

## Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species

Li *et al.*

RESEARCH

Open Access

# Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species

Shengbin Li<sup>1†</sup>, Bo Li<sup>2†</sup>, Cheng Cheng<sup>1,2,4†</sup>, Zijun Xiong<sup>2</sup>, Qingbo Liu<sup>1</sup>, Jianghua Lai<sup>1</sup>, Hannah V Carey<sup>5</sup>, Qiong Zhang<sup>2,6</sup>, Haibo Zheng<sup>1</sup>, Shuguang Wei<sup>1</sup>, Hongbo Zhang<sup>1</sup>, Liao Chang<sup>1,2</sup>, Shiping Liu<sup>2</sup>, Shanxin Zhang<sup>1</sup>, Bing Yu<sup>1</sup>, Xiaofan Zeng<sup>2</sup>, Yong Hou<sup>2</sup>, Wenhui Nie<sup>7</sup>, Youmin Guo<sup>1</sup>, Teng Chen<sup>1</sup>, Jiuqiang Han<sup>1</sup>, Jian Wang<sup>2,8</sup>, Jun Wang<sup>2,9,10</sup>, Chen Chen<sup>11</sup>, Jiankang Liu<sup>1,4</sup>, Peter J Stambrook<sup>12</sup>, Ming Xu<sup>13</sup>, Guojie Zhang<sup>2,9</sup>, M Thomas P Gilbert<sup>14,15</sup>, Huanming Yang<sup>2,6,8,16\*</sup>, Erich D Jarvis<sup>17\*</sup>, Jun Yu<sup>1,3\*</sup> and Jianqun Yan<sup>1\*</sup>

## Abstract

**Background:** Nearly one-quarter of all avian species is either threatened or nearly threatened. Of these, 73 species are currently being rescued from going extinct in wildlife sanctuaries. One of the previously most critically-endangered is the crested ibis, *Nipponia nippon*. Once widespread across North-East Asia, by 1981 only seven individuals from two breeding pairs remained in the wild. The recovering crested ibis populations thus provide an excellent example for conservation genomics since every individual bird has been recruited for genomic and demographic studies.

**Results:** Using high-quality genome sequences of multiple crested ibis individuals, its thriving co-habitant, the little egret, *Egretta garzetta*, and the recently sequenced genomes of 41 other avian species that are under various degrees of survival threats, including the bald eagle, we carry out comparative analyses for genomic signatures of near extinction events in association with environmental and behavioral attributes of species. We confirm that both loss of genetic diversity and enrichment of deleterious mutations of protein-coding genes contribute to the major genetic defects of the endangered species. We further identify that genetic inbreeding and loss-of-function genes in the crested ibis may all constitute genetic susceptibility to other factors including long-term climate change, over-hunting, and agrochemical overuse. We also establish a genome-wide DNA identification platform for molecular breeding and conservation practices, to facilitate sustainable recovery of endangered species.

**Conclusions:** These findings demonstrate common genomic signatures of population decline across avian species and pave a way for further effort in saving endangered species and enhancing conservation genomic efforts.

## Background

The International Union for Conservation of Nature (IUCN) and Bird Life Species has recognized over 20% of approximately 10,000 extant bird species as being threatened. As of 2014, the IUCN RedList has declared five, 1,373, and 959 species as extinct in the wild, threatened,

and near threatened, respectively. Between 1988 and 2008, the conservation status of 235 species was upgraded to higher categories of endangerment, as compared to just 32 species that were downgraded [1]. Furthermore, historical records document the extinction of at least 150 avian species since the 16th century. The principal threats leading to avian population decline have been linked to man-made environmental disasters, including over-hunting, habitat loss, pesticide abuse, and invasive species introduction [2]. To combat the ongoing decline, conservation efforts have been made, such as protective legislation, habitat restoration, captive breeding, and reintroduction,

\* Correspondence: yanghm@genomics.cn; jarvis@neuro.duke.edu; junyu@big.ac.cn; jqyan@mail.xjtu.edu.cn

<sup>†</sup>Equal contributors

<sup>2</sup>BGI-Shenzhen, Shenzhen 518083, P. R. China

<sup>17</sup>Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA

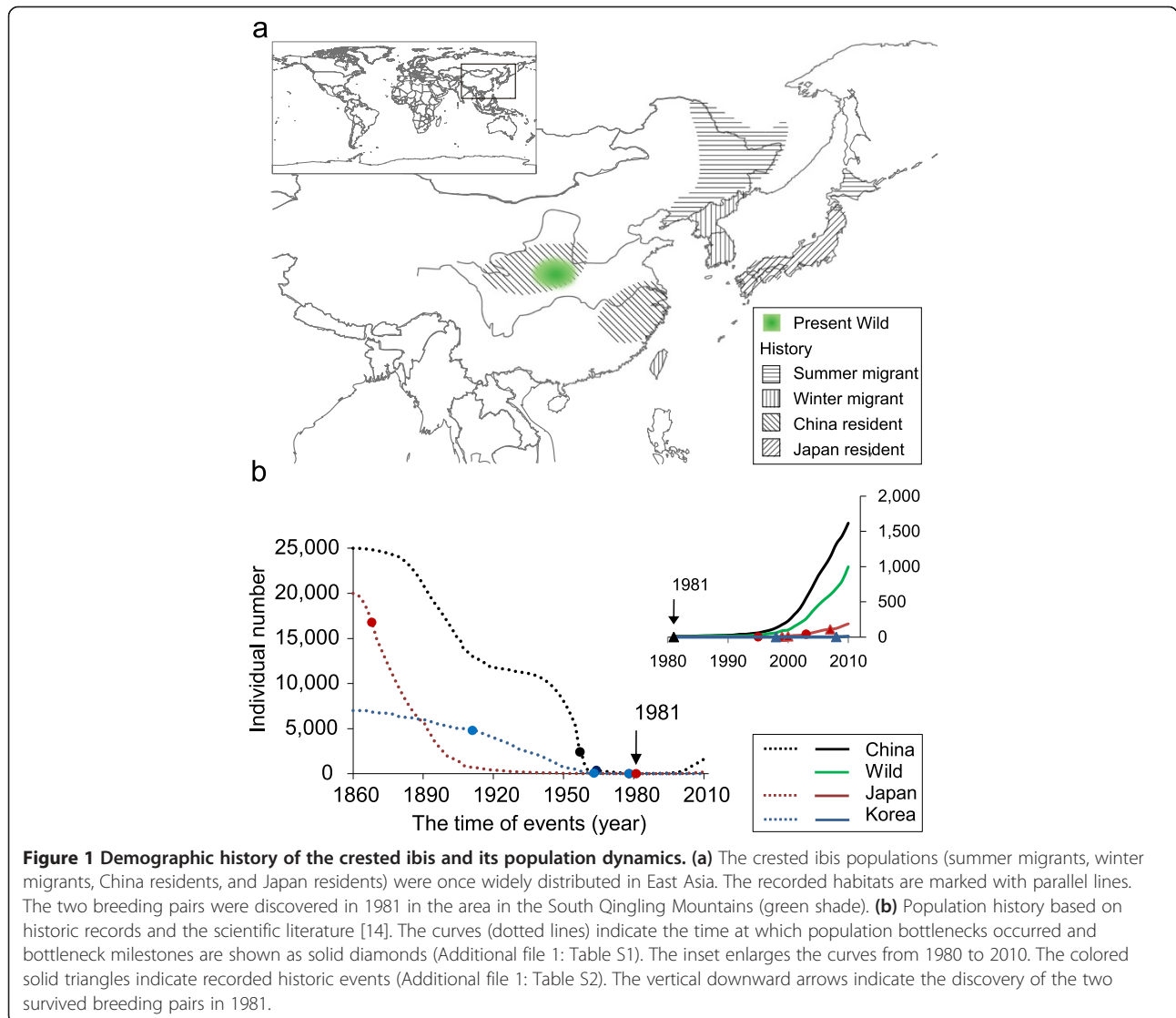
<sup>1</sup>Xi'an Jiaotong University, Xi'an 710049, P. R. China

Full list of author information is available at the end of the article

and all are responsible for the successful recovery of 49 species that were near-extinct between 1994 and 2004 [3].

Recent conservation genetic studies [4-8] have demonstrated that small populations are susceptible to allelic drift, leading to allele loss/fixation, and the process can be accelerated by inbreeding. Likewise, in small captive populations, rapid genetic deterioration, such as inbreeding depression and genetic adaptation to artificial environment, can also occur [8]. Deleterious mutation tends to accumulate due to reduced selective strength [4]. Furthermore, extinction rate in small wild populations increases significantly as heterozygosity decreases [5]. Several genetic studies have attempted to characterize this effect from conservation-related bottlenecks among avian species, albeit based on limited markers of allozymes or microsatellites [9,10]. It has been proposed that studies using up-to-date and more informative markers at a genome-scale will be necessary [11].

One of the most recently endangered bird species in the world is the crested ibis (*Nipponia Nippon*; IUCN Red Data Book, BirdLife International 2001). It was once widespread across Northeast Asia, with a range encompassing China, Russia, Korea, and Japan until the 1860s (Figure 1a). Suffering from over-hunting and habitat loss, the crested ibis populations had finally collapsed in the late 19th and early 20th centuries, to the extent that it was thought to be completely extinct from the wild, when the last five birds were taken into captivity in Japan in 1981 [2] (Figure 1b; Additional file 1: Tables S1, S2). Although extensive private and government-supported efforts were made to revive the species in Japan, the final five individuals taken into captivity died without producing offspring. However, through an extensive international survey effort aimed at saving the species from extinction, a remarkable discovery was made later in 1981. After over 17 years of no sightings in China (since one loner bird seen in



the wild in 1964), seven individuals belonging to two breeding couples (four adults, of which one couple had three chicks) were found at the edge of their normal habitats in a remote location in the southern foothills of the Qinling Mountains (Yangxian County, Shanxi). To protect these seven birds, China immediately established the Yangxian Nature Reserve, and a conservation program for both wild and captive birds was started. Monitor stations were set up for every nest, GPS-based systems for individual identification and tracking, and guidelines that control pesticides and habitat destruction were introduced [12-14]. As a result, the two breeding pairs became what can be considered the 'Adam and Eve' for the recovered crested ibis population that has gone through two phases in over 33 years: initially, a small size (<100) between 1981 to 2001, and up to over 2,000 individuals currently (Figure 1b).

To provide genome-scale insights into the near-extinction and rescue, we sequenced the genomes of multiple individuals from both the crested ibis ( $n = 9$ ; from the rescued population) and its co-habitant, not-endangered close-relative, the little egret (*Egretta garzetta*;  $n = 6$ ; from the same order *Pelecaniformes*; diverged approximately 57 million years ago) [15]. We compared their genome sequences with those of 41 other avian species described in companion publications in this issue and elsewhere [16], which include seven Endangered + Vulnerable (EV) species listed by the IUCN within the recent past (crowned crane, *Balearica regulorum*; macQueen's bustard, *Chlamydotis macqueenii*; brown mesite, *Mesitornis unicolor*; kea, *Nestor notabilis*; dalmatian pelican, *Pelecanus crispus*; white-tailed eagle, *Haliaeetus albicilla*; and bald eagle, *Haliaeetus leucocephalus*) and 31 Least Concern (LC) species (Additional file 1: Table S3). We found common genomic signatures among the endangered or recently endangered species and that in the ibis was associated with feeding behavior, climate change, environmental hazard, and man-made disasters. We also found that the ibis populations are rapidly evolving and possess greater genetic diversity than expected in the recovery process. To better assist protection and recovery efforts for the crested ibis, we developed technical platforms and molecular tools, which may also be useful for the rescue and protection of other endangered wildlife.

## Results

### The reference genome assemblies and annotations of the crested ibis and its cohabitant the little egret

With the crested ibis being our focused EV species, we first assembled its reference genome *de novo* from high-coverage (approximately 156 $\times$ ) sequence reads generated from a series of libraries constructed with various insert sizes, using SoapDenovo [17]. The assembly reached a contig N50 size of 67 kb and a scaffold N50 size of

10.7 Mb (Table 1 and Additional file 1: Table S4), and upon adding 282 Gb single-molecule optical mapping data, its super-scaffold N50 size increased to 26 Mb. Using the conserved chromosomal organization of the chicken and zebra finch genomes, we anchored 1 Gb super-scaffolds to the chromosomes, which constitute 82% of the estimated genome size. The final reference genome represents one of the more complete assemblies in the current avian genome study [18]. We validated the reference genome with alignment to eight fully-assembled fosmid sequences (98.4% alignment; Additional file 1: Table S5) and transcriptomic data (RNA-seq from two other blood samples; 95.0% alignment; Additional file 1: Table S6). Using similar procedures, we also generated a reference genome (approximately 70 $\times$  coverage) from a male little egret, which has N50 contig and scaffold sizes of 24 kb and 3.0 Mb, respectively (Additional file 1: Table S4). Our annotation efforts predicted 17,163 and 17,032 genes for the crested ibis and the little egret, respectively (Additional file 2: Figure S1). For the crested ibis, approximately 55% of the predicted genes were validated based on about 107 million mRNA sequences from blood ( $\geq 1$  RPKM (reads per kilobase per million)). In addition, the repeat contents of the two birds constitute 6.6% and 7.4% of the genome length for the ibis and egret genomes, respectively (Additional file 1: Table S7), similar to that of the zebra finch [19] but less than that of the chicken [20]. The comparative genome assembly statistics and annotations of the other 41 EV and LC avian species are reported in [16]. Our study represents the first effort of comparative genomic analyses based on the genome information generated from these EV and LC species.

### Low heterozygosity among the EV species, and its ongoing loss but with signs of increased diversity in the recovered crested ibis population

Genetic diversity has been shown to buffer species against widespread epidemics of infectious agents and parasites; its decrease has been thought to have detrimental effects on population health and survival [21,22]. Prior studies have proposed that EV species have low genetic diversity [9,10], but all were based a limited number of neutral genetic markers. We measure genetic diversity of the EV species by aligning high-quality reads from the genome sequences of individual birds against their reference genomes [23-25]. We first compared two to three species in the same order according to the phylogeny based on whole genome sequences from a companion study [15], but with different IUCN conservation statuses. In all cases, the EV species within the pairing scheme showed significant heterozygosity reduction relative to the LC control species, with the severest found in the crested ibis and the kea (Table 2). The numbers of heterozygous SNP loci are 478,836 (heterozygosity,  $0.43 \times 10^{-3}$ ) in the crested

**Table 1 Data statistics of the crested ibis and the little egret**

Crested ibis			Little egret	
<i>De novo</i> sequence				
Insert size (bp)	Total data (Gb)	Sequence coverage (x genome)	Total data (Gb)	Sequence coverage (x genome)
170-800	141	113	53	42
2-20 × 10 <sup>3</sup>	54	43	36	29
Total	195	156	89	71
Assembly				
	N50 (kb)	Total size (Gb)	N50 (kb)	Total size (Gb)
Contigs	67	1.24	24	1.16
Scaffolds	10,687	1.26	3,051	1.21
Annotation				
	Number	Total length (Mb)	Number	Total length (Mb)
Repeats	840,220	80	861,413	86
Genes	17,179	343	17,032	326
Coding exons	148,003	24	139,080	23
Resequencing <sup>a</sup>				
	Total data (Gb)	Sequence coverage (x genome)	Total data (Gb)	Sequence coverage (x genome)
Each individual	25	20	30	24
All individuals	200	160	149	119
SNP calling				
	Total SNPs	Mutation rate <sup>b</sup> (θ <sub>s</sub> ) (×10 <sup>-3</sup> )	Total SNPs	Mutation rate <sup>b</sup> (θ <sub>s</sub> ) (×10 <sup>-3</sup> )
Autosome	1,188,930	0.33	8,876,153	2.91
Coding exons	13,459	0.17	91,820	1.45

<sup>a</sup>Paired-end libraries with insert size of 500 bp were used for resequencing. Samples were from eight crested ibis and five little egrets.

<sup>b</sup>The population mutation parameter  $\theta_s = K/aL$ ,  $a = 1 + 1/2 + \dots + 1/(n-1)$ , where K is the number of variant sites found by sequencing n chromosomes in a region of length L.

ibis genome, which is approximately 6 times less than in the little egret genome (2,683,899; heterozygosity,  $2.51 \times 10^{-3}$ ; Figure 2a). This finding was confirmed in analyses that compared all eight EV and 32 LC species regardless of phylogenetic relationship, showing

that the average heterozygosity rate of the EV species is  $1.08 \times 10^{-3}$ , significantly smaller than that of the LC species of  $2.49 \times 10^{-3}$  (Figure 2b; Additional file 1: Table S8). Our analyses on heterozygous SNPs of protein-coding sequences confirmed similar heterozygosity

**Table 2 Heterozygous SNPs in nine representative avian species**

Common name	Latin name	Order	Conservation status	Heterozygosity of whole genome (10 <sup>-3</sup> ) <sup>a</sup>	Heterozygosity of whole exons (10 <sup>-3</sup> ) <sup>a</sup>	NS/S <sup>b</sup>
Crested ibis	<i>Nipponia nippon</i>	<i>Ciconiiformes</i>	EV (EN)	0.43	0.24	0.69
Little egret	<i>Egretta garzetta</i>	<i>Ciconiiformes</i>	LC	2.51	0.92	0.41
Dalmatian pelican	<i>Pelecanus crispus</i>	<i>Pelecaniformes</i>	EV (VU)	0.6	0.34	0.84
Great black cormorant	<i>Phalacrocorax carbo</i>	<i>Pelecaniformes</i>	LC	1.39	0.56	0.57
Kea	<i>Nestor notabilis</i>	<i>Psittaciformes</i>	EV (VU)	0.91	0.45	0.63
Budgerigar	<i>Melopsittacus undulatus</i>	<i>Psittaciformes</i>	LC	4.31	2.21	0.32
Bald eagle	<i>Haliaeetus leucocephalus</i>	<i>Accipitriformes</i>	EV (Once EN)	0.43	0.22	0.84
White-tailed Eagle	<i>Haliaeetus albicilla</i>	<i>Accipitriformes</i>	EV (Once VU)	0.4	0.21	0.80
Turkey vulture	<i>Cathartes aura</i>	<i>Accipitriformes</i>	LC	1.18	0.59	0.58

We classified endangered (EN) and vulnerable (VU) into endangered-vulnerable (EV) for the convenience of genetic analysis.

<sup>a</sup>P < 0.05.

<sup>b</sup>P < 0.01 (t test, EV versus LC).

LC, Least Concern; NS/S, the ratio of non-synonymous/synonymous heterozygosity.

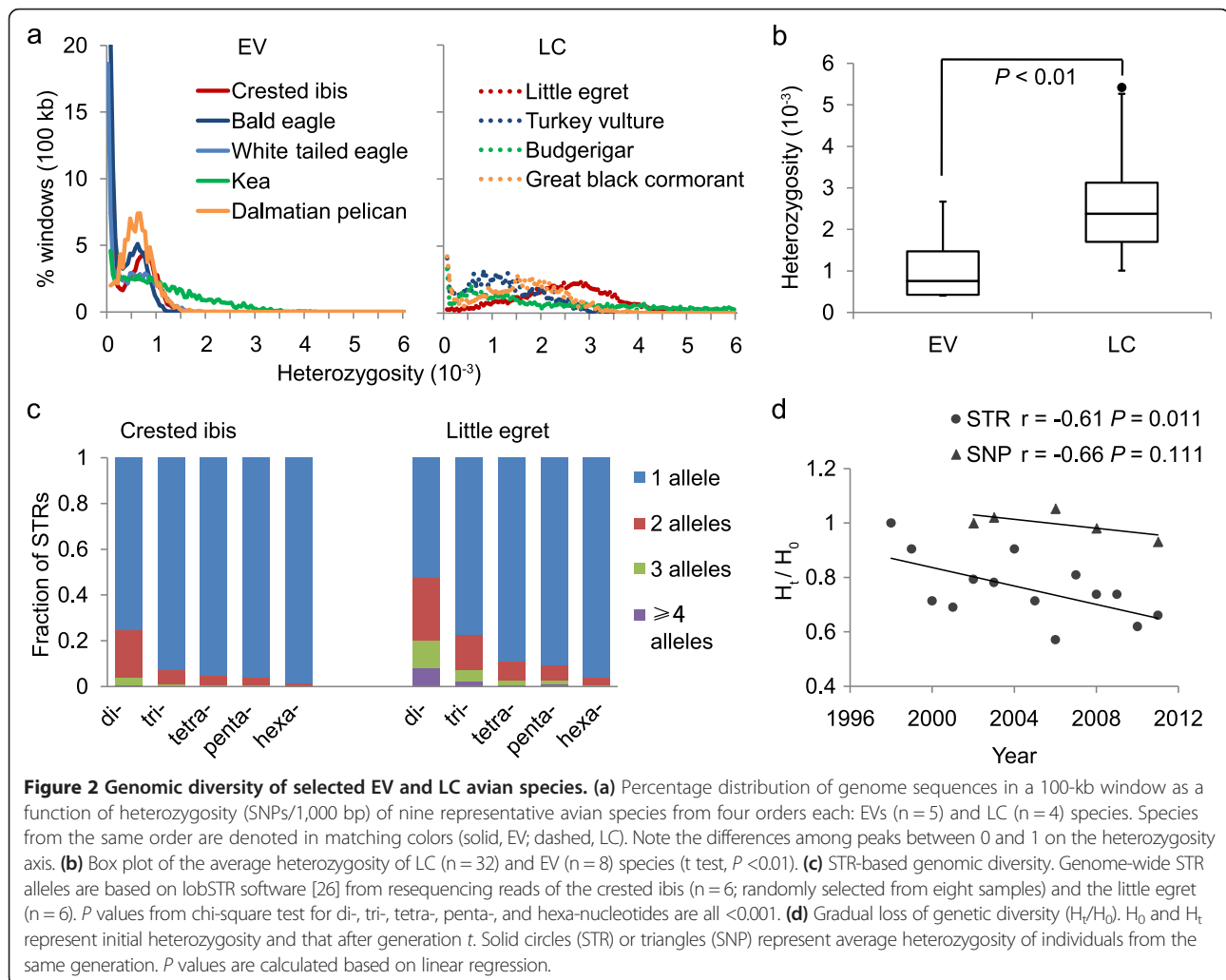


reduction in the EV species (Table 1; Additional file 2: Figure S2).

To confirm this genomic signature at the population level, we analyzed SNP and STR calls, using the moderate-coverage genome sequences (approximately 20×) of eight crested ibis and five little egret individuals, which were sampled from the same populations as used for the reference genomes. We found a dramatic one-eighth (8 times less) SNP density reduction of the autosomes in the crested ibis population as compared to the little egret population (Additional file 1: Tables S9 and S10). The average frequency of short (1 to 2 bp) STR loci for the crested ibis genomes (0.7%; 2-bp, ≥4 alleles; n = 6) was an order of magnitude lower than that of the little egret genomes (8.0%, n = 6; Figure 2c). The longer STR loci (3- to 6-bp) also showed magnitudes lower frequencies, similar to what were seen in the crested ibis (Figure 2c).

Taking advantage of the extensive identity-tracking data from the living crested ibis populations, we asked if there is still ongoing heterozygosity reduction over time

during species recovery. By analyzing 31 well-defined STR loci (4-bp unit), we estimated  $H_t/H_0$  [4,27] (heterozygosity at generation  $t$ /initial heterozygosity) from 105 individuals (Additional file 1: Table S11) and found a negative  $H_t/H_0$  correlation ( $r = -0.61$ ) with population recovery time (with a slope of 0.017 units lost per year; Figure 2d). The SNP-based  $H_t/H_0$  of the eight re-sequenced ibis individuals also displays a strong negative correlation with population recovery time ( $r = -0.66$ ), although there was no significant change with time, most likely due to the limited data points (Figure 2d). To further investigate the genetic basis of this heterozygosity reduction, we calculated fixation index among four sub-populations derived from the two original breeding pairs, including their offspring kept in the original Yangxian Nature Reserve population. We found remarkably large fixation indices among the sub-populations (Additional file 2: Figure S3), despite that the first population split occurred only about 20 years ago. This points to the presence of signs for rapidly increased genomic diversity between separated



populations, even though its smaller population size and physical isolation all lead to ongoing heterozygosity reduction.

### Accumulation of deleterious mutations in the threatened species

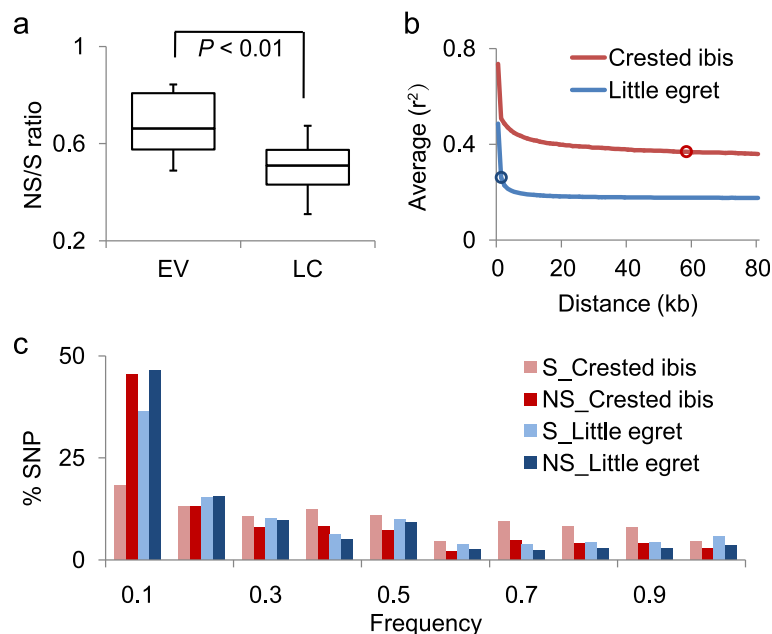
Non-synonymous changes often lead to functional, sometimes deleterious, changes in proteins [28], and inbreeding is thought to contribute to the increase of deleterious mutations in a population [29]. We wondered if there are genes bearing more non-synonymous mutations in EV species and assessed the ratio of non-synonymous/synonymous (NS/S) heterozygous SNPs between the two haploid sequences within a diploid genome of a given species. We found that the eight EV species show much higher NS/S (0.68, median) than their closely related LC species (Table 2) or the combined ratio of all 32 LC species (0.50; Figure 3a, Additional file 1: Table S8), with the highest found in the two eagles and dalmation pelican, followed by the kea and the crested ibis (Table 2). Similarly, at the population level, the NS/S ratio of the eight other crested ibis individuals ranges from 0.66 to 0.70, as compared to 0.44 to 0.48 for the five little egret individuals, and this difference is significant (Additional file 1: Table S10).

We tested whether the higher NS/S ratios can be attributed to stronger linkage disequilibrium due to inbreeding in a small effective population [30], using the multiple

sequenced individuals. The crested ibis population has a slow linkage disequilibrium (LD) decay with reduced  $r^2$  correlation coefficient at half of its maximum and at a distance of approximately 60 kb as compared to the little egret population with a distance of approximately 1 kb (Figure 3b). A similar slow LD decay has been observed in highly-inbred domestic species, such as horse [31] and dog [32]. Furthermore, the synonymous SNP fraction of the derived alleles at a low frequency of 0.1 in the crested ibis population is just half that of the non-synonymous SNPs, whereas the two values are either higher or comparable in the little egret (Figure 3c). It appears that the decreased proportion of low-frequency synonymous SNPs relative to non-synonymous SNPs is a result of inbreeding fixation in the small crested ibis population.

### Genes involved in brain function and cytochrome P450 metabolism have allelic fixation in the recovered crested ibis population

To find out whether the SNP fixation we observed in the recovered ibis population was randomly distributed among the genomes or specific to certain segments and genes, we utilized a method that identifies differences in rates of fixed SNPs, which are assumed, but does not necessarily have to occur by selective sweeps within the genomes [33]. Specifically, we calculated heterozygosity ( $H_p$ ) and its  $Z$  transformations,  $ZH_p$  (Figure 4, see Materials and methods) in 500-kb sliding windows ( $n = 2,514$ ) along



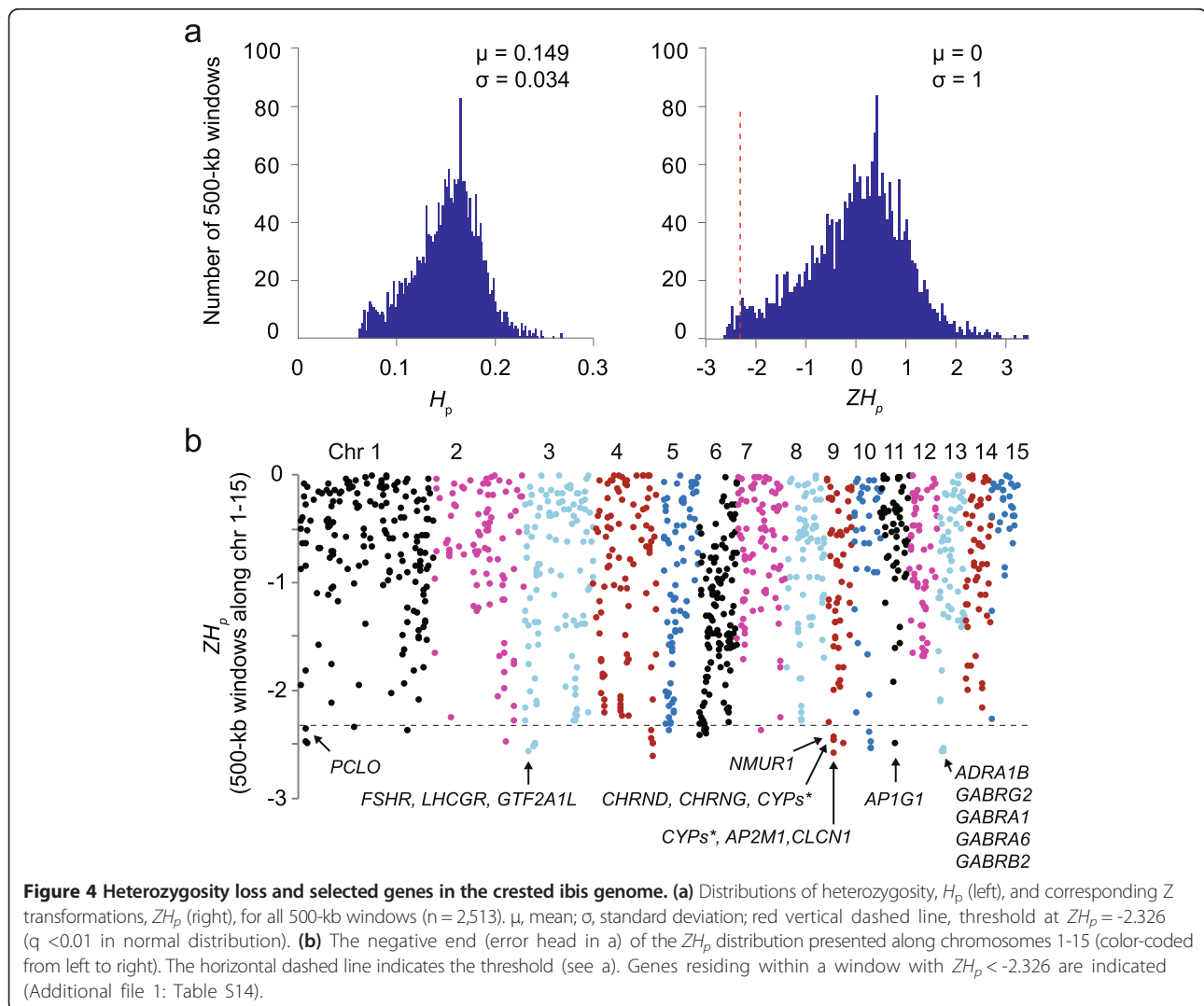
**Figure 3 Accumulation of deleterious mutations.** (a) Box plot of NS/S (non-synonymous/synonymous) ratio (based on heterozygous SNPs) in the LC ( $n = 32$ ) and EV ( $n = 8$ ) species (t test,  $P < 0.01$ ). (b) LD (Linkage disequilibrium) decay of the crested ibis and the little egret genomes. Open circles denote distances where the  $r^2$  correlation coefficient reduces to half of its maximum (approximately 60 kb for the crested ibis and approximately 1 kb for the little egret). (c) SNP fractions as derived allele frequencies in populations of the crested ibis ( $n = 9$ ) and the little egret ( $n = 6$ ). NS, non-synonymous; S, synonymous.

whole genomes (except for sex-chromosome scaffolds) for the most and least frequently observed alleles at all SNP positions. From the distribution of observed  $H_p$  and  $ZH_p$  (Figure 4a), we defined a threshold of fixed SNPs ( $ZH_p$  score equal to -2.326 or less,  $P < 0.01$ , Figure 4b). The smallest  $H_p$  values represent the least frequently observed alleles. We found approximately 1.4% of the windows ( $n = 36$ ) had a  $ZH_p$  score -2.326 or less (Figure 4b), and thus were significant outliers for the whole genome. We examined all candidate genes that resided in these regions (Additional file 1: Tables S12), and categorized them according to Gene Ontology (GO) terms (Additional file 1: Tables S13 and S14). Seven categories were statistically enriched, and nearly all involved in brain function: postsynaptic membrane, synapse part, GABA-A receptor activity, GABA signaling pathway, extracellular ligand-gated ion channel activity, chloride transport, and clathrin adaptor complex. Of the genes involved, 10 were over-represented in the neuroactive ligand-receptor interaction

pathway (KEGG map04080) that processes information from exogenous signals using neurotransmitters and signaling molecules, including epinephrine, norepinephrine, neuromedin, melanin-concentrating hormone, GABA, and acetylcholine (Additional file 1: Table S13). In addition, four of the cytochrome P450 genes stood out, which function in metabolism of hormones and toxins, including arachidonic acid (map00590) and linoleic acid (map00591) metabolism pathways that process essential fatty acids and play an important role in immune modulation [34]. These findings suggest that the fixation in the genome is not random.

#### Population bottlenecks of the crested ibis and immune genes

It is possible that the alleles became fixed through a bottleneck affect reducing the polymorphisms in the genome driven by genetic drift. Population bottleneck refers to a sharp population size reduction due to environmental

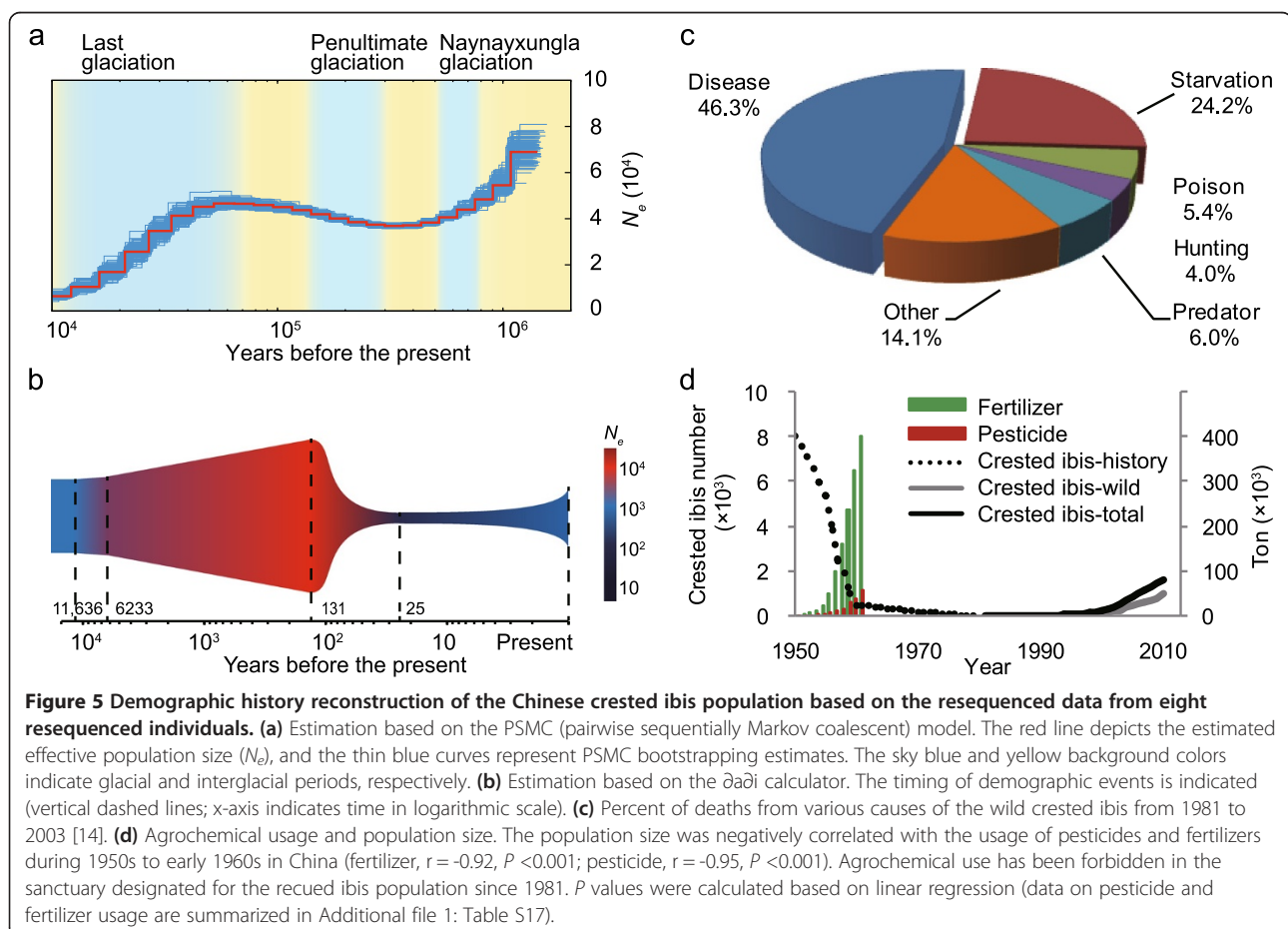




events or human activities. With sequence data from whole genomes, and of multiple individuals, we can calculate population bottlenecks more reliably than with several genes and neutral markers. We reconstructed the crested ibis' demographic history using our sequence data based on a pairwise sequential Markovian coalescent (PSMC) model [35] and a diffusion approximation method for demographic inference (*∂a∂i*) [36] (see Materials and methods). This analysis revealed two ancient and one recent bottleneck (Figure 5a). The two ancient events (1 to 0.01 MYA) occurred during the Naynayxungla glaciation (0.78 to 0.50 MYA) [37] and the last glaciation (Ice Age, 0.01 MYA), resulting in a precipitous population drop to approximately 10,000 individuals at the end of the last glaciation. The recent event leads to a drastic population reduction throughout the last 100 years, terminating approximately 25 years ago with the human assisted recovery effort (Figure 5b, Additional file 1: Table S15). This timeframe is not only consistent with the known historic records of the ibis species [2], but also reflects a blend of global climate change [37], intensified human activity [3], and remarkably, population recovery after conservation

efforts commenced some 30 years ago. This result is very different from the little egret in which population decreased during the Penultimate glaciation (0.3 to 0.13 MYA), but only slightly decreased in the last glaciation (Additional file 2: Figure S4).

To search for possible genetic causes for the recorded animal deaths among the recovered population, we scrutinized the records and found that the crested ibis population in the wild, while in the human-assisted recovery, has still been suffering from parasitic infection and other diseases, which account for 46.3% of total deaths from 1981 to 2003 (Figure 5c) [13]. Since the major histocompatibility complexes (MHC) and its genetic variants are critical for immunity [22], we analyzed the nucleotide sequences of the classical MHC class II  $\beta$  gene (*BLB*) antigen binding region (encoding a protein for presenting antigenic peptides to helper T cells). This gene shows a much lower genetic diversity (1 locus with  $\geq 3$  amino acid alleles) in the antigen binding domain than its homolog of the little egret (6 loci with  $\geq 3$  amino acid alleles; Additional file 2: Figure S5). Such a low level of genetic diversity in the *BLB* and other MHC



genes may result in defective immunity of the crested ibis population.

#### Historic agrochemical overuse and relevant mutated genes in threatened avian species

Overuse of various nondegradable agrochemicals has been suggested as one of the principal reasons for the population decline of seven of the eight EV species examined: bald eagle [38], white-tailed eagle [39], kea [40], Adele penguin [41], emperor penguin [42], chimney swift [43], and the Dalmatian pelican [44]. To investigate whether agrochemical use also contributed to the decline of the crested ibis survival, we first examined the relevant historical evidence. In Japan, during the Meiji Restoration (in the late 19th century), traditional protection measures were disregarded and rampant hunting rapidly reduced the crested ibis population to the extent that by the time when the species was enlisted for protection on the hunting ordinances in 1908, it was almost extinct there [2]. Although relevant evidence was poorly documented in Korea, Northern China, and Russia in the first half of the 20th century, in central China, the crested ibis was common in Gansu and Shaanxi Provinces before 1950 but nearly extinct by the end of 1950s. We found a negative correlation between the estimated crested ibis population size in the Gansu and Shaanxi Provinces and the amount of fertilizers and pesticides used in the region (Figure 5d). These findings suggest that overuse of agrochemicals may be associated with very dramatic and obvious decline in the crested ibis population of the region from which our genomes were sequenced.

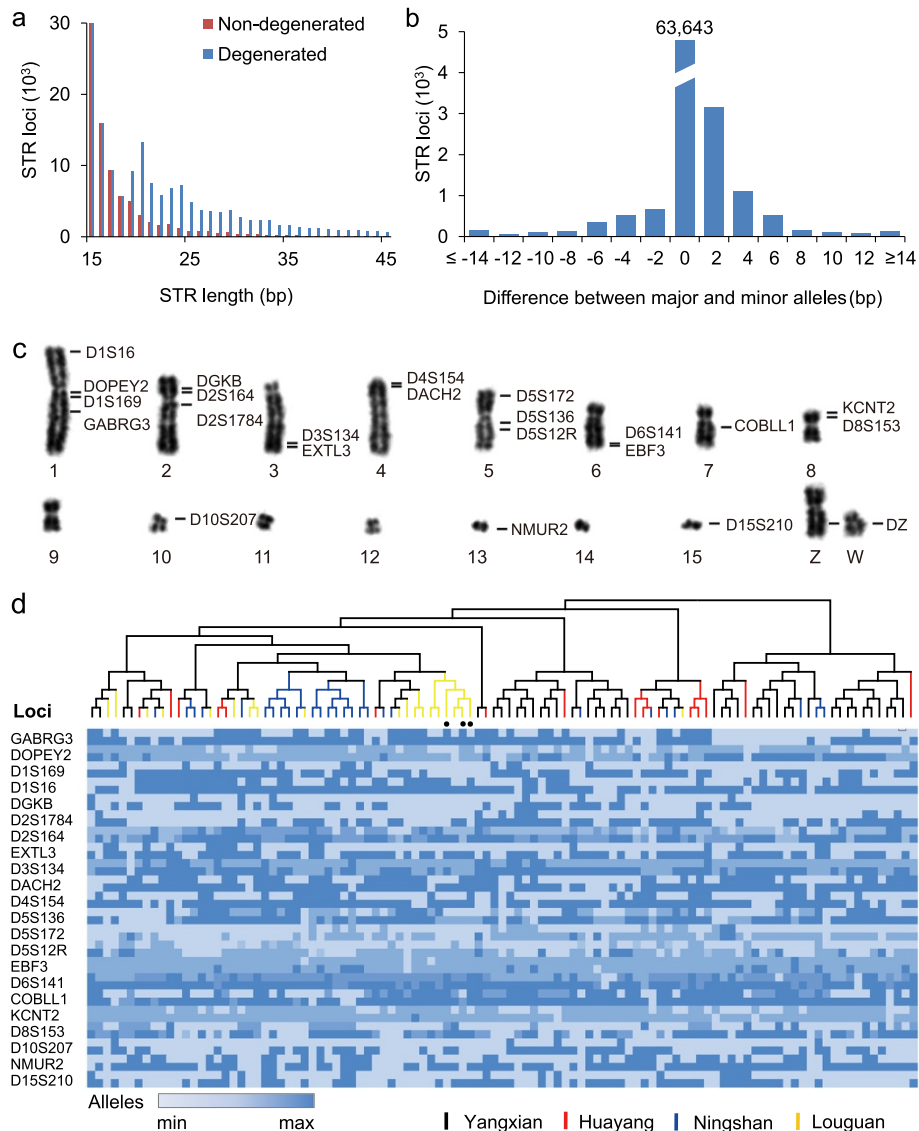
We compared 6,332 orthologs genes among EV (all are carnivorous species;  $n = 8$ ) and also to the LC carnivorous species ( $n = 15$ ), since carnivorous species are also often apex predators and more sensitive to agrochemicals [38,39,41,44]. We identified 44 genes that have a significantly higher rate of being inactivated (null mutations that alter protein structure) in the EV carnivorous species, and only nine genes with a significantly higher rate in the LC carnivorous species (Fisher's exact test,  $P < 0.05$ ; Additional file 1: Table S16). Among them, 17 genes are metabolism related enzymes; for instance, one of them, *SLCO1A2*, a sodium-independent transporter mediating cellular uptake of organic anions in the liver [45], has lost its function in three threatened species (37.5%) and in none (0%) of the LC species. Another, *HACL1*, catalyzing a carbon-carbon cleavage reaction, is necessary for the catabolism of phytaoic acid in carnivores [46], which has lost its function in three threatened species. *CHIA*, which degrades chitin-chitotriose and participates in the defense against pathogens [47], has lost its function in five threatened species. These findings suggest that carnivorous EV species have a greater genetic susceptibility to agrochemicals.

#### Genome-wide STR profiling of the crested ibis population for marker-assisted breeding

Conservation and rescue of a species from near-extinction has been partly assisted with having genetic loci to track individuals to reduce inbreeding [11]. However, there have been limited number of markers that can do so, including for the crested ibis. Further, changing conservation practice from small-scale captive breeding to STR-assisted breeding in large Nature Reserves can theoretically be benefitted by genome-scale approaches for genetic markers. In this study, we identified approximately 166,000 degenerate STR loci (sequences containing insertions, deletions, or mismatches) from the crested ibis genome sequences (Figure 6a) and tested them against a population of nine crested ibis individuals (18 haplotypes). On average, minor STR alleles exhibited a 4-bp difference from their major alleles and 18% of the allelic differences differed by  $>5$  bp over their major alleles (Figure 6b). Focusing on the 4-bp STRs, we confirmed heterozygosity for 300 such loci within and across some of the crested ibis sub-populations (105 individuals) and selected a set of 23 informative STR loci, including 22 autosomal loci and one sex-linked locus (distinguishing male from female with a 31-bp insertion in W chromosome) to establish a DNA identification profiling (DIP) platform (Figure 6c and d, and Additional file 2: Figure S6; Additional file 1: Table S18). The new sex marker accurately discriminated between males (ZZ) and females (ZW) (Additional file 2: Figure S7), an important advance since morphological-based gender determination for the crested ibis is quite difficult. This DIP marker set had a collective discrimination power of 0.628 and an exclusion probability of 0.994 (Additional file 1: Table S19). This platform demonstrated an estimated probability of paternity up to 99.997% on pedigree analysis in a four-generation family (Additional file 1: Table S20). Our DIP platform is now being used in reconstructing pedigree relationships, establishing individual identity for the recovering crested ibis populations and assisting non-sibling or genetically distant parental pairing.

#### Discussion

Our genome-wide analysis on the endangered crested ibis and seven other recently endangered and rescued avian species across the Neoaves phylogenetic tree provides direct evidence at a genomic-scale for support of previous hypothesis and novel insights into consequences of heterozygosity loss, deleterious mutation accumulations, population bottlenecks, and genetic drifts. The convergent inactivation (or pseudogenization) of xenobiotic metabolism related genes in the ibis and other endangered top predators suggests a reduction of adaptive genetic plasticity in these species to agrochemical overuse. However, the increasing genomic diversity among the isolated ibis populations derived from the offspring of the last wild pair identified in



**Figure 6 Genome-wide STR profiling of four ibis sub-populations. (a)** STR (units of 2 bp, 3 bp, 4 bp, 5 bp, and 6 bp) distribution as a fraction of the total repeat length. Non-degenerate STRs do not contain insertions, deletions or mismatches. **(b)** Near random distribution of allele size differences between the major and minor alleles ( $n = 9$ ). Size difference is calculated by subtracting the minor allele length from the major allele length. **(c)** Genetic markers of the ibis chromosomes typed in this study. Twenty-two representative STR and a single sex chromosome (W)-derived markers are shown here. **(d)** Individual identification based on the 22 STR loci. The colored horizontal scale bar indicates the number of repeat units (from minimum to maximum). The alleles (105 individuals) are used to construct neighbor-joining tree within sub-populations (Yangxian,  $n = 42$ ; Ningshan,  $n = 27$ ; Huayang,  $n = 16$ ; and Louguan,  $n = 20$ ). Solid circles denote the three individuals from a single family.

1981 indicate that rapidly diverging sequences in the recovering ibis population are being fixed in less than 10 generations.

Our genome-wide data are important for exploring causative factors of the near-extinction and exact demographic reconstruction of the endangered species, and both are necessary for distinguishing long-term climate change from recent human-mediated events [11,48]. In our case, we identified distant bottlenecks due to the

past glaciations and the most recent bottleneck that is clearly unrelated to global glaciations (although severe periodic temperature drops may happen to accelerate the process) but associated with some man-made factors. The man-made induced bottleneck was much more severe than the glaciations. It appears that the crested ibis is more sensitive to these environmental challenges than the little egret. Similar events may have affected endangered non-avian species, such as the giant panda

[49]. However, our observations on the EV avian species provide an avian model for conservation genomics, which is distinctly different from giant panda whose genetic diversity remains high [25,49] albeit with a similar population size (approximately 2,000 for the crested ibis vs. approximately 2,500 for the giant panda). We propose that, regardless of the past conservation success, an immediate evaluation of genetic diversity and sequence variation should be imposed for risk assessment on all endangered species.

The genetic drift for fixation of changes in brain and metabolism genes of the rescued crested ibis population in China is intriguing, and could mean either deleterious mutations in these genes or the possibility of enhanced functions for certain brain behaviors and enhanced metabolism of toxins for survival of the species. Consistent with the former possibility, the surviving crested ibis in China may have gone through a change in foraging behavior [2]. As a wading bird, the crested ibis uses a 'remote touch' mechanism to detect the movement of their prey in the mud through a series of sensory receptors [50], and either sacrifice or gain of sensing and digesting abilities are all relevant. We do not know if this change occurred demographically before or through genetic drift after the two breeding pairs were rescued in 1981.

One interpretation of the fixation findings based on the methods we used is that there have been selective sweeps for specific SNPs in specific brain and metabolism genes over the past 30 years. These sweeps could have artificially occurred due to controlled inbreeding, or naturally occur due to selection. But such selective sweeps within such a short time, for animals that reach sexual maturity at around 3 years with limited generations seems remarkable; although we see more rapid increasing diversity in the genome than expected. An alternative, more likely interpretation is that greater fixation of these alleles was already present in the two last breeding pairs before near extinction due to demographic differences [51]. This difference is difficult to test considering all the animals we sequenced are descended from the last seven individuals in the wild from one population in 1981, where all others are extinct. If a demographic explanation were true, it would mean that these alleles became fixed through a bottleneck affect reducing the polymorphisms in the genomes by genetic drift.

One question that can now be better addressed is why the crested ibis nearly suffered extinction, whereas its cohabitant, the little egret, did not? One possible reason is that while both species exploit aquatic environments, such as eating mollusks, crustaceans, fish, and frogs, the little egret also consumes plant seeds in the winter or under drought and thus still thrives strongly in the same habitat. This foraging behavioral difference is consistent with genetic differences in enzymes for food digestion.

Another possible reason is that the little egret might have become resident birds and gone through a bottleneck already by changing their foraging behavior at the same time.

Our genome-wide STR markers and its application to conservation genomics also provide a more powerful platform for breeding and tracking of endangered species both in partial captivity and in the wild. With this method, we believe that it is possible that immediate genome sequencing and evaluation of genetic diversity and loss-of-function genes for risk assessment can be done for generating rescuing strategies for other currently endangered avian species.

## Conclusions

Our study is the first, of which we are aware, to conduct genome-scale analyses of species that were endangered, including near extinct, across a vertebrate class. With a focus on the crested ibis, we were able to identify genetic associations before, during, and after the near extinction events and population bottlenecks. We confirmed some expected changes, but genome-wide, such as reduced heterozygosity, accumulation of deleterious mutations, and susceptibility to agrochemical overuse by humans. We also identified seemingly positive changes in the recovering crested ibis population, such as more rapidly increasing genetic diversity between new populations than expected, and changes in some gene families that could potentially be related to surviving extinction or recovery. Our genome-scale derived STR platform is now assisting in that recovery. We hope that the knowledge and lessons learned from this study will be applicable to not only the one-quarter of avian species that are threatened or near threatened, but to threatened species broadly.

## Materials and methods

### Sample collection

For *de novo* assembly, we extracted DNA samples from peripheral venous blood of a 3-year-old female crested ibis in the Yangxian County Reserve and a male little egret captured from the same county in southern Qinling Mountains, Shaanxi Province, China. For our resequencing effort, blood DNA samples were from eight crested ibis and five little egret individuals from the same locality. For meta-analysis of endangered and least concerned species, we used the genome sequences of 41 additional avian species (Additional file 1: Table S3) [16]. For DNA profiling, we used 105 individual crested ibis from four sub-populations of Yangxian, Huayang, Louguan, and Ningshan Counties (Additional file 1: Table S11).

### Karyotyping

Skin cells were grown in DMEM medium supplemented with 15% fetal bovine serum. Metaphase preparations



for flow sorting were generated [52]. The crested ibis chromosomes were numbered according to convention [53]. Chromosome preparations were stained with Hoechst 33258 (Sigma, St Louis, MO, USA) and Chromomycin A3 (Sigma) and then sorted [54] (MoFlo, DAKO, Glostrup, Denmark DAKO).

### Genome sequencing

For genome assembly, we constructed sequencing libraries with variable insert sizes (180 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb, and 20 kb) by following the manufacturer's instruction for Illumina's HiSeq 2000. Sequences of approximately 266 Gb and 127 Gb (reads length: 100 bp for libraries with insert size <2 kb, 45 bp for other libraries) were generated for the crested ibis and the little egret, respectively, and after quality-filtering, approximately 194 Gb (roughly 156×) and about 90 Gb (71×) sequences survived for the assembly. To achieve better contiguity, we also used a new physical mapping technology developed by the Argus System (OpGen) and its assembly software (Genome-Builder), which produced, based on KpnI digest, a large optical mapping dataset (about 282 Gb) from 34 high-density MapCards and containing 799,678 single-molecule restriction maps (>250 kb) with an average size of 353 kb.

### Genome assembly

The genome sequences for the crested ibis and little egret were assembled by using the de Bruijn graph-based assembler SOAPdenovo [17]. Prior to assembly, potential sequencing errors were removed or corrected based on k-mer frequency methodology. Reads from libraries with insert sizes ranging from 170 bp to 800 bp were split into 41-mers to construct de Bruijn graphs and contigs. The paired-end reads were aligned to construct scaffolds. Super-scaffolds for the crested ibis were constructed and aided with optical mapping data. The crested ibis chromosomes were built by using super-scaffolds based on conserved synteny between the assembly and genome data of chicken and zebra finch.

To assess the large-scale and local assembly accuracy, we also sequenced (Sanger sequencing technology) and assembled (phred-phrap-consed) eight randomly selected fosmids (average approximately 39 kb long) from a genomic library for the crested ibis (same DNA used for the reference assembly). We also assessed the completeness and accuracy of our assembly using 98,881 transcripts from blood, which were sequenced and assembled independently. A total of 94,709 assembled transcripts (>95%) were mapped to the assembly (BLASTN,  $E < 10^{-5}$ , coverage  $\geq 90\%$ ), yielding a single-base accuracy of approximately 98% for the assembled sequences with >20 reads coverage and excluding sequence gaps.

### Gene and repeat annotations

To predict genes, we used both homology-based and *de novo* methods as follows. First, we obtained protein sequences of chicken, zebra finch, and human from Ensembl (release 60) and mapped them onto the genome assemblies using Tblastn with E-value  $1e-5$ . All high-score segments were grouped into gene-like structures (genBlastA [55]). The homologous genome sequences with flanking sequences (2,000 bp) were aligned to the protein sequences by using Genewise [56] to define gene models. We clustered predicted transcripts >100 bp and took cross-species synteny into account (otherwise, a transcript with the best aligning score was used). Single-exon genes containing one or >1 frame shift errors and multi-exon genes containing >3 frame errors were not taken into account. Second, we clustered transcripts using TopHat [57] and Cufflinks [58] and aligned them (>150 bp) to SwissProt/TrEMBL database [59] with a cutoff E-value < $1e-5$ . Third, we predicted protein-coding genes (>150 bp) using Genscan [60] (gene model parameters trained with *Homo sapiens* genes) and Augustus [61] (gene model parameters trained with chicken genes) and defined TE-derived proteins (BlastP with E-value < $1e-5$  and >50% alignment).

For the reference gene set, we constructed gene models following three criteria: (1) candidate genes clustered with >100 bp overlap; (2) one cluster per gene (homology-based model > RNA-seq model > *de novo* predicted model); and (3) if not (2), 30% alignment to a known protein in the SwissProt/TrEMBL database [59] (>2 exons). Functional annotations were based on the best match principle using SwissProt, InterPro [62], and KEGG [63] databases. Treefam [64] was used to define gene family (Blastp, E-value < $1e-7$ ; Hscore >10; minimum edge density >1/3) and CAFE [65] to define gene loss and gain.

We annotated transposable elements (TEs) based on homology to RepBase sequences using RepeatProteinMask and RepeatMasker [66] with default parameters. We also constructed *de novo* repeat libraries (transposable elements) using RepeatModeler (repeatmasker.org/RepeatModeler.html) with default parameters.

### Resequencing data analysis

Resequencing reads were generated from a single-size insert library (500 bp) per individual and mapped high-quality reads onto the references with BWA [36], followed by removal of unmapped reads (average quality <10 or average map quality <20 or multiple-site reads). SNPs were called by using SOAPsnp [67] with thresholds of quality score  $\geq 20$ , sequencing depth >8X and <40X, copy number of flanking sequences  $\leq 2$ , >3 uniquely mapped reads, and distance between SNPs  $\geq 5$  bp.

We calculated the correlation coefficient ( $r^2$ ) of alleles at SNP locus after setting -maxdistance 300 -dprime -minGeno



0.6 -minMAF 0.1 -hwcutoff 0.001 using the Haploview software [68]. Since sample size is an important parameter influencing LD patterns, we randomly selected five crested ibises three times to repeat the experiment and the analysis. To reconstruct ancient demographic history, we ran the PSMC program (parameters: -N30, -t15, -r5, and -p '4 + 25\*2 + 4 + 6') using autosomal sequences (scaffold length  $\geq 50$  kb and a total of 478,758 heterozygous loci). We performed bootstrapping (100 times) to estimate the variance of simulated results and estimated the neutral mutation rate  $\mu$  (mutations per base per generation) using the estimated genome-wide nucleotide divergence (10.31%) and divergence time ( $38.98 \times 10^6$ ) between the crested ibis and the little egret. Based on mean generation time (3 years for crested ibis), we calculated  $\mu = (0.1031 \times 3) / (2 \times 38.98 \times 10^6) = 3.968 \times 10^{-9}$  mutations per generation for the crested ibis.

To reconstruct recent demographic history, we used the *daði* program [36] and paired-end reads (500 bp in size) from nine samples (eight re-sequencing individuals and one *de novo* assembly individual). To minimize the effect of low-coverage sequencing, we extracted the sites that were covered by high-quality reads (at least six of nine individuals covered by  $>2X$  reads). To prepare for *daði* program, we called 1,420,399 SNPs using a published method [69]. The little egret reference genome sequence was used to infer ancestral alleles. We considered four models and chose the one with highest maximum log-likelihood value. The ancestral population size ( $N_a$ ) was estimated on the basis of the calculated  $\theta$  value and the mutation rate. Population size and corresponding time were derived from parameters scaled based on  $N_a$ .

### Purifying selection analysis

For each 500-kb window, we determined the number of reads corresponding to the most and least abundant SNP alleles ( $n_{MAJ}$  and  $n_{MIN}$ ),  $H_p = 2 \sum n_{MAJ} \sum n_{MIN} / (\sum n_{MAJ} + \sum n_{MIN})^2$ , and transformed  $H_p$  into Z scores:  $ZH_p = (H_p - \mu H_p) / \sigma H_p$  [33]. We applied a threshold of  $ZH_p = -2.326$  ( $q < 0.01$  in normal distribution) for putative selective sweeps.

### Genome-wide STR profiling

We defined STRs using Tandem Repeat Finder [70] (parameters: Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 30, and MaxPeriod = 6), which were validated in the following steps. DNA was extracted with the E.Z.N.A.<sup>™</sup> Blood DNA Kit (Omega Bio-Tek Inc., USA) according to its instruction (E.Z.N.A.<sup>™</sup> Blood DNA Isolation Protocols, Revised June 2009). All DNA samples were quantified with the TIANamp Genomic DNA Kit. PCR amplification was performed in a reaction volume of 25  $\mu$ L with MicroAmp<sup>®</sup> reaction tubes (Applied Biosystems, CA, USA; the GeneAmp<sup>®</sup> PCR Systems 9700 with gold-plated silver or silver 96-well blocks). Amplified

products were separated in ABI3730 DNA Genetic Analyzer 48-capillary array system (Applied Biosystems) according to the manufacturer's protocol. The genotypes were analyzed by using Genemapper 3.5 (Applied Biosystems).

### Data availability

Genome data of crested ibis and little egret are uploaded to NCBI (PRJNA232572 and PRJNA232959). The raw reads in the SRA (SRP035852 and SRP035853). The NCBI accession numbers of the assembled genomes of all species are described in Additional file 1: Table S3.

### Additional files

**Additional file 1: Table S1.** The important events for crested ibis.

**Table S2.** Other evidence to estimate the number of crested ibis. **Table S3.** Summary of avian species and genome accessions. **Table S4.** More detailed summary of genome assemblies of the reference crested ibis and little egret genomes. **Table S5.** Fosmid sequencing and assembly. **Table S6.** Summary of transcriptome sequencing. **Table S7.** Transposable elements (TE) in the genomes of the crested ibis and the little egret. **Table S8.** Heterozygous SNPs in 40 avian species. **Table S9.** Resequencing summary of the eight crested ibis and the five little egret individual animals. **Table S10.** SNPs of the crested ibis and the little egret. **Table S11.** Sample information of 105 crested ibis individuals. **Table S12.** Details of genes in the windows of low  $ZH_p$  score in crested ibis genome. **Table S13.** Enrichment GO terms of genes in the windows of low  $ZH_p$  score in crested ibis genome. **Table S14.** Details of genes in the windows of low  $ZH_p$  score in crested ibis genome. **Table S15.** Results from model variations based on the *daði* calculations. **Table S16.** Pseudogenes in threatened and near threatened avian carnivorous species. **Table S17.** Pesticide and fertilizer usage number in China (Shaanxi and Gansu province). **Table S18.** STR summary of DIP (crested ibis DNA identification profiling). **Table S19.** Population parameters of STR loci used in DIP. **Table S20.** Paternity test of a crested ibis pedigree.

**Additional file 2: Figure S1.** Circular display of genomic information from the crested ibis. **Figure S2.** Box plot of the average heterozygosity in protein coding region of the genomes between LC ( $n = 32$ ) and EV ( $n = 8$ ) species ( $t$  test  $P = 0.005$ ). **Figure S3. (a)** Sampling sites. **(b)** Pairwise  $F_{st}$  among four sub-populations.  $F_{st}$ , fixation index, range 0-1. A value of 0 implies that the two populations are interbreeding freely, whereas a value of 1 implies that the two populations do not share any genetic diversity.

**Figure S4.** Demographic history. Estimation of the little egret based on the PSMC (pairwise sequentially Markov coalescent) model. The red line depicts the estimated effective population size ( $N_e$ ), and the thin gray curves represent PSMC bootstrapping estimates. The sky blue and yellow background colors indicate glacial and interglacial periods, respectively.

**Figure S5.** Genetic variation in the peptide binding region (red rectangle) of MHC class II  $\beta$  gene (*BLB*) of crested ibis (upper) and little egret (bottom). The consensus sequences of nine individuals are given at the top of each alignment. There are 12 loci and 18 loci maintaining  $\geq 2$  amino acid alleles for the nine crested ibis and the nine little egrets, respectively. There are one loci and six loci maintaining  $\geq 3$  amino acid alleles for the nine crested ibis and the nine little egrets, respectively. **Figure S6.** An example of STR genotyping. The PCR amplified products were separated by the ABI3730 Genetic Analyzer. Lanes 1 and 47 represent size markers and the remaining lanes depict PCR products of individual samples. Loci were labeled in multiple colors. **Figure S7.** Sex discriminating locus of crested ibis. Electrophoresis is in the top panel. The sequence of the sex marker is in the bottom panel. The green bases represent primer sequences. The indels in Z and W chromosomes were marked in red.

### Abbreviations

*BLB*: MHC class II  $\beta$  gene; *CHIA*: Chitinase, acidic; DIP: DNA identification profiling (DIP) platform; EN: Endangered; EV: Combined EN and VU; GABA:

$\gamma$ -aminobutyric acid; GO: Gene ontology; *HACL1*: 2-hydroxyacyl-CoA lyase 1; *H<sub>p</sub>*: Heterozygosity in 500-kb sliding windows;  $H_i/H_0$ : Heterozygosity at generation  $t$ /initial heterozygosity; IUCN: International Union for Conservation of Nature; KEGG: Kyoto Encyclopedia of Genes and Genomes; LC: Least concern; LD: Linkage disequilibrium; MHC: Major histocompatibility complexes; MYA: Million years ago; NS/S: Nonsynonymous/synonymous; PSMC: Pairwise sequential Markovian coalescent; RPKM: Reads per kilobase per million; *SLCO1A2*: Solute carrier organic anion transporter family member 1 A2; SNP: Single nucleotide polymorphisms; STR: Short-tandem repeat; VU: Vulnerable; *ZHp*: Z transformations of *H<sub>p</sub>*.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

SBL, BL, and CC conceived the study. JQY, JY, and HMY, managed the project. QBL, HBZ, and SGW prepared samples. BL, ZJX, SXZ, and GJZ performed genome sequencing, assembly, annotation, and transcriptomes analyses. SBL, JHL, QBL, CC, LC, HBZ, and BY carried out genetic analyses. BL, SBL, CC, and LC discussed the data. All authors contributed to data interpretation. SBL, BL, CC, EDJ, TG, and GJZ wrote the manuscript. EDJ, TG, GJZ also supervised and provided data in addition to those of the crested ibis and the little egret for comparative analysis from the avian phylogenomics project, and TG, GJZ, EDJ, and JY edited the manuscript. All authors read and approved the final manuscript.

#### Acknowledgments

We thank Laurie Goodman for help in editing the manuscript and Zhaozheng Guo for overseeing experimental bioethics. We thank Jing Liu, Hao Li, Liang Yang, Xiaojing Wei, Fang Cao, and Rong Su in Xi'an Jiaotong University for help in experiment. The Ibis and Egret sequencing and analyses part of this project was supported financially with funds from the state forestry administration of the People's Republic of China (No. 2012717), Higher Education Innovation Fund of Xi'an Jiaotong University, and Dr. Xi Guang Assistant to the President, Kwang-Hua Education Foundation (No. 2010822), and Dr. Yan Hong Assistant to the President, Dr. Chu Longfei and Dr. Ding Haihua in Shaanxi branch of State Forestry Administration, the Breeding and Releasing programs of crested ibis population (No. 2010, No. 2008), the Shenzhen Municipal Government of China (Nos. CXB201108250096A, JC201005260191A), ShenZhen Engineering Laboratory for Genomics-Assisted Animal Breeding, and China National GeneBank-Shenzhen. EDJ's contribution was supported by HHMI.

#### Author details

<sup>1</sup>Xi'an Jiaotong University, Xi'an 710049, P. R. China. <sup>2</sup>BGI-Shenzhen, Shenzhen 518083, P. R. China. <sup>3</sup>CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, P.R. China. <sup>4</sup>Center for Mitochondrial Biology and Medicine, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology and Frontier Institute of Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. <sup>5</sup>University of Wisconsin School of Veterinary Medicine, Madison, Wisconsin, USA. <sup>6</sup>BGI-Northwest, BGI-Shenzhen, Xi'an 710000, P. R. China. <sup>7</sup>Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, P. R. China. <sup>8</sup>James D. Watson Institute of Genome Science, Hangzhou, China. <sup>9</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark. <sup>10</sup>King Abdulaziz University, Jeddah, Saudi Arabia. <sup>11</sup>School of Biomedical Sciences, University of Queensland, Queensland, Australia. <sup>12</sup>University of Cincinnati College of Medicine, Cincinnati, Ohio, USA. <sup>13</sup>College of medicine, University of Chicago, Chicago, USA. <sup>14</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, ØsterVoldgade 5-7, 1350 Copenhagen, Denmark. <sup>15</sup>Department of Environment and Agriculture, Trace and Environmental DNA Laboratory, Curtin University, Perth, Western Australia 6845, Australia. <sup>16</sup>Prince Aljawhra Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>17</sup>Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA.

#### References

1. Attenborough D: *Help us build a brighter future: the need to save the world's most threatened birds*. Cambridge: BirdLife International; 2008.
2. Collar NJ, Andreev AV, Chan S: *Threatened Birds of Asia: The BirdLife International Red Data Book*. 5th edition. Cambridge: BirdLife International; 2001.
3. Brooke Mde L, Butchart SH, Garnett ST, Crowley GM, Mantilla-Beniers NB, Stattersfield AJ: **Rates of movement of threatened bird species International IUCN red list categories and toward extinction**. *Conserv Biol* 2008, **22**:417-427.
4. Frankham R: **Genetics and extinction**. *Biol Conserv* 2005, **126**:131-140.
5. Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I: **Inbreeding and extinction in a butterfly metapopulation**. *Nature* 1998, **392**:491-494.
6. Keller LF, Waller DM: **Inbreeding effects in wild populations**. *Trends Ecol Evol* 2002, **17**:230-241.
7. Lande R: **Mutation and conservation**. *Conserv Biol* 1995, **9**:782-791.
8. Woodworth L, Montgomery M, Briscoe D, Frankham R: **Rapid genetic deterioration in captive populations: causes and conservation implications**. *Conserv Genet* 2002, **3**:277-288.
9. Evans SR, Sheldon BC: **Interspecific patterns of genetic diversity in birds: correlations with extinction risk**. *Conserv Biol* 2008, **22**:1016-1025.
10. Spielman D, Brook BW, Frankham R: **Most species are not driven to extinction before genetic factors impact them**. *Proc Natl Acad Sci U S A* 2004, **101**:15261-15264.
11. Kohn MH, Murphy WJ, Ostrander EA, Wayne RK: **Genomics and conservation genetics**. *Trends Ecol Evol* 2006, **21**:629-637.
12. Liu YZ: **Rediscovery of crested ibis *Nipponia nippon* in Qinling Mountain**. *Chin J Zool* 1981, **27**:237.
13. Shi D, Cao Y: *Chinese Crested Ibis*. Beijing: China Forestry Press; 2001.
14. Ding C: *Research on the Crested Ibis*. Shanghai: Shanghai Scientific and Technological Educational Publishing House Press; 2004.
15. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, et al: **Using whole genomes to resolve the tree of life of modern birds**. *Science* 2014, **346**:1320-1331.
16. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, et al: **Comparative genomics across modern bird species reveal insights into avian genome evolution and adaptation**. *Science* 2014, **346**:1311-1320.
17. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J: **De novo assembly of human genomes with massively parallel short read sequencing**. *Genome Res* 2010, **20**:265-272.
18. Yarmolinsky DA, Zuker CS, Ryba NJ: **Common sense about taste: from mammals to insects**. *Cell* 2009, **139**:234-244.
19. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TA, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, et al: **The genome of a songbird**. *Nature* 2010, **464**:757-762.
20. International Chicken Genome Sequencing Consortium: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution**. *Nature* 2004, **432**:695-716.
21. Reed DH, Frankham R: **Correlation between fitness and genetic diversity**. *Conserv Biol* 2003, **17**:230-237.
22. Messaoudi I, Patino JAG, Dyall R, LeMaout J, Nikolich-Zugich J: **Direct link between mhc Polymorphism, T cell avidity, and diversity in immune defense**. *Science* 2002, **298**:1797-1800.
23. Fang X, Zhang Y, Zhang R, Yang L, Li M, Ye K, Guo X, Wang J, Su B: **Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque**. *Genome Biol* 2011, **12**:R63.
24. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, Yim SH, Zhao X, Kasaikina MV, Stoletzki N, Peng C, Polak P, Xiong Z, Kiezun A, Zhu Y, Chen Y, Kryukov GV, Zhang Q, Peshkin L, Yang L, Bronson RT, Buffenstein R, Wang B, Han C, Li Q, Chen L, et al:

Received: 14 October 2014 Accepted: 2 December 2014

Published online: 11 December 2014

- Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 2011, **479**:223–227.
25. Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W, Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder OA, Leung FC, Zhou Y, Cao J, Sun X, Fu Y, et al: **The sequence and de novo assembly of the giant panda genome.** *Nature* 2009, **463**:311–317.
  26. Gymrek M, Golan D, Rosset S, Erlich Y: **lobSTR: A short tandem repeat profiler for personal genomes.** *Genome Res* 2012, **22**:1154–1162.
  27. Crow JF, Kimura M: *An Introduction to Population Genetics Theory.* New York: Harper & Row; 1970.
  28. Reumers J, Schymkowitz J, Ferkinghoff-Borg J, Stricher F, Serrano L, Rousseau F: **SNPEffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs.** *Nucleic Acids Res* 2005, **33**:D527–D532.
  29. Mettler LE, Gregg TG: *Population genetics and evolution.* Englewood Cliffs (NJ): Prentice-Hall, Inc.; 1969.
  30. Biswas S, Akey JM: **Genomic insights into positive selection.** *Trends Genet* 2006, **22**:437–446.
  31. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blöcker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MC, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guérin G, et al: **Genome sequence, comparative analysis, and population genetics of the domestic horse.** *Science* 2009, **326**:865–867.
  32. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, Decaprio D, Gnerre S, et al: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**:803–819.
  33. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, Hallböök F, Besnier F, Carlborg O, Bed'hom B, Tixier-Boichard M, Jensen P, Siegel P, Lindblad-Toh K, Andersson L: **Whole-genome resequencing reveals loci under selection during chicken domestication.** *Nature* 2010, **464**:587–591.
  34. Lands B: **Consequences of essential fatty acids.** *Nutrients* 2012, **4**:1338–1357.
  35. Li H, Durbin R: **Inference of human population history from individual whole-genome sequences.** *Nature* 2011, **475**:493–496.
  36. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD: **Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data.** *PLoS Genet* 2009, **5**:e1000695.
  37. Zheng B, Xu Q, Shen Y: **The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation.** *Quaternary Int* 2002, **97–98**:93–101.
  38. Donaldson GM, Shutt JL, Hunter P: **Organochlorine contamination in bald eagle eggs and nestlings from the Canadian Great Lakes.** *Arch Environ Contam Toxicol* 1999, **36**:70–80.
  39. Helander B, Olsson A, Bignert A, Asplund L, Litzen K: **The role of DDE, PCB, coplanar PCB and eggshell parameters for reproduction in the white-tailed sea eagle (*Haliaeetus albicilla*) in Sweden.** *Ambio* 2002, **31**:386–403.
  40. Youl JM: **Lead exposure in free-ranging kea (*Nestor notabilis*), takahē (*Porphyrio hochstetteri*) and Australasian harriers (*Circus approximans*) in New Zealand.** Thesis. Palmerston North: Massey University; 2009.
  41. Shinsuke Tanabe AS, Hidaka H, Tatsukawa R: **Transfer rates and pattern of PCB isomers and congeners and p, p'-DDE from mother to egg in Adelie Penguin (*Pygoscelis adeliae*).** *Chemosphere* 1986, **15**:9.
  42. Corsolini S, Borghesi N, Ademollo N, Focardi S: **Chlorinated biphenyls and pesticides in migrating and resident seabirds from East and West Antarctica.** *Environ Int* 2011, **37**:1329–1335.
  43. Nocerla JJ, Blais JM, Beresford DV, Finity LK, Grooms C, Kimpe LE, Kyser K, Michelutti N, Reudink MW, Smol JP: **Historical pesticide applications coincided with an altered diet of aerially foraging insectivorous chimney swifts.** 2012, **279**:3114–3120.
  44. Albanis TA, Hela GD, Hatzilakos D: **Organochlorine residues in eggs of *Pelecanus crispus* and its prey in wetlands of Amvrakikos Gulf, North-western Greece.** *Chemosphere* 1995, **31**:9.
  45. Badagnani I, Castro RA, Taylor TR, Brett CM, Huang CC, Stryke D, Kawamoto M, Johns SJ, Ferrin TE, Carlson EJ, Burchard EG, Giacomini KM: **Interaction of Methotrexate with organic-anion transporting polypeptide 1A2 and its genetic variants.** *J Pharmacol Exp Ther* 2006, **318**:521–529.
  46. Foulon V, Antonenkov VD, Croes K, Waelkens E, Mannaerts GP, Van Veldhoven PP, Casteels M: **Purification, molecular cloning, and expression of 2-hydroxyphytanoyl-CoA lyase, a peroxisomal thiamine pyrophosphate-dependent enzyme that catalyzes the carbon-carbon bond cleavage during alpha-oxidation of 3-methyl-branched fatty acids.** *Proc Natl Acad Sci U S A* 1999, **96**:10039–10044.
  47. Maizels RM: **Infections and allergy - helminths, hygiene and host immune regulation.** *Curr Opin Immunol* 2005, **17**:656–661.
  48. Steiner CC, Putnam AS, Hoeck PEA, Ryder OA: **Conservation genomics of threatened animal species.** *Annu Rev Anim Biosci* 2013, **1**:261–281.
  49. Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, Zhu L, Li D, Zhang X, Chen Q, Zhang H, Zhang Z, Jin X, Zhang J, Yang H, Wang J, Wang J, Wei F: **Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation.** *Nat Genet* 2012, **45**:67–71.
  50. Cunningham SJ, Alley MR, Castro I, Potter MA, Cunningham M, Pyne MJ: **Bill morphology of ibises suggests a remote-tactile sensory system for prey detection.** *Auk* 2010, **127**:308–316.
  51. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C: **Genomic scans for selective sweeps using SNP data.** *Genome Res* 2005, **15**:1566–1575.
  52. Yang F, Carter NP, Shi L, Ferguson-Smith MA: **A comparative study of karyotypes of muntjacs by chromosome painting.** *Chromosoma* 1995, **103**:642–652.
  53. Ladjali-Mohammed K, Bitgood JJ, Tixier-Boichard M, Ponce De Leon FA: **International system for standardized avian karyotypes (ISSAK): standardized banded karyotypes of the domestic fowl (*Gallus domesticus*).** *Cytogenet Cell Genet* 1999, **86**:271–276.
  54. Ng BL, Carter NP: **Factors affecting flow karyotype resolution.** *Cytometry A* 2006, **69**:1028–1036.
  55. She R, Chu JS, Wang K, Pei J, Chen N: **GenBlastA: enabling BLAST to identify homologous gene sequences.** *Genome Res* 2009, **19**:143–149.
  56. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Res* 2004, **14**:988–995.
  57. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**:1105–1111.
  58. Roberts A, Pimentel H, Trapnell C, Pachter L: **Identification of novel transcripts in annotated genomes using RNA-Seq.** *Bioinformatics* 2011, **27**:2325–2329.
  59. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**:45–48.
  60. Salamov AA, Solovyev WV: **Ab initio gene finding in *Drosophila* genomic DNA.** *Genome Res* 2000, **10**:516–522.
  61. Stanke M, Waack S: **Gene prediction with a hidden Markov model and a new intron submodel.** *Bioinformatics* 2003, **19**:215–225.
  62. Mulder N, Apweiler R: **InterPro and InterProScan: tools for protein sequence classification and comparison.** *Methods Mol Biol* 2007, **396**:59–70.
  63. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27–30.
  64. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, Wong GK, Zheng W, Dehal P, Wang J, Durbin R: **TreeFam: a curated database of phylogenetic trees of animal gene families.** *Nucleic Acids Res* 2006, **34**:D572–D580.
  65. De Bie T, Cristianini N, Demuth JP, Hahn MW: **CAFE: a computational tool for the study of gene family evolution.** *Bioinformatics* 2006, **22**:1269–1271.
  66. Steinberg S, Chen L, Wei L, Moser A, Moser H, Cutting G, Braverman N: **The PEX Gene Screen: molecular diagnosis of peroxisome biogenesis disorders in the Zellweger syndrome spectrum.** *Mol Genet Metab* 2004, **83**:252–263.
  67. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K: **SNP detection for massively parallel whole-genome resequencing.** *Genome Res* 2009, **19**:1124–1132.
  68. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**:263–265.
  69. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Cornelissen TS, Zheng H, Liu T, He W, Li K, Luo R,

Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, *et al*: **Sequencing of 50 human exomes reveals adaptation to high altitude.** *Science* 2010, **329**:75–78.

70. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**:573–580.

doi:10.1186/s13059-014-0557-1

**Cite this article as:** Li *et al.*: Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology* 2014 **15**:557.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

