pairs in QH, contrasting with positive correlations dominating the *EPAS1* co-expressed pairs in KZ-MN (Fig. 2b). Third, within the functional (gene co-expression) module, the connectivity of *EPAS1* transcripts in QH is significantly higher than in KZ-MN, except for one transcript that features comparable connectivity between the two populations (Fig. 3a). The regulation mode of *EPAS1* detected in our study is more complex than previously thought. *EPAS1* transcripts were inferred to potentially have pleiotropic effects in that each of them could potentially regulate ~29 transcripts on average. Different *EPAS1* transcripts were even found to have opposite transcriptional effects on the same target transcript (e.g. Fig. 3b). This transcriptional effect was also found to be population specific. We screened 24 transcripts whose expression levels were highly negatively correlated with that of *EPAS1* in QH ( $r < -0.5$ ), and compared with the same transcripts expressed in KZ-MN. We were surprised to find that all but three of these transcripts had a positive correlation with *EPAS1* expression in KZ-MN. Previous functional studies showed that most ( $N = 19$ ) have

essential roles in cellular response to hypoxia and all were upregulated under hypoxic conditions (Table S11, Supporting information). Such dominating negative correlations with *EPAS1* in QH are expected to reduce the chances of cell death, because overexpression of these 19 co-expressed genes could potentially induce apoptosis (Arany *et al.* 1996; Rolfs *et al.* 1997; Tan & Hagen 2013; Yu *et al.* 2009; Mo *et al.* 2012; Zhu *et al.* 2016; Bae *et al.* 2002; Wang *et al.* 2013; Jiang *et al.* 2015; Liu *et al.* 2007; Lichner *et al.* 2012; Yuan *et al.* 2010; Santos *et al.* 1998; Xiang *et al.* 2010; Wang *et al.* 2014a,b; Li *et al.* 2007; Guo *et al.* 2001; Kvietikova *et al.* 1995). Our results, therefore, indicate that QH sakers have evolved a robust system for buffering against hypoxia pressure.

Haemoglobins (Hb) are major oxygen-carrying molecules for animals. In our study, *HBA* (also known as  $\alpha^4$ ) was found to be the most highly expressed genes in the blood transcriptome (not shown). Interestingly, this gene was found to be positively selected both when comparing QH and KZ-MN (Table 2) and East and West sakers (Table S6, Supporting information), which is unique among the studied QTP species reported so far (Fig. S1, Supporting information). However, no significant expression difference was detected between the 'highland' (i.e. QH or East) and 'lowland' populations (i.e. KZ-MN or West). A positively selected site in *HBA* was found in the intron, and therefore may involve transcriptional regulation through regulating alternative splicing, which warrants future research.

#### Low-temperature adaptation

Low temperatures on the Qinghai-Tibetan Plateau impose another physiological limit for sakers. Studies have suggested that energy conservation may be a key strategy for organisms to survive at low temperatures (Strukie 1976). In our study, significantly fewer genes and transcripts were expressed in the blood transcriptomes of saker falcons living at relative higher altitude. The number of upregulated DETs was also found to significantly fewer at high-altitude sakers. These results, therefore, suggest that sakers inhabiting higher altitudes have a conservative blood RNA expression profile.

It has been suggested that the magnitude and activity of immune interactions has, in general, a positive correlation with temperature (Catalána *et al.* 2012), due to higher pathogen diversity or activity at higher temperatures, but to date, little is known about its underlying molecular mechanism. Among the 37 SNPs showing selection signatures between QH and KZ-MN, three were found in the exons of extracellular domain ( $\beta 1$ ) of *MHC class II B*, which has been previously demonstrated to bind peptides (Silva & Edwards 2009). The saker *MHC class II B* has been shown to be a single

copy gene (Zhan *et al.* 2015). In Table S5, Supporting information, the common alleles of selected SNP loci in KZ-MN showed reduced heterozygosity compared with those in QH. In addition, the selected SNPs in KZ-MN were fixed or nearly fixed (Table S5, Supporting information), suggesting selection sweeps in KZ-MN or relaxed selection in QTP sakers. Additional supporting evidence came from our *DAF* analysis where strong directional selection signals were detected for both SNPs in KZ-MN (1, 1 and 0.94). Interestingly, all the three SNP were also found to be under directional selection in a previous Sanger sequencing-based study using different saker samples ( $N = 17$  for QH, 64 for KZ-MN) (Fig. S14, Supporting information; Zhan *et al.* 2015). It is noted that this observation is unlikely to be linked to population structure or demographically mediated differences in genomewide diversity as the same study showed contrasting patterns of genetic diversity for intronic SNPs, for which KZ-MN possessed higher observed and expected heterozygosity values than QH (Table S12, Supporting information) and a complete lack of genetic structure when exonic SNPs were excluded (Zhan *et al.* 2015).

Contrary to *EPAS1*, the expression of *MHC class II B* was found to be significantly higher in KZ-MN than in QH (Table 2). In addition, five of seven upregulated TFs in KZ-MN were involved in immunity (Table S8, Supporting information). Furthermore, the number of transcripts co-expressed with *MHC class II B* in KZ-MN was found to be twice of that in QH (Fig. 3a). We also tested for an association between *MHC class II B* gene expression levels and twenty environmental variables related to temperature, precipitation and elevation, and found expression to be best explained by mean annual temperature (Table S9, Supporting information). This, however, should be considered as a conservative estimate because our association analysis is individual-based and the expression levels of some QH sakers even were higher than that of some KZ-MN sakers (data not shown). Taken together, our results suggest that KZ-MN sakers living at lower altitudes have a higher immune capacity. One hypothesis that has been put forward to explain this observation is that species occupying high lands with their associated low temperatures are expected to have lower parasite burdens load or prevalence of pathogens (Zamora-Vilchis *et al.* 2012), resulting in relaxed selection at the *MHC* for QH saker falcons.

It is also noted that we could not refute the possibility that hypoxia is a cofactor for the downregulation of saker *MHC class II B*, as expected from the blunted effect of long-term hypoxia exposure on immune responses (e.g. Facco *et al.* 2005). Due to the lack of ambient oxygen pressure data, we could not directly

test this effect of hypoxia in the sakers, but the contribution of hypoxia to *MHC class II B* gene expression in QH sakers could be little because there is no significant correlation between the *MHC class II B* gene expression differences and elevation, which is in direct proportion to reduction of oxygen partial pressure (Altshuler & Dudley 2006).

## Conclusion

Qinghai–Tibetan Plateau adaptation has become a research hot spot for evolutionary biologists as some key genes (e.g. *EPAS1*) have been found to underlie low-oxygen adaptations, mostly through genomic inference (Fig. S1, Supporting information). These studies, however, lack functional support (e.g. gene expression). Our study is among the first to contextualize QTP adaptation at both the sequence and transcriptome levels, and more importantly, shows synergetic responses to extreme plateau environments between DNA polymorphisms and RNA expression patterns in *EPAS1* and *MHC Class II B*.

For hypoxia adaptation on the QTP, we detected significant positive selection in the coding sequence of an *EPAS1* exon, which so far has been only reported in carnivore species (Fig. S1, Supporting information). Remarkably, we found a derived *EPAS1* variant that was upregulated in the plateau population, contrasting with the reports that lower *EPAS1* expression in Tibetan migrants than lowland British in humans (Petousi *et al.* 2014). This finding suggests a distinct role of *EPAS1* for high-altitude adaptation in birds. Compared with the neighbouring lower-altitude KZ-MN population, the intramodular connectivity of the two *EPAS1* transcripts in QH was significantly higher. In contrast with KZ-MN where most co-expressed transcripts are positively correlated, this negative correlation accounts for half of co-expressed transcript pairs in QH sakers (Fig. 2b). Moreover, the top negatively correlated genes in QH are related to hypoxia responses or cell death (Table S11, Supporting information). It is known that avian blood contains both nucleated erythrocytes and hematopoietic stem cells (Moignard *et al.* 2013), expected to be under stress in hypoxic conditions. Upregulation of inhibitory transcription factors such as *EPAS1*, therefore, might alleviate hypoxic responses in their downstream targets, providing a solution for living in a low-oxygen environment.

Our study revealed a refined regulation mode for *EPAS1*, which to our knowledge has never been reported before. As a transcription factor, an *EPAS1* transcript can regulate multiple targeted transcripts, and several *EPAS1* transcripts can also interact with others to regulate the same target (Fig. 3b). Thus, the

functions of these *EPAS1* transcripts may have become diversified in this context. One hypothesis is that different *EPAS1* transcripts have distinct DNA binding domains, and consequently, the genes that are targeted are diverse. Our findings of complex transcriptome composition in *EPAS1* uncover an important role of transcripts in the adaptation and acclimation to high-altitude environments, which has been largely overlooked in past transcriptomic studies based on genes or unique transcripts (e.g. Cheviron *et al.* 2008; Dayan *et al.* 2015).

Different from low-oxygen adaptation, relaxed selection and downregulation in *MHC Class II B* molecules suggest an adaptive response to lower temperatures, supported by our correlation analysis of gene expression and environment parameters. A generally conservative blood transcriptome was also detected in high-altitude saker populations (QH or East). These innovations in the pathways responding to low-oxygen and low-temperature environment, together with the population expansion inferred on the QTP (Fig. 1c, 1d), suggest a potential evolutionary route onto the extreme QTP environment and may explain how it became the largest breeding saker population in Eurasia (Dixon 2007). The adaptive function of these gene expression variations warrants future work (e.g. transplant experiment; Cheviron *et al.* 2008). We also note that the evolution of regulatory elements in the QTP saker population remains an open question.

The fragility of Qinghai–Tibetan Plateau ecosystem has attracted considerable attention during recent years (Cui & Graf 2009). Sakers and other raptors, as apex predators, play an import role by acting as dominant drivers of energy flow in the food web. Unfortunately, sakers face a range of threats on the QTP including habitat degradation, illegal trapping and electrocution on power lines (Li *et al.* 2000; Dixon *et al.* 2013, 2015). In our study, the inferred recent rapid expansion of QTP sakers and nearly unidirectional gene flow from Mongolia and Kazakhstan onto the plateau suggest that these low-temperature habitats are now vital for sakers, linked to their origin during a cold and arid period of the Earth's climate. In this context, habitat degradation and global warming may have a negative effect on adaptive plasticity and survival of sakers on the Qinghai–Tibetan Plateau, and this question needs to be carefully assessed in the future.

## Acknowledgements

This work was funded by the National Natural Science Foundation of China (No. 31522052) and the National Key Programme of Research and Development, Ministry of Science and Technology (2016YFC0503200). This study was also

supported by grants from the Breakthrough Project of Strategic Priority Program of the Chinese Academy of Sciences (XDB13000000), the Recruitment Program of Global Youth Experts of China to XZ and a Royal Society Newton Advanced Fellowship. Field sampling in MD, KZ and MN was funded by the Environment Agency – Abu Dhabi. We thank the Wildlife Science and Conservation Centre of Mongolia, A. Levin and D. Ragyov for assistance during sample collection, and Q. Dai for help in producing the GIS map.

## References

- Ai HS, Yang B, Li J *et al.* (2014) Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics*, **15**, 834.
- Altshuler DL, Dudley R (2006) The physiology and biomechanics of avian flight at high altitude. *Integrative and Comparative Biology*, **46**, 62–71.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.
- Arany Z, Huang LE, Eckner R *et al.* (1996) An essential role for p300/CBP in the cellular response to hypoxia. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12969–12973.
- Bae MK, Ahn MY, Jeong JW *et al.* (2002) Jab1 interacts directly with HIF-1 $\alpha$  and regulates its stability. *Journal of Biological Chemistry*, **277**, 9–12.
- Barrett JC, Fry B, Maller J *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
- Beall CM (2007) Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8655–8660.
- Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.
- Brandt SJ, Koury MJ (2009) Regulation of LMO2 mRNA and protein expression in erythroid differentiation. *Haematologica*, **94**, 447–448.
- Breiman L (2001) Random forests. *Machine Learn*, **45**, 5–32.
- Catalána TP, Wozniak A, Niemeyer HM *et al.* (2012) Interplay between thermal and immune ecology: effect of environmental temperature on insect immune response and energetic costs after an immune challenge. *Journal of Insect Physiology*, **58**, 310–317.
- Chen F, Zhang W, Liang Y *et al.* (2012) Transcriptome and network changes in climbers at extreme altitudes. *PLoS ONE*, **7**, e31645.
- Chevron ZA, Brumfield RT (2011) Genomic insights into adaptation to high-altitude environments. *Heredity*, **108**, 354–361.
- Chevron ZA, Whitehead A, Brumfield RT (2008) Transcriptional variation and plasticity in rufous-collared sparrows (*Zonotrichia capensis*) along an altitudinal gradient. *Molecular Ecology*, **17**, 4556–4569.
- Cui XF, Graf HF (2009) Recent land cover changes on the Tibetan Plateau: a review. *Climatic Change*, **94**, 47–61.
- Dayan D, Crawford DL, Oleksiak MF (2015) Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of *Fundulus heteroclitus*. *Molecular Ecology*, **24**, 3345–3359.
- Dixon A (2007) Saker Falcon breeding population estimates. *Part 1: Europe. Falco*, **29**, 4–12.
- Dixon A, Ma M, Gunga A *et al.* (2013) The problem of raptor electrocution in Asia: case studies from Mongolia and China. *Bird Conservation International*, **23**, 520–529.
- Dixon A, Ma M, Batbayar N (2015) Importance of the Qinghai-Tibetan plateau for the endangered Saker Falcon *Falco cherrug*. *Forktail*, **31**, 37–42.
- Dixon A, Li XH, Rahman L *et al.* (2017) Characteristics of home range areas used by Saker Falcons (*Falco cherrug*) wintering on the Qinghai-Tibetan Plateau. *Bird Conservation International*, in press.
- Duforet-Frebourg N, Luu K, Laval G *et al.* (2016) Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 Genomes data. *Molecular Biology and Evolution*, **33**, 1082–1093.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Excoffier L, Dupanloup I, Huerta-Sanchez E *et al.* (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.
- Facco M, Zilli C, Siviero M *et al.* (2005) Modulation of immune response by the acute and chronic exposure to high altitude. *Medicine and Science in Sports and Exercise*, **37**, 768–774.
- Fariello MI, Boitard S, Naya H, *et al.* (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, **193**, 929–941.
- Favre A, Päckert M, Pauls SU *et al.* (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews*, **90**, 236–253.
- Ferguson-Lees J, Christie DA (2001) *Raptors of the World*. London: Christopher Helm.
- François O, Martins H, Caye K *et al.* (2016) Controlling false discoveries in genome scans for selection. *Molecular Ecology*, **25**, 454–469.
- Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925–929.
- Ge RL, Cai Q, Shen YY *et al.* (2013) Draft genome sequence of the Tibetan antelope. *Nature Communications*, **4**, 1858.
- Guo K, Searfoss G, Krolikowski D *et al.* (2001) Hypoxia induces the expression of the pro-apoptotic gene BNIP3. *Cell Death and Differentiation*, **8**, 367–376.
- Gutenkunst RN, Hernandez RD, Williamson SH *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.
- Hewitt G (2000) The genetic legacy of the quaternary ice ages. *Nature*, **405**, 907–913.
- Hijmans RJ, Cameron SE, Parra JL *et al.* (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hwang DG, Green P (2004) Bayesian Markov Chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13994–14001.



- Jiang Y, Wang J, Tian H *et al.* (2015) Increased SUMO-1 expression in response to hypoxia: interaction with HIF-1 $\alpha$  in hypoxic pulmonary hypertension. *International Journal of Molecular Medicine*, **36**, 271–281.
- Ke D, Lu X (2009) Burrow use by Tibetan ground tits *Pseudopodoces humilis*: coping with life at high altitudes. *Ibis*, **151**, 321–331.
- Kerwin MW, Overpeck JT, Webb RS *et al.* (1999) The role of oceanic forcing in mid-Holocene northern hemisphere climatic change. *Paleoceanography*, **14**, 200–210.
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.
- Kvietikova I, Wenger RH, Marti HH *et al.* (1995) The transcription factors *ATF-1* and *CREB-1* bind constitutively to the hypoxia-inducible factor-1 (*HIF-1*) DNA recognition site. *Nucleic Acids Research*, **23**, 4542–4550.
- Lai CH, Smith AT (2003) Keystone status of plateau pikas (*Ochotona curzoniae*): effect of control on biodiversity of native birds. *Biodiversity and Conservation*, **12**, 1901–1912.
- Langfelder P, Zhang B, Horvath S (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R. *Bioinformatics*, **24**, 719–720.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Li J, Tibshirani R (2013) Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical Methods in Medical Research*, **22**, 519–523.
- Li YM, Gao Z, Li X *et al.* (2000) Illegal wildlife trade in the Himalayan region of China. *Biodiversity and Conservation*, **9**, 901–918.
- Li Y, Wang Y, Kim E *et al.* (2007) Bnip3 mediates the hypoxia-induced inhibition on mammalian target of rapamycin by interacting with Rheb. *Journal of Biological Chemistry*, **282**, 35803–35813.
- Li R, Yu C, Li Y *et al.* (2009a) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, **25**, 1966–1967.
- Li R, Li Y, Fang X *et al.* (2009b) SNP detection for massively parallel whole-genome resequencing. *Genome Research*, **19**, 1124–1132.
- Lichner Z, Mejia-Guerrero S, Ignacak M *et al.* (2012) Pleiotropic action of renal cell carcinoma-dysregulated miRNAs on hypoxia-related signaling pathways. *The American Journal of Pathology*, **180**, 1675–1687.
- Liu YV, Hubbi ME, Pan F *et al.* (2007) Calcineurin promotes hypoxia-inducible factor 1 alpha expression by dephosphorylating RACK1 and blocking RACK1 dimerization. *Journal of Biological Chemistry*, **282**, 37064–37073.
- Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package for performing genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **17**, 67–77.
- Mo JH, Choi IJ, Jeong WJ *et al.* (2012) HIF-1 $\alpha$  and HSP90: target molecules selected from a tumorigenic papillary thyroid carcinoma cell line. *Cancer Science*, **103**, 464–471.
- Moignard V, Woodhouse S, Fisher J *et al.* (2013) Transcriptional hierarchies regulating early blood cell development. *Blood Cells, Molecules and Diseases*, **51**, 239–247.
- Nemček V, Chavko J, Deuschova L (2014) Movement of satellite-tracked juvenile saker falcons (*Falco cherrug*) in SW Slovakia. *Slovak Raptor Journal*, **8**, 97–103.
- Nittinger F, Haring E, Pinsker W *et al.* (2005) Out of Africa? phylogenetic relationships between *Falco biarmicus* and the other Hierofalcons (Aves: Falconidae). *Journal of Zoological Systematics and Evolutionary Research*, **43**, 321–331.
- Nittinger F, Gamauf A, Pinsker W *et al.* (2007) Phylogeography and population structure of the saker falcon (*Falco cherrug*) and the influence of hybridization: mitochondrial and microsatellite data. *Molecular Ecology*, **16**, 1497–1517.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.
- Pertea M, Pertea GM, Antonescu CM *et al.* (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295.
- Petousi N, Croft QPP, Cavalleri GL *et al.* (2014) Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor system and blunted physiological responses to hypoxia. *Journal of Applied Physiology*, **116**, 893–904.
- Qiu Q, Zhang G, Ma T *et al.* (2012) The yak genome and adaptation to life at high altitude. *Nature Genetics*, **44**, 946–949.
- Qu Y, Zhao H, Han N *et al.* (2013) Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nature Communications*, **4**, 2071.
- Rolfs A, Kvietikova I, Gassmann M *et al.* (1997) Oxygen-regulated transferrin expression is mediated by hypoxia-inducible factor-1. *Journal of Biological Chemistry*, **272**, 20055–20062.
- Sabeti PC, Varilly P, Fry B *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.
- Santos BC, Chevaile A, Kojima R *et al.* (1998) Characterization of the Hsp110/SSE gene family response to hyperosmolality and other stresses. *American Journal of Physiology*, **274**, 1054–1061.
- Sekercioglu CH (2006) Increasing awareness of avian ecological function. *Trends in Ecology and Evolution*, **21**, 464–471.
- Sica A, Melillo G, Varesio L (2011) Hypoxia: a double-edged sword of immunity. *Journal of Molecular Medicine*, **89**, 657–665.
- Silva MC, Edwards SV (2009) Structure and evolution of a new avian MHC Class II B gene in a sub-Antarctic seabird, the thin-billed prion (Procellariiformes: *Pachyptila belcheri*). *Journal of Molecular Evolution*, **68**, 279–291.
- Simmons T, Nadel D (1998) The avifauna of the early epipalaeolithic site of Ohalo 2 (19,400 years BP), Israel: species diversity, habitat and seasonality. *International Journal of Osteoarchaeology*, **8**, 79–96.
- Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**, Article 3.
- Strukie P (1976) *Avian Physiology*. Springer-Verlag, New York, New York.
- Tan CY, Hagen T (2013) Post-translational regulation of mTOR complex 1 in hypoxia and reoxygenation. *Cellular Signalling*, **25**, 1235–1244.
- Tang H, Peng J, Wang P *et al.* (2005) Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology*, **28**, 289–301.

- Verity R, Collins C, Card DC (2017) MINOTAUR: a platform for the analysis 1 and visualization of multivariate results from genome scans with R Shiny. *Molecular Ecology Resources*, **17**, 33–43.
- Wang K, Chen Y, Ferguson SD *et al.* (2013) MTA1 and MTA3 regulate HIF1 $\alpha$  expression in hypoxia-treated human trophoblast Cell Line HTR8/Svneo. *Medical Journal of Obstetrics and Gynecology*, **1**, 1017.
- Wang GD, Fan RX, Zhai WW *et al.* (2014a) Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan Plateau. *Genome Biology and Evolution*, **6**, 2122–2128.
- Wang H, Hsieh YJ, Cheng WC *et al.* (2014b) JMJD5 regulates PKM2 nuclear translocation and reprograms HIF-1 $\alpha$ -mediated glucose metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 279–284.
- Wang MS, Li Y, Peng MS *et al.* (2015) Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular Biology and Evolution*, **32**, 1880–1889.
- Werger MJ, Werger MJA, vanStaalduinen MA (2012) *Eurasian Steppes: Ecological Problems and Livelihoods in a Changing World*. Dordrecht: Springer-Verlag.
- Xiang F, Huang Y, Shi X *et al.* (2010) Mitochondrial chaperone tumor necrosis factor receptor-associated protein 1 protects cardiomyocytes from hypoxic injury by regulating mitochondrial permeability transition pore opening. *The FEBS Journal*, **277**, 1929–1938.
- Yang Y, Wang L, Han J *et al.* (2015) Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude lizard living in the Qinghai-Tibet Plateau. *BMC Evolutionary Biology*, **15**, 101.
- Yi X, Liang Y, Huerta-Sanchez E *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75–78.
- Yu B, Miao Z, Li M *et al.* (2009) c-Jun protects hypoxia-inducible factor-1 alpha from degradation via its oxygen-dependent degradation domain in a nontranscriptional manner. *Cancer Research*, **69**, 7704–7712.
- Yuan Z, Zhang Y, Li X *et al.* (2010) HSP70 protects intestinal epithelial cells from hypoxia/reoxygenation injury via a mechanism that involves the mitochondrial pathways. *European Journal of Pharmacology*, **643**, 282–288.
- Zamora-Vilchis I, Williams SE, Johnson CN (2012) Environmental temperature affects prevalence of blood parasites of birds on an elevation gradient: implications for disease in a warming climate. *PLoS ONE*, **7**, e39208.
- Zhan X, Pan S, Wang J *et al.* (2013) Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics*, **45**, 563–566.
- Zhan X, Dixon A, Batbayar N *et al.* (2015) Exonic versus intronic SNPs: contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity*, **114**, 1–9.
- Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, **4**, Article17.
- Zhang WP, Fan ZX, Han EJ *et al.* (2014) Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genetics*, **10**, e1004466.
- Zhou XB, Lindsay H, Robinson MD (2014) Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*, **42**, e91.
- Zhu W, Guo W, Zhu JN *et al.* (2016) Hsp90aa1: a novel target gene of miR-1 in cardiac ischemia/reperfusion injury. *Scientific Report*, **6**, 24498.

## Data accessibility

The data sets generated during and/or analysed during the current study are available in the NCBI Short Read Archive database (PRJNA354494) and Dryad (doi:10.5061/dryad.67cq5).

X.Z. led the research team. X.Z., S.P. and M.W.B. designed the research. A.D., T.Z., L.D. and X.Z. sampled the birds. X.Z. extracted the RNA. S.P., R.Z., L.H., Q.W., S.D., Q.L., Z.G., Z.L., X.L. and X.Z. performed the analyses. X.Z., S.P., A.D., M.W.B. interpreted the results. X.Z. wrote the manuscript with input from M.W.B. and S.P. All authors read and approved the final manuscript.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Main effect genes involving in the hypoxia adaptation of the Qinghai-Tibetan Plateau.

**Fig. S2** Model testing to determine whether the first population divergence event resulted from the ancestral population contraction or expansion.

**Fig. S3** PCA analysis of transcript expression profiles of every saker falcons in each of three populations.

**Fig. S4** Contribution of each of the twenty environmental variables to the MHC gene expression.

**Fig. S6** Results of PCA analysis on the saker falcons. Principal component 1 and 4 are plotted.

**Fig. S7** *frappe* population structure reconstructed from the filtered SNP dataset (SNP with  $r^2 > 0.2$  excluded).

**Fig. S8** Demographic history reconstruction using *fastsimcoal2*.

**Fig. S9** The phylogenetic tree reconstructed using the SNP alleles sharing matrix between all sakers and the peregrine falcon.

**Fig. S10** Statistics of SNP allele frequencies in QH, KZ-MN and West saker populations (a, b) in comparison with human data (c) (Yi *et al.* 2010).

**Fig. S11** Positively selected SNPs between East and West saker populations along the saker reference genome.

**Fig. S12** The number of differentially expressed transcripts between QH and KZ-MN saker populations.

**Fig. S13** The number of differentially expressed transcripts between East and West saker populations.

**Fig. S14** Positive selection on the *MHC class II B* gene showed by a previous study (from Zhan *et al.* 2015).

**Table S1** Statistics of saker falcon RNA-seq data.

**Table S2** 19 bioclimatic variables used in the present study.

**Table S3** SNP annotation of 30 saker falcons used in the RNA-seq.

**Table S4** Demographic history parameters estimated by  $\partial a \partial i$ .

**Table S5** Genotypes of the selected protein-coding SNPs on the *EPAS1* and *MHC class II B* genes.

**Table S6** SNP candidates under positive selection between West and East saker populations.

**Table S7** GO enrichment for PSGs between West and East population.

**Table S8** Results of Transcription factor analysis between QH and KZ-MN saker populations.

**Table S9** Results of the variance analysis between the bioclimatic data and *MHC Class II B* gene expression.

**Table S10** GO enrichment for the upregulated transcripts in East saker population.

**Table S11** 24 transcripts showing high negative correlation with *EPAS1* in QH ( $r < -0.5$ ).

**Table S12** SNP diversity in the six geographic and one genetic population.

**Appendix S1** Methods.