

RESEARCH

Open Access



# Genomic insight into the nocturnal adaptation of the black-crowned night heron (*Nycticorax nycticorax*)

Haoran Luo, Site Luo, Wenzhen Fang, Qingxian Lin, Xiaolin Chen\* and Xiaoping Zhou\*

## Abstract

**Background:** The black-crowned night heron (*Nycticorax nycticorax*) is an ardeid bird successfully adapted to the nocturnal environment. Previous studies had indicated that the eyes of the night herons have evolved several specialized morphological traits favoring nocturnal vision. However, the molecular mechanisms of the nocturnal vision adaptation of night herons remained inattentions. In this study, the whole genome of *N. nycticorax* was sequenced and comparative analyses were performed on the vision-related and olfactory receptor (OR) genes to understand the molecular mechanisms of the visual and olfactory adaptation of night herons.

**Results:** The results indicated that a number of vision genes were under positive or relaxed selection in *N. nycticorax*, whereas a number of other vision genes were under relaxed or intensified selection in the boat-billed heron (*Cochlearius cochlearius*), which suggested that the two species adapt to nocturnality with different genetic mechanisms. The different selections acting on vision genes are probably associated with the enlargement of eye size and the enhancement of visual sensitivity in night herons. The analyses on olfactory receptor (OR) genes indicated that the total number of OR genes in the genomes of *N. nycticorax* and *C. cochlearius* were about half those in the little egret (*Egretta garzetta*), whereas the diversity of their OR genes was not remarkably different. Additionally, the number of expressed OR genes in the transcriptomes of *N. nycticorax* was also fewer than that in *E. garzetta*. These results suggest a reduced olfactory capability in night herons compared with *E. garzetta*.

**Conclusions:** Our results provided evidence that several vision genes of the night herons were subjected to different natural selections, which can contribute to a better understanding of the genetic mechanisms of visual adaptations of the night heron. In addition, the finding of the reduced number of total and expressed OR genes in night herons may reflect a trade-off between olfaction and vision.

**Keywords:** Vision genes, Olfactory receptor genes, Nocturnal adaptation, Positive selection, Relaxed and intensified selection, Ardeidae birds

## Introduction

Ardeid birds are foragers that rely on visual cues for the detection and capture of prey [1–3]. Most species within this family are diurnal, and few species, namely night herons, are nocturnal [1]. The black-crowned night heron (*Nycticorax nycticorax*) is the most common and widely distributed night heron that forages mostly at night but also forages in daytime, especially during the breeding season [4, 5]. Compared with diurnal ardeid species, the

\*Correspondence: xlchen@xmu.edu.cn; xpzhou@xmu.edu.cn

Key Laboratory of Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen 361102, People's Republic of China



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

eyes of *N. nycticorax* have many anatomical characteristics similar to those of owls, such as relatively large size [6] and a tapetum lucidum behind the retina [7]. The rod-to-cone ratio of *N. nycticorax* has not yet been studied; however, many rod cells are found in the eyes of *N. nycticorax* [8], and high rod-to-cone ratio was reported in another night heron closely related to *N. nycticorax*, the yellow-crowned night heron (*Nyctanassa violacea*) [9]. Unlike owls, a large eye size does not give rise to larger binocular fields in *N. nycticorax* compared with diurnal ardeid species. The binocular field of *N. nycticorax* is vertically long and narrow with the bill, which was argued to be associated with their foraging technique rather than their nocturnal habits [6].

Recently, whole genomes have been sequenced for diverse nocturnal species from different animal groups, and the results have greatly advanced the understanding of the genetic mechanism underlying the sensorial adaptations of the nocturnal species. For instance, a comparative genomics analysis for the Chinese forest musk deer (*Moschus berezovskii*) indicated positively selected genes distributed in phototransduction and retinol metabolism pathways may be contributed to the nocturnality of this species [10]. Additionally, comparative genomic analyses of bats [11] and owls [12] revealed that adaptive evolution of the vision genes play important roles in the eye anatomical and physiological specializations favoring nocturnal vision. Other comparative genome analyses also provided evidence that some nocturnal species adapted to nocturnality by developing other sensorial modalities, such as olfaction and hearing. For example, the kiwi (*Apteryx mantelli*) genome has a remarkably high diversity of olfactory receptor genes, which proves that kiwis have well developed olfactory acuity [13, 14] and relies more on olfactory than other sensory systems for nocturnal foraging [15]. Furthermore, comparative analyses indicated that several evolutionary signatures associated with sensory adaptations to nocturnal environment are shared by at least two nocturnal bird groups of owls, chuck-will's-widow (*Caprimulgus carolinensis*), and brown kiwi (*A. australis*) [16].

To date, only the genomes of two ardeid species, the little egret (*Egretta garzetta*) and the boat-billed heron (*Cochlearius cochlearius*), have been sequenced. *E. garzetta* is a diurnal species, and its genome shows an expansion of the olfactory receptor (OR) gene repertoire [17]. *C. cochlearius* is a strictly nocturnal forager; however, whether it relies on vision or tactile techniques for foraging is in conflict [18, 19]. The published genome has not been used to explore the genetic basis underlying the nocturnal adaptations in this species [20].

In the present study, the whole genome of *N. nycticorax* was sequenced, and comparative analyses of *E. garzetta*,

**Table 1** Statistics of *N. nycticorax* genome sequencing

| Pair-end libraries | Total data (Gb) | Read length (bp) | Sequence coverage (X) |
|--------------------|-----------------|------------------|-----------------------|
| 230 bp             | 34.38           | 150              | 27.0                  |
| 500 bp             | 35.85           | 125              | 28.2                  |
| 2 k                | 24.23           | 125              | 19.0                  |
| 5 k                | 14.78           | 125              | 11.6                  |
| 10 k               | 14.70           | 125              | 11.5                  |
| Total              | 123.93          |                  | 97.4                  |

**Table 2** Assembly statistics for *N. nycticorax* genome

| Assembly  | N50 (bp)  | Number | Total Size (bp) |
|-----------|-----------|--------|-----------------|
| Contigs   | 58,550    | 46,539 | 1,116,931,249   |
| Scaffolds | 3,016,563 | 7,611  | 1,179,048,687   |

Contig numbers were contig after scaffolding. Details of the genome assembly statistics see Supplement Table S3

*C. cochlearius*, and other birds within Pelecaniformes with high-quality genomes were performed. Previous phylogenetic studies indicated *N. nycticorax* is more closely related to *E. garzetta* than to *C. cochlearius*, suggesting nocturnal behavior of *N. nycticorax* and *C. cochlearius* have evolved independently [21]. We tested the natural selections acting on the vision genes of *N. nycticorax* and *C. cochlearius*, with the aim to improve the understanding of the molecular vision adaptations to nocturnality in night herons. The number of OR genes between *N. nycticorax* and *E. garzetta* at the genomic and transcriptome levels were also compared. The total number of OR genes in a genome are believed to be positively correlated to the olfactory bulb ratio of a species and therefore can be a good indicator for the olfactory abilities of the species [17]. The OR gene expansion in *E. garzetta* genome suggests enhanced olfactory abilities [17]. *N. nycticorax* has an olfactory bulb ratio similar to that of *E. garzetta* [22]. Here, we investigated whether the OR gene expansion is exhibited in *N. nycticorax*.

## Results

### Genome assembly and annotation

A total of 123.93 Gb (~97.4-fold coverage) high-quality sequences were obtained from five paired-end and mate-pair libraries (Table 1). The final size of the assembled *N. nycticorax* genome is 1179.04 Mb in length, which cover about 93% of the 1272.61 Mb genome size estimated by the 17 K-mer distribution (Supplement Figure S1). The contig and scaffold N50 sizes in our assembly are 58.55 and 3016.563 kb, respectively (Table 2), and the longest scaffold is 17,752 kb. The Core Eukaryotic

Genes Mapping Approach (CEGMA) analysis indicated that 69.35% (172) complete and 12.5% (31) partial Core Eukaryotic Genes (CEGs) could be identified in our assembly. The Benchmarking Universal Single-Copy Orthologs (BUSCO) results showed that 93.0% complete (7724 single copies and 30 duplicated) and 2.2% (183) fragmented BUSCOs could be identified in our assembly, and 4.8% (401) were considered missing.

We observed that 8.39% of the whole assembly are repetitive sequences, which included 0.75% tandem repeat sequences, 0.66% DNA repeat elements, 5.15% long interspersed nuclear elements (LINE), 0.14% short interspersed nuclear elements (SINE), and 1.42% long terminal repeat elements (LTR). The functional annotation revealed that the assembly contains 13,361 predicted protein-coding genes. Among which, 13,358 (99.98%) predicted protein-coding genes were well annotated by the SwissProt, TrEMBL, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and InterPro databases. Additionally, 335 candidate microRNA genes and 183 candidate tRNAs with total lengths of 28,473 and 13,666 bp, respectively, were also identified.

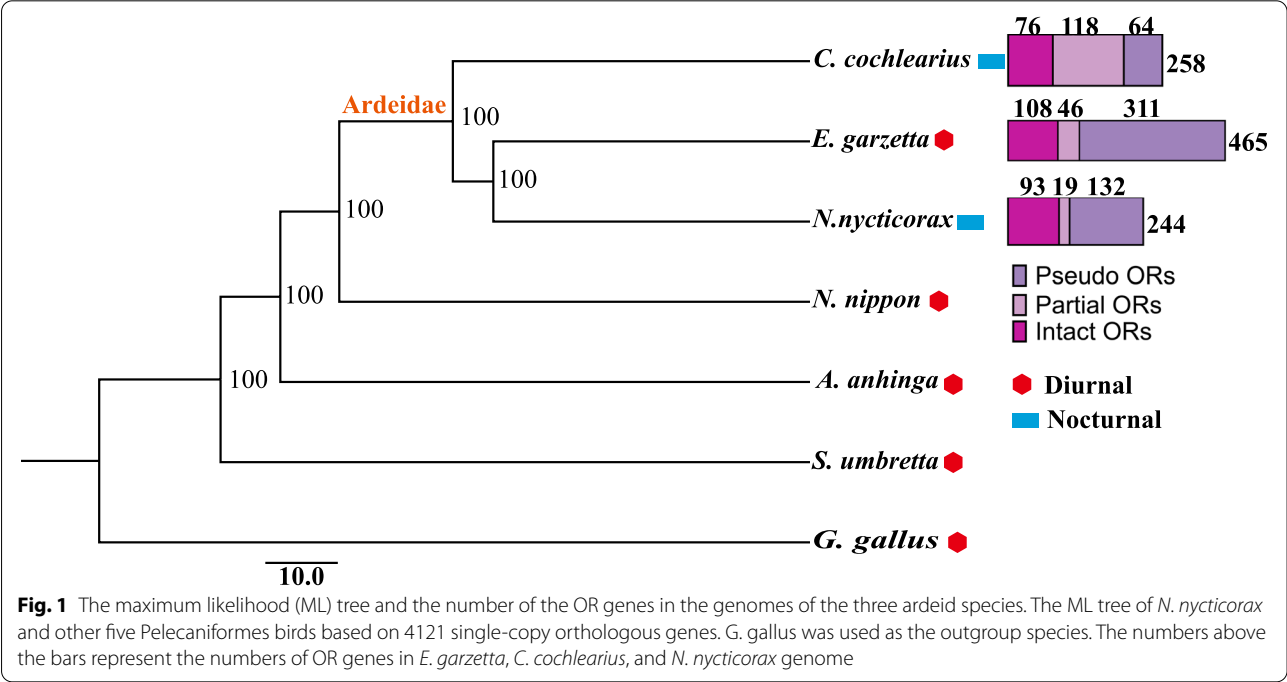
Selective analysis of vision-related genes

To provide a phylogenetic frame of reference for selection analyses, we firstly constructed a maximum likelihood (ML) tree using the IQTREE based on the protein coding sequences of 4121 single-copy orthologous gene from *N. nycticorax*, *C. cochlearius*, *E. garzetta*, *Nipponia*

*nippon*, *Anhinga anhinga*, *Scopus umbretta* and *Gallus gallus* (Fig. 1). Based on the ML tree, Phylogenetic Analysis by Maximum Likelihood (PAML) branch model, Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED), and RELAX were used to detect signatures of selection in each of the 216 orthologous vision genes (Supplement Table S2.xlsx), which were yielded by EggNog and OrthoFinder searches and had sequences more than 99 bp in length.

From the 216 orthologous genes, a total of 21 positively selected genes were identified in *N. nycticorax*, of which three genes (*CCDC66*, *CDON* and *RPL24*) were identified by both PAML branch model and BUSTED, 18 genes were identified only by BUSTED (*CHD7*, *CRYBA1*, *CTNS*, *FAT3*, *IMPG1*, *IMPG2*, *MDM1*, *MFN2*, *MITF*, *OLFM3*, *OPN4*, *RRM1*, *SH3PXD2B*, *SKI*, *SLC7A11*, *UNC119*, *VEGFA*, and *WDR19*). In *C. cochlearius* lineage, no positively selected gene was identified by PAML branch model and one positively selected gene (*MED1*) was identified by BUSTED.

Because high rate of false positive in detection of positive selection due to relaxed selection, previous studies suggested using a combination of approaches to distinguish between positive and relaxed selection, arguing that a gene with both intensified and positive selection signatures likely constitutes targets of truly positive selection, while a gene with both relaxed and positive selection signatures likely has experienced relaxation of selective constraint [23, 24]. Among the positively



selected genes mentioned above, 13 genes (*CCDC66*, *CDON*, *CHD7*, *CTNS*, *FAT3*, *IMPG2*, *MFN2*, *OPN4*, *RPL24*, *SH3PXD2B*, *SLC7A11*, *UNC119*, and *VEGFA*) in *N. nycticorax* were identified under intensified selection by the RELAX. Conversely, five genes (*CRYBA1*, *IMPG1*, *OLFM3*, *RRM1*, and *SKI*) in *N. nycticorax* and *MED1* in *C. cochlearius* were identified as under relaxed selection by both PAML branch model and RELAX.

In *N. nycticorax*, PAML branch model and RELAX both further identified *GNA11* under relaxed selection and *ATP8A2* under intensified selection, and respectively identified two (*GNAT1* and *PDGFRB*) and 15 genes (*ADAMTS18*, *EPAS1*, *FGFR2*, *FOXP2*, *GNAQ*, *HDAC1*, *HPS1*, *NHS*, *OPA1*, *PAX6*, *RPE65*, *SLC4A10*, *SMAD3*, *TTC8*, and *UCHL3*) under relaxed selection.

In *C. cochlearius*, RELAX further detected relaxed selection signatures in 13 additional genes (*ALDH1A3*, *CACNB2*, *CCDC66*, *EPAS1*, *FOXP2*, *JAG1*, *MAP3K1*, *OPA1*, *OPN4*, *PDE6B*, *SLC4A10*, *SLC7A11*, *SMAD3*, and *SOX14*). Additionally, five (*CCDC66*, *CRB1*, *PDGFRB*, *RNF2* and *TULP3*) and seven genes (*ADAMTS18*, *CRB1*, *EPHB1*, *MEGF11*, *MYO7A*, *SLC25A25*, and *WDR19*) were respectively identified as under intensified selection by PAML branch model and RELAX, of which *CRB1* was shared by PAML branch model and RELAX.

Comparing between *N. nycticorax* and *C. cochlearius*, five genes (*EPAS1*, *FOXP2*, *OPA1*, *SMAD3*, *SLC4A10*) overlapped with similar selection signatures, while six genes (*ADAMTS18*, *CCDC66*, *OPN4*, *PDGFRB*, *SLC7A11*, and *WDR19*) overlapped with different or opposite selection signatures. The details of results of different selective analyses and gene functions are presented in Tables 3 and 4, and supplement Table 4.

### Olfactory receptor genes analysis

The total number of OR genes identified from *N. nycticorax*, *C. cochlearius*, and *E. garzetta* were 244, 258, and 465, respectively. The proportions of intact, partial, and pseudo OR genes were 38.1% (93), 7.8% (19), and 54.1% (132) in *N. nycticorax*; 29.5% (76), 45.7% (118), and 24.8% (64) in *C. cochlearius*; and 23.2% (108), 9.9% (46), and 66.9% (311) in *E. garzetta* (Fig. 1). *E. garzetta* had the highest number of total and intact OR genes but the lowest proportion of intact OR genes among the three species owing to the large number of pseudogenes.

The intact OR genes in *N. nycticorax*, *C. cochlearius* and *E. garzetta* could be classified into 13, 7, and 10 subfamilies, respectively. As previously reported for other birds [17], the OR14 subfamily were the most abundant in all the three species,  $n=61$ ,  $n=65$ , and  $n=75$ , respectively (Table 5).

The average Shannon entropy ( $H$ ) values estimated from the Type I Class II ( $\gamma$ ) OR genes were  $0.492 \pm 0.414$

in *N. nycticorax*,  $0.499 \pm 0.413$  in *C. cochlearius*, and  $0.489 \pm 0.430$  in *E. garzetta*. The Wilcoxon signed-rank test indicated that the two night herons had no significantly higher  $H$  values than *E. garzetta* ( $P=0.490$  in *N. nycticorax* vs. *E. garzetta*,  $P=0.376$  in *C. cochlearius* vs. *N. nycticorax*).

To test positive selection in the OR14 subfamily, Genetic Algorithm Recombination Detection (GARD) was firstly used to detect the presence of recombination in the OR14 genes. The results revealed two breakpoints at the alignment nucleotide positions 509 and 719 in *E. garzetta*; five breakpoints at positions 207, 331, 464, 573, and 642 in *N. nycticorax*; and three breakpoints at positions 369, 507, and 631 in *C. cochlearius*. Based on the inferred recombination breakpoints, Single-likelihood Ancestor Counting (SLAC), Mixed Effects Model of Evolution (MEME), and Fast Unconstrained Bayesian Approximation (FUBAR) were used to infer signatures of positive selection and all three analytical methods identified positive selection in amino acid positions 52 (near transmembrane domain (TM) 2), 107 (in TM3), 196 (closest site near TM5), 204 (in TM5), 251 (in TM6), and 283 (in TM7) in *E. garzetta* (Supplement Table S5). Additionally, positive selection was identified in amino acid positions 93 (the closest site near TM3) and 110 (in TM3) in *N. nycticorax* (Supplement Table S6) and in amino acid positions 16, 47, 107 (in TM3), 172, 196 (closest site near TM5), 204 (in TM5), 207 (in TM5), and 283 (in TM7) in *C. cochlearius* (Supplement Table S7) by all three analytical methods. These results indicated that *E. garzetta* and *C. cochlearius* share four positive selection positions, and the other positions were species specific.

The transcriptomes of the olfactory epitheliums (OEs) of *N. nycticorax* and *E. garzetta* were studied to confirm that the identified intact OR genes were actually expressed. The results indicated that 61 genes (61 in nestlings and 30 in adults) had detectable expression in *N. nycticorax*, and 132 OR genes (113 in nestlings and 118 in adults) had detectable expression in *E. garzetta*. Notably, the nestling *N. nycticorax* expressed more OR genes than the adults, where 31 of the 61 expressed genes had no detectable expression in adult OEs. In *E. garzetta*, 14 expressed OR genes had no detectable expression in nestling OEs, and 19 had no detectable expression in adult OEs. The expression levels of ORs were quite low in both species, wherein the average expression levels ranged from 0.161 transcripts per kilobase million (TPM) in the nestling *E. garzetta* OE to 0.566 TPM in the nestling *N. nycticorax* OE (Fig. 2, Supplement Fig. 2, and Supplement Table S3.xlsx).

**Table 3** The results of selective analyses of the vision genes in *N. nycticorax*

| Gene     | PAML- $\omega$ 0 | PAML- $\omega$ 1 | PAML-P           | PAML-Q           | BUSTED-P         | BUSTED-Q         | RELAX-K      | RELAX-P            | RELAX-Q          |
|----------|------------------|------------------|------------------|------------------|------------------|------------------|--------------|--------------------|------------------|
| ADAMTS18 | 0.2385           | 0.2385           | 5.158E-01        | 8.936E-01        | 0.5              | 0.46323          | 0            | 0.005              | 0.0231           |
| ATP8A2   | <b>0.2646</b>    | <b>0.0001</b>    | <b>9.826E-05</b> | <b>3.301E-03</b> | 0.5              | 0.50859          | <b>1.17</b>  | <b>0.002</b>       | <b>0.0136</b>    |
| CCDC66   | <b>0.3254</b>    | <b>2.5296</b>    | <b>1.639E-06</b> | <b>5.679E-05</b> | <b>1.054E-18</b> | <b>6.646E-17</b> | <b>3.86</b>  | <b>6.22154E-14</b> | <b>1.730E-12</b> |
| CDON     | <b>0.2713</b>    | <b>1.0674</b>    | <b>0</b>         | <b>0</b>         | <b>4.277E-18</b> | <b>1.796E-16</b> | <b>10.54</b> | <b>2.398E-26</b>   | <b>1.335E-24</b> |
| CHD7     | 0.0817           | 0.0817           | 9.238E-01        | 9.900E-01        | <b>1.837E-20</b> | <b>2.315E-18</b> | <b>35.59</b> | <b>0.0001</b>      | <b>0.0015</b>    |
| CRYBA1   | <b>0.0477</b>    | <b>0.2343</b>    | <b>2.176E-04</b> | <b>4.310E-03</b> | <b>0.0001</b>    | <b>0.0013</b>    | <b>0.32</b>  | <b>3.122E-06</b>   | <b>3E-05</b>     |
| CTNS     | 0.4024           | 0.4024           | 6.838E-02        | 3.833E-01        | <b>0.0011</b>    | <b>0.0108</b>    | <b>10.39</b> | <b>0</b>           | <b>0</b>         |
| EPAS1    | 0.2276           | 0.2276           | 8.291E-01        | 1                | 0.5              | 0.4295           | <b>0.11</b>  | <b>0.002</b>       | <b>0.008</b>     |
| FAT3     | 0.0913           | 0.0913           | 8.275E-02        | 3.955E-01        | <b>1.030E-09</b> | <b>2.165E-08</b> | <b>50</b>    | <b>1.99E-11</b>    | <b>2.768E-10</b> |
| FGFR2    | 0.0738           | 0.0738           | 3.157E-01        | 6.267E-01        | 0.5              | 0.4632           | <b>0.44</b>  | <b>0.003</b>       | <b>0.0151</b>    |
| FOXP2    | 0.1829           | 0.1829           | 1.735E-01        | 4.966E-01        | 0.5              | 0.4632           | <b>0</b>     | <b>0.011</b>       | <b>0.0394</b>    |
| GNA11    | <b>0.0001</b>    | <b>0.3321</b>    | <b>6.565E-04</b> | <b>1.137E-02</b> | 0.5              | 0.4632           | <b>0.15</b>  | <b>0.008</b>       | <b>0.0318</b>    |
| GNAQ     | 0.1753           | 0.1753           | 4.591E-02        | 2.204E-01        | 0.5              | 0.5085           | <b>0</b>     | <b>0.004</b>       | <b>0.0218</b>    |
| GNAT1    | <b>0.0217</b>    | <b>0.1128</b>    | <b>5.745E-04</b> | <b>1.025E-02</b> | 0.5              | 0.525            | 0.89         | 0.648              | 1                |
| HDAC1    | 0.0411           | 0.0411           | 1.530E-01        | 4.903E-01        | 0.5              | 0.4632           | <b>0</b>     | <b>0.017</b>       | <b>0.05</b>      |
| HPS1     | 0.1729           | 0.1729           | 9.458E-01        | 9.900E-01        | 0.019            | 0.1408           | <b>0.61</b>  | <b>0.01</b>        | <b>0.0371</b>    |
| IMPG1    | <b>0.2561</b>    | <b>0.6748</b>    | <b>2.942E-04</b> | <b>4.943E-03</b> | <b>8.481E-12</b> | <b>2.760E-10</b> | <b>0.5</b>   | <b>8.008E-13</b>   | <b>2.186E-11</b> |
| IMPG2    | 0.3791           | 0.3791           | 7.494E-01        | 1                | <b>2.423E-11</b> | <b>2.290E-10</b> | <b>39.63</b> | <b>0.013</b>       | <b>0.0464</b>    |
| MDM1     | 0.3542           | 0.3542           | 9.218E-01        | 9.900E-01        | <b>0.002</b>     | <b>0.018</b>     | 1.18         | 0.476              | 0.6307           |
| MFN2     | 0.0505           | 0.0505           | 2.719E-01        | 5.703E-01        | <b>8.329E-05</b> | <b>0.0011</b>    | <b>5.49</b>  | <b>6.713E-05</b>   | <b>0.0006</b>    |
| MITF     | 0.0183           | 0.0183           | 7.086E-01        | 9.484E-01        | <b>4.630E-10</b> | <b>1.167E-08</b> | 7.68         | 0.761              | 0.8066           |
| NHS      | 0.1380           | 0.1380           | 1.199E-01        | 4.903E-01        | 0.5              | 0.4632           | <b>0.09</b>  | <b>0.001</b>       | <b>0.0055</b>    |
| OLFM3    | <b>0.1720</b>    | <b>0.9096</b>    | <b>4.110E-07</b> | <b>9E-06</b>     | <b>3.985E-10</b> | <b>2.510E-09</b> | <b>0</b>     | <b>1.894E-12</b>   | <b>3.373E-11</b> |
| OPA1     | 0.1855           | 0.1855           | 6.913E-02        | 3.833E-01        | 0.031            | 0.2055           | <b>0</b>     | <b>0.001</b>       | <b>0.0055</b>    |
| OPN4     | 0.0361           | 0.0361           | 5.956E-02        | 3.589E-01        | <b>0.0001</b>    | <b>0.0014</b>    | <b>13.37</b> | <b>9.919E-13</b>   | <b>2.208E-11</b> |
| PAX6     | 0.1595           | 0.1595           | 9.914E-01        | 9.900E-01        | 0.5              | 0.4632           | <b>0.15</b>  | <b>0.007</b>       | <b>0.0288</b>    |
| PDGFRB   | <b>0.0564</b>    | <b>0.4014</b>    | <b>1.792E-04</b> | <b>4.139E-03</b> | 0.5              | 0.4632           | 5.54         | 0.853              | 0.8183           |
| RPE65    | 0.1523           | 0.1523           | 1.923E-01        | 5.030E-01        | 0.5              | 0.4632           | <b>0.09</b>  | <b>0.001</b>       | <b>0.0055</b>    |
| RPL24    | <b>0.0089</b>    | <b>1.4164</b>    | <b>0</b>         | <b>0</b>         | <b>1.963E-07</b> | <b>3E-06</b>     | <b>29.42</b> | <b>6.06E-12</b>    | <b>9.635E-11</b> |
| RRM1     | <b>0.0198</b>    | <b>0.6076</b>    | <b>8.728E-06</b> | <b>2.419E-04</b> | <b>0.0024</b>    | <b>0.0209</b>    | <b>0.07</b>  | <b>0.002</b>       | <b>0.0106</b>    |
| SH3PXD2B | 0.1368           | 0.1368           | 5.E-02           | 3.300E-01        | <b>1.790E-10</b> | <b>5.641E-09</b> | <b>17.08</b> | <b>1.529E-12</b>   | <b>2.838E-11</b> |
| SKI      | <b>0.0631</b>    | <b>0.6871</b>    | <b>0</b>         | <b>0</b>         | <b>8.824E-08</b> | <b>1.5E-06</b>   | <b>0.2</b>   | <b>3.898E-23</b>   | <b>1.446E-21</b> |
| SLC4A10  | 0.1424           | 0.1424           | 3.260E-01        | 6.365E-01        | 0.417            | 0.4632           | <b>0</b>     | <b>9.367E-05</b>   | <b>0.0008</b>    |
| SLC7A11  | 0.2571           | 0.2571           | 9.602E-01        | 1                | <b>7.536E-05</b> | <b>0.0003</b>    | <b>2.37</b>  | <b>0.001</b>       | <b>0.0059</b>    |
| SMAD3    | 0.0281           | 0.0281           | 7.100E-01        | 9.484E-01        | 0.5              | 0.4632           | <b>0</b>     | <b>0.0002</b>      | <b>0.0021</b>    |
| TTC8     | 0.2426           | 0.2426           | 2.313E-02        | 1.554E-01        | 0.5              | 0.5085           | <b>0</b>     | <b>0.001</b>       | <b>0.0091</b>    |
| UCHL3    | 0.1590           | 0.1590           | 1.460E-01        | 3.774E-01        | 0.5              | 0.5085           | <b>0</b>     | <b>0.007</b>       | <b>0.0318</b>    |
| UNC119   | 0.0264           | 0.0264           | 4.437E-03        | 6.833E-02        | <b>0.001</b>     | <b>0.0105</b>    | <b>50</b>    | <b>0.006</b>       | <b>0.0256</b>    |
| VEGFA    | 0.6017           | 0.6017           | 8.976E-01        | 9.9E-01          | <b>0.003</b>     | <b>0.0236</b>    | <b>46.91</b> | <b>3.583E-05</b>   | <b>0.0003</b>    |
| WDR19    | 0.1118           | 0.1118           | 4.362E-01        | 1                | <b>5.819E-21</b> | <b>1.099E-19</b> | 0.97         | 0.889              | 0.85             |

Significant results were highlighted in bold

## Discussion

This study presents the first genome assembly of *N. nycticorax* and attempted to uncover the potential molecular basis that underlies the adaptation of night herons' visual and olfactory systems to nocturnality. Consistent with the expectation that night herons would show signatures of adaptive evolution associated with their visual

systems, we detected different selective pressures acting on a number of different vision genes in *N. nycticorax* and *C. cochlearius*. Additionally, we found relatively higher number of genes with positive selection and relaxed selection in *N. nycticorax*, whereas relatively higher number of genes with intensified selection in *C. cochlearius*. These results indicated different evolutionary mechanisms



**Table 4** The results of the selective analyses of the vision genes in *C. cochlearius*

| Gene     | PAML- $\omega$ 0 | PAML- $\omega$ 1 | PAML-P            | PAML-Q            | BUSTED-P  | BUSTED-Q        | RELAX-K      | RELAX-P          | RELAX-Q          |
|----------|------------------|------------------|-------------------|-------------------|-----------|-----------------|--------------|------------------|------------------|
| ADAMTS18 | 0.2385           | 0.2385           | 4.639E-03         | 8.160E-02         | 0.5       | 5.21E-01        | <b>1.16</b>  | <b>0.002</b>     | <b>0.0184</b>    |
| ALDH1A3  | 0.0315           | 0.0315           | 8.359E-02         | 4.524E-01         | 0.002     | 8.39E-02        | <b>0.32</b>  | <b>0.006</b>     | <b>0.0378</b>    |
| CACNB2   | 0.1851           | 0.1851           | 1.489E-01         | 5.051E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>0.001</b>     | <b>0.0108</b>    |
| CCDC66   | <b>0.5134</b>    | <b>0.0175</b>    | <b>2.815E-06</b>  | <b>9.902E-05</b>  | 0.5       | 5.21E-01        | 1            | 1                | 0.8932           |
| CRB1     | <b>0.1249</b>    | <b>0.0212</b>    | <b>3.556E-07</b>  | <b>1.668E-05</b>  | 0.0035    | 8.39E-02        | <b>2.51</b>  | <b>5.349E-05</b> | <b>0.0011</b>    |
| EPAS1    | 0.2276           | 0.2276           | 1.937E-01         | 6.102E-01         | 0.5       | 5.00E-01        | <b>0</b>     | <b>0.001</b>     | <b>0.0073</b>    |
| EPHB1    | 0.0485           | 0.0485           | 2.102E-02         | 2.201E-01         | 0.5       | 5.21E-01        | <b>15.9</b>  | <b>1.384E-05</b> | <b>0.0003</b>    |
| FOXP2    | 0.1829           | 0.1829           | 9.362E-02         | 4.841E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>0.005</b>     | <b>0.0374</b>    |
| JAG1     | 0.0328           | 0.0328           | 1.439E-01         | 5.051E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>0.007</b>     | <b>0.0418</b>    |
| MAP3K1   | 0.2610           | 0.2610           | 5.521E-02         | 3.237E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>0.001</b>     | <b>0.0108</b>    |
| MED1     | <b>0.0319</b>    | <b>0.1132</b>    | <b>1.036E-03</b>  | <b>2.429E-02</b>  | 6.801E-05 | <b>9.64E-03</b> | <b>0.64</b>  | <b>0.0004</b>    | <b>0.0062</b>    |
| MEGF11   | 0.1747           | 0.1747           | 2.070E-02         | 2.201E-01         | 0.5       | 5.21E-01        | <b>46.41</b> | <b>0.002</b>     | <b>0.0184</b>    |
| MYO7A    | 0.0001           | 0.0001           | 1                 | 1                 | 0.5       | 5.00E-01        | <b>15.7</b>  | <b>0.006</b>     | <b>0.0176</b>    |
| OPA1     | 0.1855           | 0.1855           | 3.263E-02         | 3.061E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>7.505E-06</b> | <b>0.0002</b>    |
| OPN4     | 0.0361           | 0.0361           | 1.507E-01         | 5.051E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>5.853E-07</b> | <b>3.503E-05</b> |
| PDGFRB   | <b>0.0754</b>    | <b>0.0077</b>    | <b>4.477E-04</b>  | <b>1.260E-02</b>  | 0.5       | 5.21E-01        | 1.12         | 0.594            | 0.7279           |
| RNF2     | <b>0.0442</b>    | <b>0.0203</b>    | <b>0</b>          | <b>0</b>          | 0.5       | 5.21E-01        | 1.21         | 0.88             | 0.8917           |
| SLC25A25 | 0.0099           | 0.0099           | 4.839E-01         | 7.612E-01         | 0.003     | 8.39E-02        | <b>18.88</b> | <b>7.371E-10</b> | <b>8.822E-08</b> |
| SLC4A10  | 0.1424           | 0.1424           | 3.054E-01         | 6.319E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>8.979E-05</b> | <b>0.0015</b>    |
| SLC7A11  | 0.2571           | 0.2571           | 4.125E-03         | 9.096E-02         | 0.5       | 5.00E-01        | <b>0</b>     | <b>0.002</b>     | <b>0.0098</b>    |
| SMAD3    | 0.0281           | 0.0281           | 5.467E-01         | 8.013E-01         | 0.5       | 5.21E-01        | <b>0</b>     | <b>0.0003</b>    | <b>0.0054</b>    |
| SOX14    | 0.0151           | 0.0151           | 9.751E-01         | 1                 | 0.5       | 5.21E-01        | <b>0</b>     | <b>7.505E-06</b> | <b>0.0002</b>    |
| TULP3    | <b>0.2277</b>    | <b>0.1016</b>    | <b>7.272E-230</b> | <b>5.116E-228</b> | 0.02      | 4.05E-01        | 1.58         | 0.131            | 0.3038           |
| WDR19    | 0.1118           | 0.1118           | 4.995E-02         | 3.671E-01         | 0.5       | 5.00E-01        | <b>2.3</b>   | <b>0.003</b>     | <b>0.0110</b>    |

Significant results were highlighted in bold

of visual adaptation between *N. nycticorax* and *C. cochlearius*, which were possibly due to that the nocturnalities of *N. nycticorax* and *C. cochlearius* are of independent origin, and *N. nycticorax* and *C. cochlearius* have different diel activity patterns and degrees of dependence on vision [4, 19, 21].

Adaptive evolution of the vision genes facilitated by positive selection was widely reported in many other animals which rely on vision in dim-light environments, such as mole rats [25], Chinese forest musk deer [10], okapi [26], bats [11], and owls [27], although the positively selected genes varied cross species. Similar to these findings, 16 vision genes were found to undergo positive selection in *N. nycticorax*, suggesting that positive selection has played an important role in the adaptive evolution of the vision genes and thereby contributed to the morphological or functional modifications of the eyes favoring night vision. Specifically, *CDON* and *MITF* are associated with eye size, deleterious mutations of which have been correlated with microphthalmia [28, 29]. *CTNS*, *MFN2*, and *SH3PXD2B* are involved in the development and maintenance of cornea and lens [30–34]. Positive selections in these genes may be responsible for

the enlarged eye size, cornea, and lens of *N. nycticorax*, which allow more light to enter the eyes in dim conditions [35, 36]. The remaining positively selected genes are all retina-associated, involved in the development and maintenance of retina (*FAT3*, *OPN4*, *SLC7A11*, *VEGFA*) [37–41] and photoreceptor (*CHD7*, *CCDC66*, *IPMG2*, *WDR19*) [42–45]. Positive selections in these genes likely play important roles in the retinal modifications, such as the preponderance of rods over cones [9].

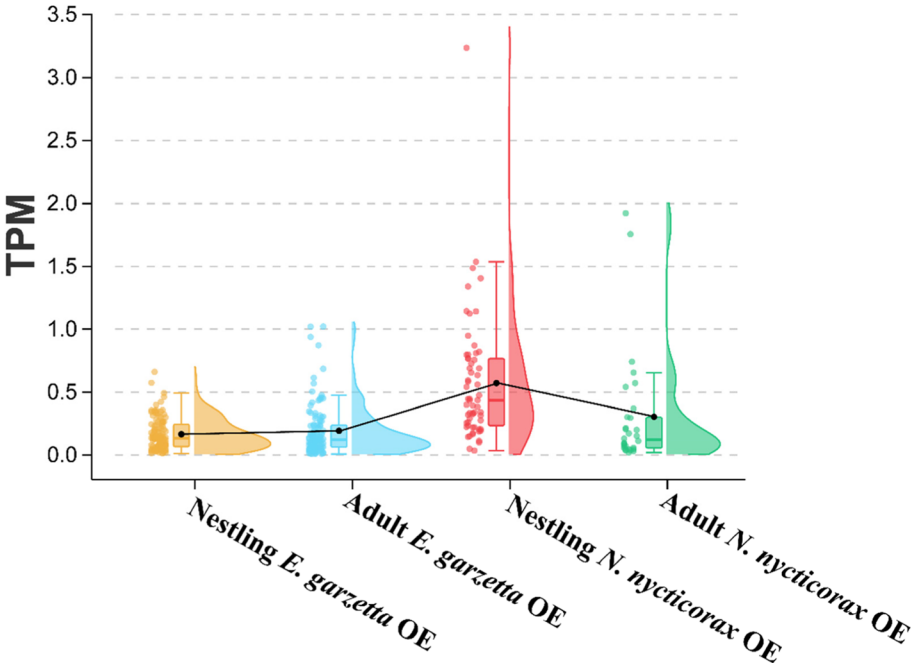
In addition to positive selection, 23 genes in *N. nycticorax* and 14 genes in *C. cochlearius* were identified as under relaxed selection, suggesting that relaxed selection is the predominant force that shaped the evolution of visual adaptation in both species. For example, *GNAT1* in *N. nycticorax* and *PDE6B* in *C. cochlearius* are involved in the rod phototransduction pathway [46, 47], which is responsible for dim light vision [27, 48]. Relaxed selection of these two genes may have allowed them to explore new functions increasing visual sensitivity in *N. nycticorax* and *C. cochlearius*, respectively. The other relaxed selection genes, such as the five genes shared by *N. nycticorax* and *C. cochlearius*, are associated with morphologies or functions of (*FOXP2*) [49], lens (*SMAD3*) [50],

**Table 5** Classes of Intact ORs

| Classes             | Subfamily | <i>N. nycticorax</i> | <i>C. cochlearius</i> | <i>E. garzetta</i> |
|---------------------|-----------|----------------------|-----------------------|--------------------|
| Type I-class II (γ) | 1         |                      |                       |                    |
|                     | 2         | 3                    | 1                     | 3                  |
|                     | 4         | 6                    | 1                     | 6                  |
|                     | 5         | 5                    | 5                     | 6                  |
|                     | 6         | 4                    | 1                     | 3                  |
|                     | 7         |                      |                       |                    |
|                     | 8         | 1                    |                       |                    |
|                     | 9         |                      |                       |                    |
|                     | 10        | 1                    | 1                     | 3                  |
|                     | 11        | 2                    |                       | 2                  |
|                     | 13        |                      |                       |                    |
|                     | 14        | 61                   | 65                    | 75                 |
|                     | total     | 83                   | 74                    | 98                 |
| Type I-class I      | 51        | 2                    |                       | 1                  |
|                     | 52        |                      | 2                     | 7                  |
|                     | 55        | 4                    |                       |                    |
|                     | 56        |                      |                       |                    |
|                     | total α   |                      |                       | 8                  |
|                     | δ         | 6                    |                       |                    |
|                     | ε         | 1                    |                       |                    |
| Type II             | ζ         |                      |                       |                    |
|                     | η         | 1                    |                       | 2                  |
|                     | θ         | 2                    |                       |                    |

cornea (*SLC4A10*) [51], retina (*EPAS1*) [52], and optic nerve (*OPA1*) [53, 54], respectively. Relaxed selection of these genes may contribute to adjust the morphologies or functions of different eye components to enhance night vision as the finding in a previous study of barn owl (*Tyto alba*) [35]. In the barn owl, the unique phenotypes of the scotopic-adapted eyes are hypothesized to have evolved through relaxed or intensified selection in genes with predominant roles in different eye elements [35]. Additionally, the authors suggested that some genes have coevolved with contrasting selective signature due to adaptive compensation [35]. *N. nycticorax* and *C. cochlearius* separately have one and 11 genes with signatures of intensified selection, which may also be explained by the same evolutionary mechanism.

Turning to the olfactory receptor gene repertoires, the total number and genetic diversity of ORgenes are commonly used as proxy for olfactory capability [55–57]. Fewer number of OR genes in the genomes of *N. nycticorax* and *C. cochlearius* along with fewer number of expressed OR genes in the transcriptomes of *N. nycticorax* reflected reduced olfactory capabilities in night herons compared with *E. garzetta*. Night heron reduced olfactory capabilities may be a trade-off with their large eyes [58–60], because large eyes are energetically costly to develop and maintain [61]. An animal with increased investment in vision modality



**Fig. 2** Raincloud plot of ORs expression levels. Each plot represents the expression levels of ORs. Quartiles are shown in the error bars, and the cloud showed the expression distributions. Black dots in the error bars represent the mean values of each group, black line was used to connection mean values

could decrease investment in other sensory modalities, including olfaction [62].

The transcriptome results showed that the number of expressed OR genes in *E. garzetta* is higher than the number of intact genes annotated in the genome. This finding reflected that some functional genes were identified as partial OR genes because of the genome assembly quality or the expression of some pseudogenes. The transcription of OR pseudogenes in the OE was also reported in humans [63] and other animals [64, 65]. Conversely, the number of expressed genes in *N. nycticorax* is fewer than the number of intact genes annotated in the genome, revealing that a part of the OR gene is transcriptionally inactive [66]. Additionally, some OR genes in *N. nycticorax* and *E. garzetta* are specifically expressed in nestling or adult samples, which suggests that these OR genes play different roles in specific developmental stages. Similar findings were also reported in other animals, such as the Chinese perch (*Siniperca chuatsi*) [65] and the Leach's storm-petrel (*Oceanodroma leucorhoa*) [67].

The expression of OR genes in the OE of *N. nycticorax* and *E. garzetta* indicated that they play olfactory roles despite of their low expression levels. The study of Leach's storm-petrel [67], a highly olfactory forager, also indicated the low expression levels of OR genes in the OE. Therefore, the low expression levels of OR genes may be common in birds. Additionally, the positively selective sites exhibited in the expanded OR14 subfamily in the night herons and *E. garzetta* further supported that OR genes are ecologically functional, consistent with the findings in other birds [16].

## Conclusions

This study provides the first analysis of the nocturnal adaptation of night herons from vision-related and OR genes. The different selective signatures detected in the vision genes provided evidence that the visual systems of the night herons have undergone adaptive evolution for enhanced night vision. Additionally, adaptive evolution of the vision genes in *N. nycticorax* is predominantly driven by positive and relaxed selections, while that in *C. cochlearius* is predominantly driven by relaxed and intensified selections. The comparative analyses of OR gene repertoires in the genomes and transcriptome between night herons and *E. garzetta* indicated reduced olfactory capacity in night herons, but olfactory capacity of night herons still has ecological adaptation. Future studies on olfactory function in different life activities in night herons and other diurnal egrets would provide a better understanding of the roles of OR genes in ardeid birds.

## Materials and methods

### Sample collection

*N. nycticorax* and *E. garzetta* were captured in Xiamen, Fujian, China. Each species includes three nestlings and three adults. To reduce potential kinships between samples, each nestling was sampled from different nest and adult samples were collected from a location different from the nestlings. The body lengths of the nestlings of each species were less than 15 cm. The adults were identified by their breeding plumage. Muscle sample was collected from one *N. nycticorax* and stored at  $-80^{\circ}\text{C}$  until DNA extraction for whole-genome sequencing. Olfactory epithelium (OE) tissues were collected from each individual and stored in RNeasy Lysis Buffer (Qiagen, Venlo, The Netherlands) at  $4^{\circ}\text{C}$  until RNA extraction.

### DNA and RNA extraction and sequencing

Genomic DNA was extracted using QIAGEN® Puregene Tissue Core Kit A according to the manufacturer's instructions (Qiagen, Beijing, China). Two short insert libraries (230 and 500 bp) were constructed using the Illumina TruSeq DNA Library Preparation Kit (Illumina, San Diego, USA), and three mate pair libraries (2, 5, and 10 kbp) were constructed using the Nextera Mate Pair Sample Preparation Kit (Illumina, San Diego, USA). The libraries were sequenced on the Illumina HiSeq 2500 sequencing platform at Novogene (Beijing). Cutadapt [68] was used to remove the adaptor in the sequencing data and Trimmomatic [69] was used to remove low-quality reads with a minimum quality score of 20.

RNA was extracted from each OE sample using Invitrogen TRIzol (Invitrogen, Vilnius, Lithuania). The 150 bp paired-end RNA sequencing libraries were generated using the NEBNext Ultra RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). All libraries were sequenced on the Illumina HiSeq 2500 sequencing platform at Novogene (Beijing). FastQC was used to evaluate the sequencing quality, and Trimmomatic [69] was used to remove adaptors and low-quality reads. The clean data for each sample was over 6 GB.

### Genome assembly

Jellyfish [70] was used to count of k-mers from cleaned sequencing reads to estimate genome size of *N. nycticorax* genome. The shotgun assembly of *N. nycticorax*'s whole genome was performed using SOAPdenovo2 [71]. All assembly steps were performed by following the official guideline, including read error correction, de Bruijn graph construction using short-insert-size library data, contig construction, realignment of the linkages between the contigs, and the creation of scaffolds from the short-insert-size paired ends and long-distance



paired ends. The completeness of the assembled genome was assayed using Core Eukaryotic Genes Mapping Approach (CEGMA) [72], using 248 highly conserved core eukaryotic genes (CEGs). The completeness of the assembled genome was also evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.3 [73] with Hmmersearch (hmmer.org) 3.1 and Aves (odb10) dataset, which contains 8338 universal single-copy genes in birds.

### Genome annotation

Repetitive sequences in *N. nycticorax* genome were identified by de novo and homology-based approaches. In the de novo approach, RepeatModeler (<https://www.repeatmasker.org>) was used to identify interspersed repeats and generate de novo repeat libraries. RepeatMasker (<https://www.repeatmasker.org>) was run with the de novo libraries to identify interspersed repeats. Tandem Repeats Finder [74] was used to identify tandem repeats. In the homolog-based approach, RepeatMasker was used to search the interspersed repeats against the Repbase2 repeat database. Repeated proteins were identified by using RepeatProteinMasker.

To predict protein coding genes, proteins of *Homo sapiens*, *Taeniopygia guttata*, *G. gallus*, and *Meleagris gallopavo* were downloaded and aligned to *N. nycticorax* genome using TBLASTN [75] with an E-value cutoff of  $1E-5$ . Subsequently, the homologous sequences of *N. nycticorax* genome were aligned against the matching proteins using Genewise2 [76] to predict the gene structure. Functions of the protein coding genes were annotated by blasting in KEGG [77], GO [78], SwissProt [79], and TrEMBL [79] databases. Gene domains were determined by InterProScan [80]. Moreover, tRNAscan-SE [81, 82] and INFERNAL [83] were used to annotate microRNAs and tRNAs in *N. nycticorax* genome.

### Selection of vision-related genes

The protein sequences of other Pelecaniformes birds (*E. garzetta* [GCA\_000687185.1], *N. nippon* [GCA\_000708225.1], *C. cochlearius* [VWPP000000000], *A. anhinga* [WBMU000000000], and *S. umbretta* [VZTL000000000]) and *G. gallus* (GRCg6a-GCA\_000002315.5) were downloaded from Birds 10 K Phase I and II databases on the basis of sequence depth higher than  $50\times$  and the availability of official predicted protein sequences. OrthoFinder2 [84] was used to search single-copy orthologous genes in *N. nycticorax* and the six birds with an E-value of  $1e-10$  in the all-against-all BLASTP step. A total of 4121 single-copy orthologous genes were obtained and used to construct a maximum likelihood (ML) tree with IQTREE [85] with substitution model automatically selected by ModelFinder [86].

The ML tree (Fig. 1) was used in the downstream nature selection analysis of vision-related genes.

The orthologous genes of the seven birds involved in “eye development processes” (GO:0001654), “visual perception” (GO:0007601), “detection of light stimulus” (GO:0009583), and the rod phototransduction pathway (KEGG map04744) were re-annotated by EggNog-mapper V2.1.2 [87] with hmmer [88] search engine (E-value  $1e-5$ ) in EggNog V5 Aves (8782) database [89] and Basic Local Alignment Search Tool (BLAST) search engine [90] in the NR database. The results obtained 299 vision-related orthologous groups from the EggNOG and NR databases (Supplement Table S1.xlsx). The nucleotide sequences of the orthologous genes were aligned using ParaAT 2.0 [91] and alignments less than 99 bp (33AA) were filtered out. Finally, there was 216 candidate vision related orthologous genes remaining for the test of selections, of which, 159 orthologous genes had sequences for all the seven birds and 57 genes missed in one species of *N. nippon*, *C. cochlearius*, *A. anhinga*, or *S. umbretta*. For each orthologous gene, if multiple transcript variants were available in a species, the longest CDS of these orthologous genes were used in the downstream analysis of the selection.

Based on the ML tree, the branch model in the CodeML in PAML [92] was used to test for signatures of the positive selection acting on the vision genes of *N. nycticorax* and *C. cochlearius*, which were separately labeled as foreground branch using EasyCodeML [93]. The branch model tests one ratio model (same  $\omega$  ratio of all branches) against two ratio model (different  $\omega$  ratios between foreground background branches) based on likelihood ratio tests (LRT). The P-values calculated by LRT were then adjusted by false discovery rate (FDR) correction [94] with a cutoff of 0.05. Genes with  $\omega_1$  higher than 1 and  $FDR < 0.05$  were considered as positive selection genes.

Considering that positive selection sometimes acts only on a few sites and within a short evolutionary time period, we also used BUSTED [95] in Hyphy [96] to identify the vision related gene with evidence of positive selection at a fraction of sites. For analyses, BUSTED classifies sites to three rate classes ( $\omega_1 \leq \omega_2 \leq \omega_3$ ) representing strong and weak conservation and positive selection and estimates the proportion of sites belonging to each  $\omega$  class occurring in both foreground and background branches (unconstrained model). Positive selection is then detected by comparing unconstrained model to constrained model (null model,  $\omega_3 = 1$  on the foreground branch), and the significance was tested using LRT. For our analyses, *N. nycticorax* and *C. cochlearius* branch were respectively used as the foreground branch and the remaining branches were treated as the background branches. FDR correction was used to adjust

$P$ -values. Genes with adjusted  $P$ -values less than 0.05 and  $\omega_3$  in foreground branch significantly greater than background branches were assigned as positive selected genes.

We further used two methods were used to test relaxed or intensified selection acting on the night herons. The first method was followed the procedure described in the barn owl research [35]. Briefly, a significantly higher  $\omega_{\text{Foreground}}$  than  $\omega_{\text{Background}}$  ( $P < 0.05$ ,  $\text{FDR} < 0.05$ ) in the branch model test suggests a relaxed selection in the foreground branch. A  $\omega_{\text{Foreground}}$  significantly lower than  $\omega_{\text{Background}}$  suggests the intensified selection in the foreground branch. The second method was RELAX [24] in HyPhy [96], based on the branch-site model. RELAX tests for relaxed or intensified selection were performed using parameter  $k$ , where  $k > 1$  suggests intensified selection in the test branch and that the distribution of  $\omega$  categories is close to 1 compared with that of the background, whereas  $k < 1$  suggests relaxed selection in the test branch and that the distribution of  $\omega$  categories is far from 1. The setting of test branch and background branch was the same in all the approaches,

#### OR genes annotation and analysis

A custom BLAST database was created using the known nucleotide sequences of OR genes downloaded from *H. sapiens*, *Canis lupus familiaris*, *Bos taurus*, *T. guttata*, *G. gallus*, *Anolis carolinensis*, and *M. gallopavo* to identify the OR genes in genomes of *N. nycticorax*, *C. cochlearius*, and *E. garzetta*. The genomes of the three ardeid birds were aligned to the custom database using TBLASTN [75] with an E-value cut-off of 10. A candidate OR gene that best hit with the smallest E-value was retained. Genewise2 [76] was used to search 750 bp upstream and 750 bp downstream of the retained candidate ORs for finding the open reading frame (ORF). Based on the Genewise results, the candidate OR genes with normal start codons, stop codons, and more than 650 bp size that can code for seven transmembrane (TM) domains were identified as intact genes, the candidate OR genes without a start and/or stop codon were identified as partial genes, and the candidate OR genes with frameshift mutations and/or premature stop codons were identified as pseudogenes. Subsequently, OR genes were distinguished from non-OR G protein-coupled receptors (GPCRs) using the neighbor-joining tree constructed in MEGA X software [97] using the candidate intact OR sequences and six non-OR GPCR sequences from *Frizzled* [98]. The candidate intact ORs that were clustered with non-OR GPCRs were filtered out. The candidate partial and pseudo-OR genes were subjected to BLAST, and those

that had the top BLAST hits to non-OR sequences were omitted.

The Shannon entropy ( $H$ ) [99] of Type I Class II ( $\gamma$ ) OR genes was calculated by BioEdit [100] to investigate the diversity of ORs in the three ardeid birds. The Type I Class II ( $\gamma$ ) OR sequences of the three ardeid birds were aligned using the Muscle program [101]. Gaps were excluded, and  $H$  was separately averaged across all positions for each ardeid species.  $H \geq 2.0$  is considered a variable position,  $H \leq 2$  is considered a conserved position, and  $H \leq 1.0$  is considered a highly conserved position [102].

Because phylogeny-based selection detection methods tend to overestimate the extent of positive selection among members of multigene families due to the presence of recombination or gene conversion [103], the analysis of positively selective sites on genes in OR family 14 were performed on the Datamonkey website [104] using the methods that allow to test for positive selection in the presence of recombination. Specifically, GARD [105] was used to generate multiple phylogenies based on putative nonrecombinant fragments to avoid the misleading estimation of selection caused by recombination and gene conversion. Then, SLAC [106], MEME [107], and FUBAR [108] methods were used in HyPhy (2.5.32) [95] to infer the signatures of the positive selection. Sites were considered under positive selection when they were detected by all the three methods with  $Q < 0.05$  in MEME,  $P < 0.1$  in SLAC, and Bayesian posterior probability  $> 0.95$  in FUBAR. TM-helix in OR14 was predicted by TOPCONS [109]. Site positions followed *G. gallus* XM\_001236559.6 (NCBI *G. gallus* annotation release 105-GRCg7b, chromosome assembly level) by sequence alignments.

#### OR gene expression analysis

HISAT2 [110] was used to map the high-quality reads to genomes of *E. garzetta* [111] and *N. nycticorax*. There were about 60% to 75% of sequencing reads can be mapped to the reference genomes (Supplement Table S1 and S2). The mapping results of each sample transcripts were assembled by StringTie [112, 113], and then input to TACO [114] to reconstruct a consensus transcriptome by merging the biological repeat individual transcriptomes. The TACO results include the transcript of each gene and TPM of each group's meta-assembly transcriptome merged from the biological repeats. This meta-assembly process can improve the precision of the transcript abundance [114]. OR genes expression levels were extracted from the TACO results according to functional annotation results.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08904-y>.

**Additional file 1: Supplement Fig 1.** K-mer distribution. The peak of 17-kmer located in Depth=45 and have 58,326,594,430 kmers. The raw genome size estimated by 17-kmer was  $58,326,594,430/45 \approx 1296.15\text{Mb}$ , and the revised genome size was  $1272.61\text{Mb}$  by raw genome size\* (1-E), where E represent the error rate estimate by Kdepth=1 rate. **Supplement Fig 2.** Expression numbers of ORs. Numbers of Expressed ORs in *E. garzetta*, and *N. nycticorax*. The numbers above the bars represent the expression numbers of ORs. **Supplement Table 1.** Transcriptome reads mapping in *N. nycticorax*. **Supplement Table 2.** Transcriptome reads mapping in *E. garzetta*. **Supplement Table 3.** Genome assembly statistics details. **Supplement Table 4.** Function of the visual adaptive evolution genes. **Supplement Table 5.** Positively selected sites in intact OR14 of the *E. garzetta*. **Supplement Table 6.** Positively selected sites in intact OR14 of the *N. nycticorax*. **Supplement Table 7.** Positively selected sites in intact OR14 of the *C. cochlearius*.

**Additional file 2: Supplement Table S1.** xlsx

**Additional file 3: Supplement Table S2.** xlsx

**Additional file 4: Supplement Table S3.** xlsx

## Acknowledgements

We thank Dr. Wei Xu (Xiamen University), Quanzhou Jinjiang International Airport and Egret Naturel Reserve on Xiamen Dayu Island for samples collection. We thank KGSsupport for its linguistic assistance during the preparation of this manuscript.

## Authors' contributions

H.L., X.Z., and X.C. conceived the project and designed the experiments. S.L. performed genome sequencing and assembly. H.L. performed RNA sequencing and data analyses. W.F. and Q.L. participated in the project design and provided samples. H.L., X.Z., and X.C. wrote the manuscript and revised the manuscript. All authors read and approved the final manuscript.

## Funding

This research was supported by the National Natural Science Foundation of China (Nos. 42076107 and 41676123).

## Availability of data and materials

The draft genome assembly data are available at GenBank with the accession number: JAKFQP000000000. The associated BioProject number are PRJNA796267. RNA-Seq data are available at GEO with the accession number: GSE197206. The genomic annotation files for genomic and transcriptomic analyses are available at <https://github.com/hrluo93/4Hero-n5-genome-anno.git>

## Declarations

### Ethics approval and consent to participate

This study was approved by the Ethics Committee for Animal Experimentation of the Xiamen University. The experiments were conducted in accordance with ethical guidelines.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 5 May 2022 Accepted: 20 September 2022

Published online: 03 October 2022

## References

- Kushlan JA, Hancock JA. Herons (Ardeidae) (Bird Families of the World). New York: Oxford University Press; 2005.
- Martin GR, Katzir G. Visual fields and eye movements in herons (Ardeidae). Brain Behav Evol. 1994;44(2):74–85.
- Martin GR. Through birds' eyes: insights into avian sensory ecology. J Ornithol. 2012;153(Suppl 1):S23–48.
- Maccarone AD, Hamilton BL. Diurnal and nocturnal foraging activity by black-crowned night-herons (*Nycticorax nycticorax*) at an artificial weir. Waterbirds. 2014;37(2):220–4.
- McNeil R, Drapeau P, Pierotti R. Nocturnality in colonial waterbirds: occurrence, special adaptations, and suspected benefits. In: Power DM (ed) Current ornithology, vol 10. New York: Plenum Press; 1993. p. 187–246.
- Katzir G, Martin GR. Visual fields in the black-crowned night heron *Nycticorax nycticorax*: nocturnality does not result in owl-like features. Ibis. 2008;140(1):157–62.
- Kushlan J. Feeding ecology of wading birds Wading Birds. New York: National Audubon Society; 1978.
- Gondo M, Ando H. Comparative and histophysiological study of oil droplets in the avian retina. Kobe J Med Sci. 1995;41(4):127–39.
- Rojas LM, McNeil R, Cabana T, Lachapelle P. Behavioral, morphological and physiological correlates of diurnal and nocturnal vision in selected wading bird species. Brain Behav Evol. 1999;53(5):227–42.
- Zhou C, Zhang W, Wen Q, Bu P, Gao J, Wang G, Jin J, Song Y, Sun X, Zhang Y, et al. Comparative genomics reveals the genetic mechanisms of musk secretion and adaptive immunity in Chinese forest musk deer. Genome Biol Evol. 2019;11(4):1019–32.
- Davies KTJ, Yohe LR, Almonte J, Sanchez MKR, Rengifo EM, Dumont ER, Sears KE, Davalos LM, Rossiter SJ. Foraging shifts and visual preadaptation in ecologically diverse bats. Mol Ecol. 2020;29(10):1839–59.
- Espindola-Hernandez P, Mueller JC, Carrete M, Boerno S, Kempenaers B. Genomic evidence for sensorial adaptations to a nocturnal predatory lifestyle in Owls. Genome Biol Evol. 2020;12(10):1895–908.
- Steiger SS, Fidler AE, Kempenaers B. Evidence for increased olfactory receptor gene repertoire size in two nocturnal bird species with well-developed olfactory ability. BMC Evol Biol. 2009;9:117.
- Le Duc D, Renaud G, Krishnan A, Almen MS, Huynen L, Prohaska SJ, Ongyerth M, Bitarello BD, Schiöth HB, Hofreiter M, et al. Kiwi genome provides insights into evolution of a nocturnal lifestyle. Genome Biol. 2015;16:147.
- Martin GR, Wilson KJ, Martin Wild J, Parsons S, Fabiana Kubke M, Corfield J. Kiwi forego vision in the guidance of their nocturnal activities. PLoS ONE. 2007;2(2): e198.
- Cho YS, Jun JH, Kim JA, Kim HM, Chung O, Kang SG, Park JY, Kim HJ, Kim S, Kim HJ, et al. Raptor genomes reveal evolutionary signatures of predatory and nocturnal lifestyles. Genome Biol. 2019;20:181.
- Khan I, Yang Z, Maldonado E, Li C, Zhang G, Gilbert MT, Jarvis ED, O'Brien SJ, Johnson WE, Antunes A. Olfactory receptor subgenomes linked with broad ecological adaptations in sauropsida. Mol Biol Evol. 2015;32(11):2832–43.
- Kushlan JA. Feeding Repertoire of the Boat-Billed Heron (*Cochlearius cochlearius*). Waterbirds. 2009;32(3):408–14.
- Mock DW. Feeding methods of the boat-billed heron, a deductive hypothesis. Auk. 1975;92(3):590–2.
- Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth BC, et al. Dense sampling of bird diversity increases power of comparative genomics. Nature. 2020;587(7833):252–7.
- Sheldon FH, Jones CE, McCracken KG. Relative patterns and rates of evolution in heron nuclear and mitochondrial DNA. Mol Biol Evol. 2000;17(3):437–50.
- Cobb S. The size of the olfactory bulb in 108 species of birds. Auk. 1968;85(1):55–61.
- Schneider K, Adams CE, Elmer KR. Parallel selection on ecologically relevant gene functions in the transcriptomes of highly diversifying salmonids. BMC Genomics. 2019;20(1):1010.
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32(3):820–32.

25. Davies KT, Bennett NC, Tsagkogeorga G, Rossiter SJ, Faulkes CG. Family wide molecular adaptations to underground life in African mole-rats revealed by phylogenomic analysis. *Mol Biol Evol*. 2015;32(12):3089–107.
26. Ishengoma E, Agaba M, Cavener DR. Evolutionary analysis of vision genes identifies potential drivers of visual differences between giraffe and okapi. *PeerJ*. 2017;5: e3145.
27. Wu Y, Hadly EA, Teng W, Hao Y, Liang W, Liu Y, Wang H. Retinal transcriptome sequencing sheds light on the adaptation to nocturnal and diurnal lifestyles in raptors. *Sci Rep*. 2016;6:33578.
28. Zhang W, Mulieri PJ, Gaio U, Bae GU, Krauss RS, Kang JS. Ocular abnormalities in mice lacking the immunoglobulin superfamily member Cdo. *FEBS J*. 2009;276(20):5998–6010.
29. Tassabehji M, Newton VE, Read AP. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet*. 1994;8(3):251–5.
30. Simpson J, Nien CJ, Flynn K, Jester B, Cherqui S, Jester J. Quantitative in vivo and ex vivo confocal microscopy analysis of corneal cystine crystals in the Ctns knockout mouse. *Mol Vis*. 2011;17:2212–20.
31. Zhang J, Qu B, Yu W, Zhu Y, Yan X, Shen H, Zhao J. Role of surface ectoderm-specific mitofusin 2 in the corneal morphologic development of mice. *Am J Transl Res*. 2019;11(6):3620–8.
32. Zhao J, Wu X, Wu D, Yu Y, Yu Y, Wang Y, Fu Q, Zhang J, Yao K. Embryonic surface ectoderm-specific mitofusin 2 conditional knockout induces congenital cataracts in mice. *Sci Rep*. 2018;8(1):1522.
33. Maas SM, Kayserili H, Lam J, Apak MY, Hennekam RC. Further delineation of Frank-ter Haar syndrome. *Am J Med Genet A*. 2004;131(2):127–33.
34. Iqbal Z, Cejudo-Martin P, de Brouwer A, van der Zwaag B, Ruiz-Lozano P, Scimia MC, Lindsey JD, Weinreb R, Albrecht B, Megarbane A, et al. Disruption of the podosome adaptor protein TKS4 (SH3PXD2B) causes the skeletal dysplasia, eye, and cardiac abnormalities of Frank-Ter Haar Syndrome. *Am J Hum Genet*. 2010;86(2):254–61.
35. Borges R, Fonseca J, Gomes C, Johnson WE, O'Brien SJ, Zhang G, Gilbert MTP, Jarvis ED, Antunes A. Avian binocularity and adaptation to nocturnal environments: genomic insights from a highly derived visual phenotype. *Genome Biol Evol*. 2019;11(8):2244–55.
36. Lisney TJ, Iwaniuk AN, Bandet MV, Wylie DR. Eye shape and retinal topography in owls (Aves: Strigiformes). *Brain Behav Evol*. 2012;79(4):218–36.
37. Deans MR, Krol A, Abaira VE, Copley CO, Tucker AF, Goodrich LV. Control of neuronal morphology by the atypical cadherin Fat3. *Neuron*. 2011;71(5):820–32.
38. Vugler AA, Semo M, Joseph A, Jeffery G. Survival and remodeling of melanopsin cells during retinal dystrophy. *Vis Neurosci*. 2008;25(2):125–38.
39. Lin B, Koizumi A, Tanaka N, Panda S, Masland RH. Restoration of visual function in retinal degeneration mice by ectopic expression of melanopsin. *Proc Natl Acad Sci U S A*. 2008;105(41):16009–14.
40. Ananth S, Babu E, Veeranan-Karmegam R, Bozard Baldowski BR, Boettger T, Martin PM. Induction of the cystine/glutamate exchanger SLC7A11 in retinal pigment epithelial cells by the antipsoriatic drug monomethylfumarate. *Invest Ophthalmol Vis Sci*. 2013;54(3):1592–602.
41. Marneros AG, Fan J, Yokoyama Y, Gerber HP, Ferrara N, Crouch RK, Olsen BR. Vascular endothelial growth factor expression in the retinal pigment epithelium is essential for choriocapillaris development and visual function. *Am J Pathol*. 2005;167(5):1451–9.
42. Patten SA, Jacobs-McDaniels NL, Zaouter C, Drapeau P, Albertson RC, Moldovan F. Role of Chd7 in zebrafish: a model for CHARGE syndrome. *PLoS ONE*. 2012;7(2): e31650.
43. Gerding WM, Schreiber S, Schulte-Middelmann T, de Castro MA, Atorf J, Akkad DA, Dekomien G, Kremers J, Dermietzel R, Gal A, et al. Ccdc66 null mutation causes retinal degeneration and dysfunction. *Hum Mol Genet*. 2011;20(18):3620–31.
44. Acharya S, Foletta VC, Lee JW, Rayborn ME, Rodriguez IR, Young WS 3rd, Hollyfield JG. SPACRCAN, a novel human interphotoreceptor matrix hyaluronan-binding proteoglycan synthesized by photoreceptors and pinealocytes. *J Biol Chem*. 2000;275(10):6945–55.
45. Jiang ST, Chiou YY, Wang E, Chien YL, Ho HH, Tsai FJ, Lin CY, Tsai SP, Li H. Essential role of nephrocystin in photoreceptor intraflagellar transport in mouse. *Hum Mol Genet*. 2009;18(9):1566–77.
46. Naeem MA, Chavali VR, Ali S, Iqbal M, Riazuddin S, Khan SN, Husnain T, Sieving PA, Ayyagari R, Riazuddin S, et al. GNAT1 associated with autosomal recessive congenital stationary night blindness. *Invest Ophthalmol Vis Sci*. 2012;53(3):1353–61.
47. Manes G, Cheguru P, Majumder A, Bocquet B, Sénéchal A, Artemyev NO, Hamel CP, Brabet P. A truncated form of rod photoreceptor PDE6  $\beta$ -subunit causes autosomal dominant congenital stationary night blindness by interfering with the inhibitory activity of the  $\gamma$ -subunit. *PLoS One*. 2014;9(4):e95768.
48. Wu Y. Widespread nocturnality of living birds stemming from their common ancestor. *BMC Evol Biol*. 2019;19(1):189.
49. Prueitt RL, Zinn AR. A fork in the road to fertility. *Nat Genet*. 2001;27(2):132–4.
50. Wu G, Glickstein S, Liu W, Fujita T, Li W, Yang Q, Duvoisin R, Wan Y. The anaphase-promoting complex coordinates initiation of lens differentiation. *Mol Biol Cell*. 2007;18(3):1018–29.
51. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, Gershoni M, Morrey CP, Safran M, Lancet D. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*. 2017;45(D1):D877–87.
52. Ding K, Scortegagna M, Seaman R, Birch DG, Garcia JA. Retinal disease in mice lacking hypoxia-inducible transcription factor-2alpha. *Invest Ophthalmol Vis Sci*. 2005;46(3):1010–6.
53. Votruba M, Moore AT, Bhattacharya SS. Demonstration of a founder effect and fine mapping of dominant optic atrophy locus on 3q28-qter by linkage disequilibrium method: a study of 38 British Isles pedigrees. *Hum Genet*. 1998;102(1):79–86.
54. Johnston RL, Seller MJ, Behnam JT, Burdon MA, Spalton DJ. Dominant optic atrophy. *Ophthalmology*. 1999;106(1):123–8.
55. Niimura Y, Nei M. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet*. 2006;51(6):505–17.
56. Steiger SS, Fidler AE, Valcu M, Kempnaers B. Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proc Biol Sci*. 2008;275(1649):2309–17.
57. Le Duc D, Schoneberg T. Adaptation to nocturnality - learning from avian genomes. *BioEssays*. 2016;38(7):694–703.
58. Niven JE, Laughlin SB. Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol*. 2008;211(Pt 11):1792–804.
59. Bullmore E, Sporns O. The economy of brain network organization. *Nat Rev Neurosci*. 2012;13(5):336–49.
60. Aiello LC, Wheeler P. The Expensive-Tissue Hypothesis: The brain and the digestive system in human and primate evolution. *Curr Anthropol*. 1995;36(2):199–221.
61. Laughlin SB, de Ruyter van Steveninck RR, Anderson JC. The metabolic cost of neural information. *Nat Neurosci*. 1998;1(1):36–41.
62. Ozer I, Carle T. Back to the light, coevolution between vision and olfaction in the "Dark-flies" (*Drosophila melanogaster*). *PLoS ONE*. 2020;15(2): e0228939.
63. Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y. Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol*. 2007;8(5):R86.
64. Young JM, Shykind BM, Lane RP, Tonnes-Priddy L, Ross JA, Walker M, Williams EM, Trask BJ. Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biol*. 2003;4(11):R71.
65. Lv LY, Liang XF, He S. Genome-wide identification and characterization of olfactory receptor genes in Chinese perch, *Siniperca chuatsi*. *Genes (Basel)*. 2019;10(2):178.
66. Ferreira T, Wilson SR, Choi YG, Risso D, Dudoit S, Speed TP, Ngai J. Silencing of odorant receptor genes by G protein betagamma signaling ensures the expression of one odorant receptor per olfactory sensory neuron. *Neuron*. 2014;81(4):847–59.
67. Sin SYW, Cloutier A, Nevitt G, Edwards SV. Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics*. 2022;220(2):iyab210.
68. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;17(1):138–43.
69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
70. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764–70.



71. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 2010;20(2):265–72.
72. Parra G, Bradnam K, Korff I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics.* 2007;23(9):1061–7.
73. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2.
74. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
75. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res.* 2016;44(D1):D646–653.
76. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res.* 2004;14(5):988–95.
77. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
78. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. *Gene Ontol Consortium Nat Genet.* 2000;25(1):25–9.
79. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000;28(1):45–8.
80. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics.* 2001;17(9):847–8.
81. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 2019;1962:1–14.
82. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016;44(W1):W54–57.
83. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013;29(22):2933–5.
84. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;16:157.
85. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
86. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9.
87. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825–9.
88. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011;7(10):e1002195.
89. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47(D1):D309–14.
90. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
91. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun.* 2012;419(4):779–81.
92. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
93. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol.* 2019;9(7):3891–8.
94. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika.* 2006;93(3):491–507.
95. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. Gene-wide identification of episodic selection. *Mol Biol Evol.* 2015;32(5):1365–71.
96. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol.* 2020;37(1):295–9.
97. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547–9.
98. Lagerstrom MC, Hellstrom AR, Gloriam DE, Larsson TP, Schioth HB, Fredriksson R. The G protein-coupled receptor subset of the chicken genome. *PLoS Comput Biol.* 2006;2(6): e54.
99. Shannon CE. A mathematical theory of communication. *Bell System Tech J.* 1948;27(3): 379–423.
100. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
101. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:113.
102. Samuel L, Rita J. Shannon information as a measure of amino acid diversity. Berlin: Springer, Berlin Heidelberg; 1992.
103. Steiger SS, Fidler AE, Mueller JC, Kempenaers B. Evidence for adaptive evolution of olfactory receptor genes in 9 bird species. *J Hered.* 2010;101(3):325–33.
104. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* 2018;35(3):773–7.
105. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelck CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 2006;23(10):1891–901.
106. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22(5):1208–22.
107. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8(7): e1002764.
108. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* 2013;30(5):1196–205.
109. Tsirogas KD, Peters C, Shu N, Kall L, Elovsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 2015;43(W1):W401–407.
110. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
111. Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, Carey HV, Zhang Q, Zheng H, Wei S, et al. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biol.* 2014;15(12):557.
112. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. *StringTie and Ballgown Nat Protoc.* 2016;11(9):1650–67.
113. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5.
114. Niknafs YS, Pandian B, Iyer HK, Chinnaiyan AM, Iyer MK. TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat Methods.* 2017;14(1):68–70.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

### At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

