Demographic histories and genome-wide patterns of divergence in incipient species of shorebirds

Xuejing Wang[1§], Kathryn H. Maher[2§], Nan Zhang[1§], Pingjia Que[3], Chenqing Zheng[1,4], Simin Liu[1], Biao Wang[5], Qin Huang[1], De Chen[3], Xu Yang[4], Zhengwang Zhang[3], Tamás Székely[2], Araxi O. Urrutia[2,6*], Yang Liu[1*]

1. State Key Laboratory of Biocontrol, Department of Ecology, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, P. R. China
2. Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK
3. Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, 100875, P. R. China
4. Shenzhen Realomics Biological Technology Ltd, Shenzhen, 518000, P. R. China
5. Research Center of Precision Medicine, College of Life Sciences and Oceanography, Key laboratory of Optoelectronic Devices and System of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, P. R. China
6. Instituto de Ecología, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico.

Total words (without references): 6985

Running title: Speciation genomics of two plover species

§ These authors contributed equally to this work.

* Correspondence should be addressed to

Yang Liu,

State Key Laboratory of Biocontrol, Department of Ecology, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, P. R. China. Email: liuy353@mail.sysu.edu.cn

Araxi O. Urrutia,

Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK. Email: a.urrutia@bath.ac.uk

Abstract

Understanding how incipient species are maintained with gene flow is a fundamental question in evolutionary biology. Whole genome sequencing of multiple individuals holds great potential to illustrate patterns of genomic differentiation as well as the associated evolutionary histories. Kentish (*Charadrius alexandrinus*) and the white-faced (*C. dealbatus*) plovers, which differ in their phenotype, ecology and behaviour, are two incipient species and parapatrically distributed in East Asia. Previous studies show evidence of genetic diversification with gene flow between the two plovers. Under this scenario, it is of great importance to explore the patterns of divergence at the genomic level and to determine whether specific regions are involved in reproductive isolation and local adaptation. Here we present the first population genomic analysis of the two incipient species based on the de novo Kentish plover reference genome and resequenced populations. We show that the two plover lineages are distinct in both nuclear and mitochondrial genomes. Using model-based coalescence analysis, we found that population sizes of Kentish plover increased whereas white-faced plovers declined during the Last Glaciation Period. Moreover, the two plovers diverged allopatrically, with gene flow occurring after secondary contact. This has resulted in low levels of genome-wide differentiation, although we found evidence of a few highly differentiated genomic regions in both the autosomes and the Z-chromosome. This study illustrates that incipient shorebird species with gene flow after secondary contact can exhibit discrete divergence at specific genomic regions and provides basis to further exploration on the genetic basis of relevant phenotypic traits.

Key words: speciation, population genomics, shorebirds, gene flow, natural selection

Introduction

Understanding the conditions in which speciation occurs is a fundamental question in evolutionary biology. Of equal importance is the understanding of how newly diverged species (incipient species) are maintained, as it is likely that interspecific gene flow is a common occurrence between diverging species [1]. During allopatric speciation, a physical barrier acts to prevent gene flow across the whole genome [2] and pre- and post-zygotic mechanisms of reproductive isolation can evolve to facilitate divergence. However, gene flow across geographical barriers, or secondary contact between diverged populations is possible, which allows gene flow to recommence [3]. Even infrequent gene flow can erode species barriers [4]. In the more contentious geographical context, such as parapatric or sympatric speciation [5, 6], disentangling the relative role of gene flow and other diverging conditions and forces remains challenging [7, 8].

Whether speciation can occur with gene flow has been an area of intense investigation within the last decade [9]. In certain instances, gene flow overcomes the barriers of reproductive isolation and reverse speciation processes occur, while in other species divergence persists in spite of gene flow [10–14]. The homogenising effect of gene flow, on both a genetic and phenotypic level, can be reduced in sympatry and at secondary contact areas if the diverging species vary in niche or mate preference [7, 15–18]. Areas of elevated genetic differentiation found throughout the genome, so called ‘genomic islands’, or a subset of these regions, could be responsible for the phenotypic differences observed between species or contain potential mechanisms for reproductive isolation and hybrid incompatibility [18–22]. These regions of high genetic differentiation are often spread widely throughout the genome and can be either small [e.g. 23] or large in size [e.g. 24]. As gene flow leaves delectable signatures of divergence in the introgressed regions of genome [1], it is possible to study these patterns and infer past gene flow and also the demographic history of a species [1, 25, 26].

77 Improved estimation of demographic history makes it possible to better understand population
78 differentiation and speciation mechanisms [e.g. 27]. It allows accurate estimates of gene flow,
79 divergence time, effective population size and population changes in size throughout time. It is also
80 possible to obtain information on which is the most likely demographic history of speciation, such as
81 isolation, isolation with migration, early migration or secondary contact. Using whole-genome
82 approaches also makes it possible to screen for fast evolving regions along genomes that techniques
83 using a small number of genetic markers may miss [e.g. 28]. For example, genomic resequencing in
84 carrion (*Corvus corone*) and hooded crows (*C. cornix*) found that distinct differences in phenotype are
85 maintained by variation in less than 1% of the genome [18]. Finally, population genomics has been
86 shown to provide markedly different estimates of effective population sizes compared to the use of a
87 reduced set of molecular markers and thus could better identify species and populations at risk of
88 extinction, or populations with unique genetic structure worthy of conservation effort. Assessing how
89 secondary contact and hybridisation between distinct taxa impacts native populations is of vital
90 importance when considering how to implement effective conservation protocols [29, 30].
91 Advancements in analysis on the basis of coalescence simulations [31] using unprecedented high-
92 throughput genomic data of a single individual [32] or populations hold great potential to reveal
93 inference about demographic histories [33].

94 *Charadrius* plovers are model species for investigating breeding system evolution and have been used
95 in numerous studies to better understand mating and parenting behaviours [e.g. 34–39]. Species in
96 the Kentish plover complex (*Charadrius alexandrinus*; KP) are small shorebirds found breeding on
97 saline lakes and coasts throughout Eurasia and North Africa [40]. A previous study found no genetic
98 differentiation between several Eurasian continental populations of Kentish plover [41]. In East Asia,
99 the subspecies *C. a. alexandrinus* has a wide breeding range in the temperate zone, whereas the
100 southern subspecies *C. a. dealbatus*, (known as white-faced plovers; WFP) show distinct phenotypic
101 traits compared to northern populations. They lack the dark eye barring of the KP, and have lighter

lower ear coverts, a brighter cinnamon cap and paler plumage with white lores [42]. They also typically have longer wings, beak and tarsus and are more commonly found on sandier substrates, with more active foraging behaviour and a more upright stance [42]. While the KP and WFP share much of the same wintering range, they have largely non-overlapping breeding ranges. WFPs breed exclusively in a restricted coastal range from Fujian to Guangxi, as well as Hainan island in south-east China. KPs nest to the north of this range [43]. Previous work examining mitochondrial DNA and microsatellite markers has shown that although KPs and WFPs are phenotypically well-differentiated [42, 44], genetically they lack differentiation. More extensive microsatellite genotypes and autosomal nuclear sequences, however, showed KPs and WFPs are distinct and young lineages (diverged around 0.6 mya), and bidirectional gene flow occurred between them [43].

Here, we expand previous works by characterizing demographic histories and genomic landscape of divergence in two closely related plovers. Because the isolation-with-migration model (IM) applied in a previous study assumes gene flow throughout the entire divergence history of the two plover [43], it was unknown whether gene flow persisted in the early stages of divergence or/and also occurred after the secondary contact. The current work explores these different scenarios of gene flow using based on advanced modelling on historical demography. Further, under the model of speciation- with-gene-flow model [10–14], it is of importance to investigate the potential heterogeneous genomic landscape of incipient species. In the case of the two study species, It is possible that a small number of genomic regions involved in the phenotypic and ecological differences among them [28, 42]. Hence, we attempted to disentangle the aforementioned questions by applying whole-genome sequencing and assembly of a high-quality *de novo* reference genome of a female KP. We also re-sequenced whole genomes of 21 unrelated male genomes from five populations of KP and six populations of WFP in China as well as full mitochondrion genomes of four KP and two WFP.

Materials and Methods

Sampling collection

A single female KP (heterogametic sex) was collected using mist nets in coastal Xitou, Yangjiang county, Guangdong, China in November 2014. A muscle sample was taken from this individual, stored in RNALater (QIAGEN, USA) and transported to the sequencing centre for *de novo* whole genome sequencing (BGI-Shenzhen). In addition, twenty male KP and WFP were collected from breeding colonies at 11 sites for whole genome resequencing (Figure 1 and Supplementary Table 1), including one inland site at Qinghai Lake, and one continental island, Hainan, with the remaining sites located along the Chinese coast, starting from Hebei to Guangxi. Using males avoids systematic biases occurring caused by differences in coverage of the autosomes and Z chromosome [18]. One female WFP was collected from Hainan for resequencing in higher coverage. These individuals were captured on nests using funnel traps during the breeding season between March and July in 2014-2015 [45]. Blood samples taken from these individuals were stored in RNALater (QIAGEN, USA) at -40°C. All bird captures and sampling was performed with permission from the respective authorities (Beijing Normal University to PJQ and Sun Yat-sen University to YL) and blood and tissue collection procedures conform to the regulations of the animal experimental and medical ethics committee of Sun Yat-sen University.

de novo sequencing and assembly of Kentish Plover genome

We isolated DNA from the blood/muscle samples using Qiagen DNeasy Blood and Tissue Kit using the standard manufacturer protocols. Short-insert-sized (170 and 800bp) and mate-pair (2, 5, 10 and 20kb) DNA libraries were constructed for the KP reference genome (Supplementary Table 2). All the libraries were sequenced using Illumina Hiseq 2000 platform on paired-end data. Paired-end sequence data from the genomic DNA libraries were assembled using short oligonucleotide analysis package SOAPdenovo [46]. Final N50 contig and scaffold sizes were calculated based on a minimum length of sequence >100bp. The sequencing coverage, depth and GC content distribution were evaluated by

mapping all sequencing reads of the short-insert-sized libraries back to the scaffolds using BWA [47] with the algorithm of BWA-MEM. We also evaluated the genome assembly completeness using BUSCO's genome mode [48].

Genome annotation

We combined the homologous prediction method based on the RepBase (http://www.girinst.org/repbase) using the software RepeatMasker and RepeatProteinMask with the *de novo ab initio* prediction method based on self-sequence alignment and repetitive sequence characteristic using the software RepeatModeler (http://www.repeatmasker.org/RepeatModeler), Trf and LTR-FINDER [49]. We used homology, *ab initio* prediction, EST and RNA sequencing to identify protein-coding genes. For homology-based gene prediction, chicken, turkey and zebra finch proteins were downloaded from Ensembl Release 84 (http://www.ensembl.org/info/data/ftp/index.html) and mapped onto the repeat-masked KP genome using tblastn v.2.2.21 [50]. Then we aligned homologous genome sequences against the matching proteins using Genewise (v.wise2.1.23c,10/22/2002) [51] to define gene models. Subsequently, we used the *ab initio* gene prediction software Genescan [52] and Augustus [53] to predict protein-coding genes using parameters trained from a set of high quality homologue prediction proteins. Finally, we carried out RNA-seq on samples of muscle, blood, heart, kidney, liver and brain from the same individual for the *de novo* sequencing. We mapped RNA sequence reads data to the genome using Tophat [54], and obtained transcription-based gene structure using Cufflinks [55]. We also mapped the gene structure and EST sequences to the genome using Genewise. Finally, we merged all genes predicted by the three methods using EvidenceModeler [56], and then removed all genes with a length shorter than 50 amino acids and only with *ab initio* support and FPKM confidence <5 to generate the final gene set. Gene functions were assigned according to the best match of the alignment to the SwissProt, TrEMBL, KEGG and InterPro exogenous protein databases using BLASTP. For the non-coding RNA annotation, we used the software tRNAscan-SE [57] to find the tRNA sequence in Kentish plover genome according to the tRNA architectural

feature. Since rRNA sequences are highly conservative, we used similar species rRNA sequences as a reference to find rRNA sequence in the genome using BLASTN. We also used the software INFERNAL [58] with a Rfam family's covariance model to predict the miRNA and snRNA sequences in the genome.

Gene family analysis

To define gene families that descended from a single gene in the last common ancestor, we downloaded all protein-coding genes of 15 waterbirds (Emperor penguin, *Aptenodytes forsteri*; Adélie penguin, *Pygoscelis adeliae*; Grey-crowned crane, *Balearica regulorum*; Killdeer, *Charadrius vociferus*; Little egret, *Egretta garzetta*; Sunbittern, *Eurypyga helias*; Northern fulmar, *Fulmarus glacialis*; Red-throated loon, *Gavia stellata*; Crested ibis, *Nipponia nippon*; Dalmatian pelican, *Pelecanus crispus*; White-tailed tropicbird, *Phaethon lepturus*; Common cormorant, *Phalacrocorax carbo*; Greater flamingo, *Phoenicopterus ruber*; Great crested grebe, *Podiceps cristatus*; and Mallard, *Anas platyrhynchos*) from the GigaScience database [59]. Together with the Kentish plover, these species' genomes represent the major orders in the core waterbird radiation in the avian tree of life [60]. We chose the longest isoform to represent each gene and employed BLASTP to identify potential homologous genes using E-value<1e-10. The raw Blast results were refined using solar (an in-house software) by which the high-scoring segment pairs (HSPs) were conjoined. Similarity between protein sequences were evaluated using bit-score, followed by clustering algorithm in the Treefam pipeline [61].

Phylogenetic tree construction and divergence time estimate

The protein sequences of one-to-one orthologous genes were aligned using MUSCLE [62] with the default parameters. We then filtered the gap sites from the alignments. The trimmed protein alignments were used as a guide to align corresponding coding sequences (CDS). The phylogenetic tree was reconstructed using RaxML version v8.1.19 [63] and GTRGAMMA model. Divergence times

198 between species were calculated using MCMC tree program implemented in the PAML version 4 [64].
199 Based on the phylogenetic tree, branch-specific synonymous substitution rate was estimated for KP.

200 ***Whole-genome resequencing***

201 We extracted genomic DNA from blood samples of 20 male individuals of KP and WFP, 10 of each
202 species for 5x depth resequencing and 1 female WFP for 30x depth. For each individual, 1–3 μg of DNA
203 was sheared into fragments of 200–800 bp with the Covaris system. DNA fragments were then treated
204 according to the Illumina DNA sample preparation protocol: fragments were end repaired, A-tailed,
205 ligated to paired-end adaptors and PCR amplified with 500-bp inserts for library construction.
206 Sequencing was performed on the Illumina HiSeq 2000 platform, and 100-bp paired-end reads were
207 generated.

208 ***Read mapping and SNPs calling***

209 After quality control, the reads were mapped to the KP genome using BWA and reads having a mean
210 of approximately 5x depth for each individual and >90% coverage of the KP genome were retained for
211 SNP calling. We used GATK v 3.5 [65] program to call SNPs. SNPs were filtered using VCFtools [66] and
212 GATK by following criteria: 1) missing rate <=0.10; 2) allele frequency >0.05; 3) each 10 bp <=3 SNPs.

213 ***Mitochondrion genome analysis***

214 To infer evolutionary history from mitochondrial DNA (mtDNA), we conducted mitochondrion genome
215 sequencing. Six blood samples were selected, including four KP from four sites (1. Xinbei, Taiwan; 2.
216 Weihai, Shandong; 3. Qinghai Lake, Qinghai; 4. Zhoushan, Zhejiang) and two WFP from two sites (5.
217 Dongfang, Hainan; 6. Minjiang Estuary Fuzhou, Fujian). Gross genomic DNA was extracted by TIANamp
218 Blood Genomic DNA Extraction Kit (TIANGEN, China), following the standard extraction protocol.
219 Paired-end (PE) 150-bp sequence reads were obtained from Illumina MiSeq PE150 sequencing for each
220 sample. Novogene Ltd. (Beijing) performed the library preparation and sequencing. Consequently, we
221 obtained 31,282,372, 29,185,701, 25,092,010, and 24,383,015 clean paired-end reads for the four KP

222 from four sites, respectively; and 30,397,339 and 24,982,573 clean paired-end reads for the two WFP
223 from two sites, respectively. We mapped the clean reads to the mitochondrial genome of the pied
224 avocet, *Recurvirostra avosetta* (GenBank Accession Number: KP757766), using "Map to Reference"
225 tool in Geneious R8 (Biomatters, Auckland, New Zealand) with a medium-low sensitivity and ran 5
226 iterations. Consensus sequences were saved using a 75% masking threshold, and sites that received
227 insufficient coverage (<5x) were coded using the IUPAC ambiguity symbol N.
228 We inferred mitochondrial phylogenetic relationship between the two plovers with Bayesian Inference
229 in MrBayes v.3.2.6 [67] and maximum likelihood in RAxML v8 [63] using complete mitochondrial
230 genome sequences including the pied avocet as outgroup. MrBayes was run on the CIPRES science
231 Gateway portal [68] with Metropolis coupling (four chains) set for 10 million generations and sampling
232 every 10000 generations, using HKY nucleotide substitution model which was best-fit model tested by
233 jModelTest 2 [69]. Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used to check the
234 effective sample sizes (ESS) for parameter estimation. RAxML was also run on CIPRES with GTRCAT
235 model and 1,000 bootstrap runs. Maximum-likelihood-bootstrap proportions (MLBS) ≥ 70% were
236 considered strong support [70]. The phylogenetic trees were modified using FigTree v1.4.3
237 (http://tree.bio.ed.ac.uk/software/figtree).

238 ***Population structure and divergent history between the two plover species***

239 To infer population structuring between KP and WFP, we carried out genetic admixture analysis of the
240 resequenced individuals with ADMIXTURE 1.3 [71]. For K from 1 to 5, each analysis was performed
241 using 200 bootstraps. We applied two approaches to reconstruct the demographic history of KP and
242 WFP. First, we performed a pairwise sequentially Markovian coalescent (PSMC) model to examine
243 changes in historical effective population sizes (*Ne*) of both species [32]. This enabled us to infer
244 demographic dynamics between about 10Ka to 10 Ma. The parameters were set as: "N30 –t5 –r5 –p
245 4+30*2+4+6+10", following Nadachowska-Brzyska et al. [72]. Generation time was set to 2.5 years.

246 The estimated synonymous substitution rate was used as mutation rate, which was 8.11×10^{10} per base
247 pair per year. 100 bootstraps were performed for each analysis. In addition, we carried out a model-
248 based method using an Approximate Bayesian Computational (ABC) approach to infer the divergence
249 history between the two plovers. To achieve this, we first defined four basic demographic models,
250 including: 1) Isolation model, no gene flow during divergence; 2) Isolation with migration model,
251 constant gene flow during divergence; 3) Early migration model, gene flow only exists within the early
252 period of divergence; 4) Secondary contact model, gene flow only exists in the late period of
253 divergence. We performed two groups of simulations with different effective population size (*Ne*)
254 settings (Figure 2). In the first group (A), effective population sizes were hypothesised to be constant
255 in all the four models. In the second group (B), effective population sizes changed based on PSMC
256 results. Model illustrations and priors are shown in Figure 2 and Supplementary Table 17. We used
257 msABC [73] to perform coalescent simulations for these eight models. To obtain the observed data
258 and priors for simulation, 10kbp length loci were randomly chosen from scaffolds with coverage over
259 80%. The distance between each two loci was higher than 500 Kb to reduce the effect of linkage [27].
260 Loci with missing data of over 40% in any individual, or over 30% on average, were excluded. F_{ST},
261 *Tajima's D* [74], and LD r^2 were calculated for each locus with VCFtools 0.1.14. Loci with F_{ST} higher than
262 0.159 or *Tajima's D* >1 or <0 were excluded. 143 10-kbp loci with 14,087 SNPs were left for analysis.
263 1,000,000 simulations were executed for each model. R package "abc" [75] was used to choose the
264 model which best fitted the observed data with a tolerance of 0.005. Model selection was performed
265 using a multinomial logistic regression method, first in group A and B separately, and then between
266 groups by using two models with highest likelihood in each group. 5,000,000 simulations were used
267 for the best fitted model to estimate population parameters. We used the neural network method
268 with the Epanechnikov kernel to calculate the posterior densities [76]. The number of neural networks
269 was 50. All parameters were log-transformed and medians were used as the as the point estimates.

270 ***Detecting and annotating genomic region under selection***

271 In order to better understand the divergence patterns between the two plover species, genetic
272 parameters π, *Tajima's D*, F_{ST} and D_f (fixed differences) were calculated in 50 kb blocks with VCFtools
273 0.1.14. Blocks with a length shorter than 25kbp were excluded.
274 In order to map the KP genome scaffolds onto chromosome coordinates, we downloaded the chicken
275 (*Gallus gallus*) genome from NCBI database (GCF_000002315.4) and computed its alignment with the
276 KP reference genome using Satsuma version 3.1.0 [77]. We divided the genome into non-overlapping
277 windows of 50kb in size with the first window of each scaffold beginning with position 1 of that scaffold,
278 oriented along the chromosome. For each window, we estimated the population genomic parameters
279 calculated above. Finally, we generated the genomic landscape of population divergence in KP and
280 WFP according to the above methods.

281 To calculate *Tajima's D* per gene we used the R package PopGenome v2.2.3 [78] in R v3.3.2 [79] for KP
282 and WFP separately. VCF and complementary GFF files were loaded into PopGenome by scaffold with
283 positions with unknown nucleotides excluded (include.unknown=FALSE). Data was then split into
284 genes and neutrality statistics were calculated for coding regions. For calculating F_{ST}, samples were
285 assigned to their two respective species, KP and WFP. F_{ST} statistics were generated per gene for coding
286 regions.

287 ***Gene ontology annotations***

288 To obtain gene ontology (GO) categories, Kentish plover proteins were BLASTed against the RefSeq
289 protein database using BLASTP v.3.2.0+, with an *E*-value of 1e-5 [80]. GO terms were then assigned
290 using Blast2GO software v.4.1.9 [81] and merged with GO terms obtained from InterProScan v.5.25
291 (parameters: -f xml, -goterms, -iprlookup) [82]. GO categories were then split into groups associated
292 with biological processes and cellular components. GO categories with fewer than 50 genes were
293 grouped into a single "small" category. Genes with no GO annotation were assigned to an
294 "uncharacterised" category.

Gene ontology enrichment analyses

Genes were assigned to either autosomes or the Z chromosome, with genes of unknown location excluded from GO enrichment analysis. GO enrichment was performed for genes with high F_{ST}. Genes were considered to have high F_{ST} if they fell above the 95th percentile of F_{ST} values, using R quantile(c(.95)). Two cut-off values were used, 0.184 for autosomal genes and 0.262 for genes on the Z chromosome. GO enrichment was also performed on genes with high and low *Tajima's D* values calculated for KP and WFP separately. Genes with positive *Tajima's D* could indicate balancing selection. Regions of negative *Tajima's D* can indicate strong positive selection or selective sweeps. Genes were separated into known autosomal and Z chromosomal genes for positive and negative *Tajima's D* separately. The 95th percentile of positive values were taken as high *Tajima's D* genes and the 5th percentile were taken for the negative values. This resulted in cut-off values of $\geqslant$1.565 for high and $\leqslant$-1.179 for low value autosomal genes for KP and $\geqslant$1.669 and $\leqslant$-1.191 for WFP. $\geqslant$1.723 and $\leqslant$-1.234 cut-offs were used for the Z chromosome for KP and $\geqslant$1.835 and $\leqslant$-1.445 for WFP. GO enrichment was performed for both high and low *Tajima's D* values separately. GO enrichment analyses were performed to evaluate if any GO category was over represented in the set of genes of interest compared to equally sized samples of genes drawn randomly. The expected number of genes annotated to each GO was calculated using 1000 equally sized random families. Significance was established using Z-scores and a Benjamini-Hochberg correction was applied to adjust for multiple comparisons. GO categories were significantly enriched if the adjusted p value was < 0.05. Results were filtered to remove any category with an expected number of genes per category of < 1 and an observed number of GOs of 1. This analysis was performed using R v.3.3.2.

Results

De novo sequencing the Kentish plover genome

Muscle samples from a heterogametic sex female Kentish plover were collected from a wintering population in Guangdong, China (Figure 1). Short read DNA sequencing (125bp) was carried out using the Illumina platform (see Supplementary Figure 1 for pipeline). After filtering out low quality and clonally duplicated reads, we obtained a genome assembly from 1.81 billion reads in six paired-end and mate-pair libraries that provide 134-fold coverage with a total assembly length of 1.16Gb (Supplementary Table 2). This approximates the genome size estimated using K-mer frequency method (Supplementary Table 3 and Supplementary Figure 2). The GC content versus depth is gathered into a cluster showing the genome sequence is pure and has no pollution from other species (Supplementary Figure 3.

The contig and scaffold N50 sizes are 38.9 and 3220.7kbp, respectively, with the largest scaffold spanning 15291.1kbp (Supplementary Table 4). Although the number of scaffolds for the Kentish plover is considerably higher than that of the chicken or the zebra finch genomes, the estimated genome size for the Kentish plover (1,245,524,081bp, ~1.25 Gb) is comparable to the sequenced genomes of these two species (Supplementary Table 5). Whole genome alignment reveals that, as expected, a higher proportion of the zebra finch and KP genomes can be aligned against each other than either of them can to the more distantly related chicken with over 900Mbp that can be aligned between the two species (Supplementary Table 6). The BUSCO assessment results indicate that the KP genome assembly has high completeness (C: 94%) (Supplementary Figure 4).

Within the genome sequence of the KP, fewer than 4% of the bases were N bases (Supplementary Table 7). We estimate that repetitive sequences compose about 10% of the genome (Supplementary Table 8) with LINE transposons being the most common, making up about 8% of the genome. LTR transposons are the second most common repetitive sequences in the Kentish plover genome with 2.22% occupancy, with other DNA and LINE transposable elements occupying less than one per cent of the genome (Supplementary Table 9). Annotation of non-protein coding RNA genes revealed around 722 non-coding genes with transfer RNAs and small nuclear RNAs being the most common with over

200 copies each (Supplementary Table 10). Based on the number of RNA genes identified in heavily annotated genomes such as human and mouse, it is likely more will be found in the future as non-coding gene annotation tools improve further.

Initially, annotation of protein coding genes identified 8,893 genes. Further annotation based on homologous sequence alignment against the chicken, the zebra finch and the turkey aided to identify a final set of 15,677 genes in the Kentish plover genome (Supplementary Table 11). The quality of protein coding gene annotation is of similar quality to that of other avian genomes (Supplementary Figure 5). Of these genes, based on ortholog annotations, the vast majority of protein coding genes (15,644) were annotated against at least one functional category in one of several functional annotation databases (Supplementary Table 12).

Protein coding gene content was then compared to that of 15 previously sequenced waterbird species' genomes. Gene number identified ranged from 13,454 in the Gaviiformes to the egret with 16,585 annotated protein coding genes, with the Kentish plover having a number close to the average (Supplementary Table 13). The number of genes found in the KP genome are close to the average number of genes found in other waterbird species (Supplementary Table 13). Most KP genes have homologs in all other shorebird genomes it was compared to (12,196/15,677; Supplementary Figure 6). Single copy orthologs make up almost 40% of the genes with other orthology relationships making up most of the rest of the genes. Less than 5% of genes did not cluster into orthology groups (Supplementary Figure 7).

Phylogenetic reconstruction from coding sequence alignment places the KP as a closer relative of a previously sequenced plover genome of the killdeer *Charadrius vociferous* than to any other waterbird genome and with the most recent divergence time, as expected (Supplementary Figure 8; Supplementary Figure 9).

Genome resequencing reveals two diverging plover species with contrasting evolutionary histories

368 Blood samples were obtained from a total of 21 individuals taken from mainland and Hainan island
369 plover populations along the Chinese coast and from the inland and a high-altitude population of
370 Qinghai Lake (Figure 1 and Supplementary Table 1). Kentish and white-faced plovers have been shown
371 to present distinct phenotypic features, including facial plumage pattern (Figure 1) [44]. For the 20 low
372 depth resequenced samples, a total of 914,529,390 high quality paired-end reads were retained for
373 the further analyses (Supplementary Table 14). Genome resequencing was carried out resulting in over
374 95% genome coverage with a depth of over 4x for around 70% of the genome per individual
375 (Supplementary Table 15). After filtering, a total of 11,959,725 high quality SNPs were retained. KP and
376 WFP were found to cluster into two distinct groups based on Admixture analysis (Figure 3a). For the
377 high depth sequencing WFP, its genome assembly quality was high enough to be used for the PSMC
378 analysis.

379 In addition, we obtained 15,613 bp of the complete mitochondrial genome for Kentish and white-faced
380 plovers, except for the D-loop region, which had poor assembly quality. In the phylogenetic analysis,
381 topologies between Bayesian and ML tree were consistent (Figure 3b). Analyses clearly show that the
382 monophyletic relationship of KP and WFP is strongly supported.

383 Demographic history reconstruction of KP and WFP revealed distinct evolutionary histories of the two
384 species from approximately 10 Ma to 10 ka (Figure 3c). PSMC analysis demonstrated a similar *Ne*
385 history for both species around 1 million years ago during the glaciation, with population sizes of both
386 species rising from 0.8 to 1 million years ago then sharply declining until about 100 thousand years ago
387 during the Last Interglacial Period. KP and WFP then went through steady population size changes
388 separately. The *Ne* of KP increased greatly to 1.65 million and the *Ne* of WFP decreased to about ten
389 thousand.

390 With ABC simulations, we found that the genomic-wide polymorphism patterns in KP and WFP fit best
391 with the secondary contact model (posterior probability = 0.98, Supplementary Table 17), suggesting
392 that the two plover species experienced gene flow after secondary contact. Model selection showed

that changing *Ne* models incorporating the PSMC-inferred Ne fluctuations had much higher posterior possibilities than constant *Ne* models (Bayes factors > 10^3), which indicates that the secondary contact model based on PSMC results fitted best (Figure 2).

Demographic analyses allowed us to estimate several demographic parameter estimates, including divergence times, effective population sizes and migration rate per generation (Table 1). The two plover populations are estimated to have diverged approximately 606 000 years ago (95% CI, 277 000 ~974 000 years). The effective population size of KP is estimated to be around 2.6 times than that of WFP (median Ne_K = 296,800 and median Ne_W = 114,500). The most recent common ancestor is believed to have an effective population size, Ne_A, which is about 10-15 times greater than both modern species. Gene flow between the two groups of plover was approximately symmetric: c. 7.85 individuals per generation immigrate from WFP to KP, compared with c. 3.21 individuals per generation immigrating from KP to WFP. It must be noted however, that the ranges of the probability distributions for migration estimates are broad (Supplementary Figure 10).

Genomic regions associated with divergence between Kentish plover and White-faced Plover

Genome scans showed low genome-wide divergence between KP and WFP (Figure 4). The genome-wide F_{ST} was 0.046, π of KP was 0.00262 and π of WFP was 0.00259. 24,065 blocks of 50kbp length were scanned, 22,148 of which were located on autosomes, 1,622 were located on Z chromosome and 295 unassigned. Since the W chromosome is very short (5.16Mb in chicken genome) and only males were used for population analyses, it was not included in the genomic landscape analysis. Examining autosomes and the Z chromosome separately, it was found that the average F_{ST} of autosomal blocks was 0.043, π of KP was 0.00268 and π of WFP was 0.00266. The top 1% outlier blocks on autosomes, which have the highest F_{ST} (average 0.230, peak 0.605), were found to have much lower polymorphism, than when examined at the genome-wide scale (π_{KP} = 0.0012, π_{WFP} = 0.0011, $p < 0.001$). The Z chromosome was more divergent than autosomes (average F_{ST} = 0.089, π_{KP} = 0.0020 and π_{WFP} =

417 0.0019). The highest 1% outlier blocks on the Z chromosome had an average F_{ST} of was 0.664, and the
418 peak value was 0.741. The average π was 0.0003.

419 We identified genes which had high levels of divergence between species, by calculating population
420 statistics per gene (Figure 4). 691 autosomal genes had F_{ST} higher than 0.184 and 41 genes on the Z
421 chromosome had F_{ST} higher than 0.262 (Supplementary Table 18). GO enrichment analyses of
422 autosomal genes with high F_{ST} found that no categories were enriched for biological processes and
423 integral component of membrane was enriched for cellular components. No categories were enriched
424 when looking at the genes with high F_{ST} on the Z chromosome. A total of 339 autosomal genes were
425 found with a *Tajima's D* ≥ 1.565 (Supplementary Table 19) and 235 genes with *Tajima's D* ≤ -1.179 for
426 KP (Supplementary Table 20). No GOs were enriched after GO enrichment analysis. 325 genes had
427 values of *Tajima's D* ≥ 1.669 (Supplementary Table 21) and 216 had values of *Tajima's D* ≤ -1.191 for
428 WFP (Supplementary Table 22). GO enrichment analysis revealed an overrepresentation of high
429 *Tajima's D* genes associated with microtubule cellular component categories and genes with low
430 *Tajima's D* had an overrepresentation of genes associated with proteolysis for biological function
431 categories. For KP 18 genes on the Z chromosome had a *Tajima's D* ≥ 1.723 and 12 genes had *Tajima's*
432 *D* ≤ -1.234. GO enrichment was not performed due to the small number of high diversity genes on the
433 Z chromosome. WFP had 19 genes on the Z chromosome had *Tajima's D* ≥ 1.835 and 10 genes had
434 *Tajima's D* ≤ -1.445.

435

436 **Discussion**

437 We have utilised methods of mitochondria and nuclear whole genome *de novo* sequencing to shed
438 unprecedented insight into the evolution and demographic histories of two small plover species.
439 Although previous studies suggest KP and WFP are sufficiently differentiated at a phenotypic level to
440 justify them as two different species [42, 44], genetic studies have not always agreed [42]. Our results

441 show that KPs and WFPs have low levels of differentiation on average at the genomic level, with
442 moderate to high differentiation in some regions. Our divergence time estimates suggest that Kentish
443 and white-faced plovers diverged relatively recently, less than 0.61 million years ago.

444 Using genome wide data, we were able to model the demographic histories of the two plovers. By
445 using more complex demographic models it was possible to model demographic histories more
446 comprehensibly than would be possible using traditional population genetics techniques. ABC analysis
447 estimates the divergence time to be about 606,000 years ago during the Pleistocene. Estimates of
448 Ancestral N_e suggests that the most recent common ancestor had a much higher N_e (721,500) than
449 either modern species. Current N_e estimates for KP (296,800) were about 2.5 times that that of WFP
450 (114,500). N_e, can be important as an estimate of the health of a population in terms of conservation
451 biology. This can be particularly valuable for species where census data is lacking, such as with the WFP
452 which has no census data available from the IUCN Red List, although the relationship between N_e and
453 census N can be difficult to interpret.

454 We found contrasting demographic histories between the two plover species as demonstrated from
455 the PSMC estimates, especially through their history from 1 million to approximately 100 thousand
456 years ago. The N_e of KP was about seven times larger than WFP towards the end of the last glacial
457 period (LGP) resulting from population increases in KP and declines in WFP. Although declines in
458 population size at the start of the LGP is a common pattern in bird species [72], marked contrasting
459 patterns in population demography may reflect species differences in response to historical climate
460 fluctuation and also population divergence [83]. One possibility is that a decrease in suitable habitat
461 for WFP during the LGP could explain continued declines of the N_e of WFP whereas the KP may have
462 been able to exploit a wider range of habitats. Caution should be exercised when implementing
463 demographic approaches such as ABC and PSMC and when interpreting the results, due to heavy
464 dependence on the parameters and priors [27, 84].

465 It is unknown whether the breeding grounds of KP and WFP overlap or whether any form of
466 reproductive isolation has occurred. If a hybrid zone exists it has been estimated that this would be
467 found in a narrow range in Fujian Province [42]. It is believed that advanced reproductive isolation
468 occurs relatively late during speciation events in birds, with complete F_1 hybrid sterility often taking in
469 the order of millions of years [27, but see 85]. The recent split of the two plover species would suggest
470 that reproductive isolation may be limited. The rate of gene flow and recombination occurring
471 throughout the genome can affect the rate at which reproductive isolation occurs. Migratory
472 shorebirds are highly mobile with the potential to disperse large distances to breed, this can lead to
473 high levels of gene flow and weak population structuring in some species [41, 86–88]. We estimated
474 levels of gene flow throughout the history of the two species to determine whether historic gene flow
475 occurred as a result of secondary contact or alternatively that gene flow has occurred continuously as
476 the species diverged. Various scenarios of gene flow were modelled using ABC analysis. ABC
477 simulations best fit the secondary contact model, in which the two plover species diverged
478 allopatrically, with gene flow occurring after secondary contact. There was little support for models of
479 isolation with migration, isolation, or early migration models. Gene flow was found to be
480 approximately symmetric and bidirectional between the two species. How both phenotypic and
481 genomic differentiation is maintained in spite of gene flow is a key question in evolutionary biology.
482 We report levels of gene flow with more than five individuals introgressing between the two species
483 per generation. It is expected that very few individuals are needed to exchange genes between
484 populations to break down differentiation produced by genetic drift [89, 90]. This suggests that some
485 form of selection is acting on these two species to maintain this divergence within the populations.
486 Further study of these species at the hybrid zone would help elucidate which selective forces might be
487 maintaining these differences.

488 Despite secondary gene flow, we detected highly differentiated genomic regions that may contribute
489 to species divergence [18, 22, 91–93]. We used a window based approach to perform whole genome

490 scans and calculate various population statistics, including F_{ST}, π and *Tajima's D*, to detect areas of high
491 divergence and selection. We found that average F_{ST} across genome was 0.046. This is slightly higher
492 than levels found between carrion and hooded crows, another taxonomically debated pair of species
493 [18]. It has been suggested that areas of peak divergence contain genes involved in reproductive
494 isolation and that these areas can contain genes responsible for differences in phenotype [18–22].
495 Higher levels of F_{ST} and *Tajima's D* in KP were found in the PPP3CB gene, part of the oocyte meiosis
496 KEGG pathway and could therefore be important for promoting genetic incompatibility in females.
497 Hybrid incompatibility often the heterogametic sex due to Haldane's Rule [94], which often means in
498 birds that hybrid females are sterile and fertile males allow gene flow to occur between species [95].
499 The Z chromosome had higher levels of block-average and peak divergence, 0.089 and 0.741
500 respectively. This pattern is consistent with results from previous studies [18, e.g. 22]. Mean
501 divergence levels of sex chromosomes are often higher than on the autosomes. This could be caused
502 by the lower *Ne* of the Z chromosome compared with the autosomes, which can result in faster lineage
503 sorting and higher F_{ST} [9]. There is often an overabundance of highly differentiated loci on the sex
504 chromosomes [1]. In light of this pattern, Cruickshank and Hahn [96] proposed that speciation is not
505 necessarily the sole force contributing to divergence in genomic regions. Population bottlenecks, i.e.
506 in WFP (Figure 2c), and divergence in regions of low *Ne*, i.e. Z chromosome, can also contribute to
507 heterogeneity of genomic differentiation [9].

508 This study also provides a greater understanding about the demographic histories of two shorebird
509 species, especially linking their current population status with evolutionary context [29, 30, 97].
510 Although KP have a large census population size and are widely distributed throughout Eurasia and
511 North Africa, there is evidence that populations are in decline in East Asia. The decline in Chinese
512 plovers could, at least in part, be due to a reduction in suitable breeding and feeding locations along
513 the Chinese coast due to land reclamation and development [98]. Extremely low nest survival in
514 Kentish plovers from Bohai Bay has been reported and linked to anthropogenic disturbance [99]. This

515 work also emphasises the lower effective population size of WFP when compare to KP, which has
516 resulted from population declines since the LGM (Figure 2c). Thus, increased effort to monitor
517 population trends for this species is warranted in order to accurately assess any potential threats to
518 this species and for conservation status evaluation for the IUCN.

519 In conclusion, we produced the first high quality genome of the KP and performed whole genome
520 resequencing of two plover species relatively early in their divergence. We found multiple pieces of
521 evidence to support that the WFP and KP are distinct lineages with complex demographic histories.
522 We found evidence for gene flow between these two species due to secondary contact. Our results
523 further reveal a heterogeneous pattern of genomic differentiation with elevated divergence in the Z
524 chromosome. This suggests that some form of selection is working to maintain genetic and phenotypic
525 differences between the two species. Overall, this study provides new insights into the genomic
526 patterns between a species pair at an early stage of speciation. Further analyses of populations at the
527 hybrid zone would increase our understanding of the specific selective forces maintaining this
528 divergence.

529

530 **Acknowledgements**

531 The authors thank Qiaoyi Liang, Zhechun Zhang, Xuecong Zhang, Xin Lin and Demeng Jiang for their
532 assistance during sampling, and Shaochong Peng and Chunfa Zhou for their help in the preparation of
533 Figure 1. This work was supported by National Natural Science Foundation of China (31301875,
534 31572251 to YL, and 31600297 to PJQ) and Special Program for Applied Research on Super
535 Computation of the NSFC-Guangdong Joint Fund (the second phase) under Grant No. U1501501; a
536 National Environment Research Council Great Western Four+ Doctoral Training Partnership
537 studentship (grant number NE/L002434/1), Korner Travelling Award and a British Council and Chinese
538 Scholarship Council Newton Fund PhD Placement awarded to KHM; a Royal Society Dorothy Hodgkin

539 Research Fellowship (grant number DH071902), Royal Society research grant (grant number
540 RG0870644), a Royal Society research grant for fellows (grant number RG080272) and a NERC grant
541 (NE/P004121/1) to AOU. All sequencing data will be deposited in NCBI databases upon acceptance.

542

543 **References**

544 1. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. Mol Ecol.
545 2016;25:2337–60.

546 2. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-
547 gene-flow. Philos Trans R Soc B Biol Sci. 2012;367:451–60.

548 3. Wu CI. The genic view of the process of speciation. J Evol Biol. 2001;14:851–65.

549 4. Lindtke D, Buerkle CA. The genetic architecture of hybrid incompatibilities and their effect on
550 barriers to introgression in secondary contact. Evolution. 2015;69:1987–2004.

551 5. Bolnick DI, Fitzpatrick BM. Sympatric speciation: models and empirical evidence. Annu Rev Ecol Evol
552 Syst. 2007;38:459–87.

553 6. Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ. Sympatric speciation in the post "modern
554 synthesis" era of evolutionary biology. Evol Biol. 2012;39:158–80.

555 7. Shaner PJL, Tsao TH, Lin RC, Liang W, Yeh CF, Yang XJ, et al. Climate niche differentiation between
556 two passerines despite ongoing gene flow. J Anim Ecol. 2015;84:829–39.

557 8. Wang P, Liu Y, Liu Y, Chang Y, Wang N, Zhang Z. The role of niche divergence and geographic
558 arrangement in the speciation of Eared Pheasants (*Crossoptilon*, Hodgson 1938). Mol Phylogenet Evol.
559 2017;113:1–8.

560 9. Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. Nat
561 Rev Genet. 2017;18:87–100.

562 10. Seehausen O. Conservation: losing biodiversity by reverse speciation. Curr Biol. 2006;16:R334–7.

563 11. Taylor EB, Boughman JW, Groenenboom M, Sniatynski M, Schluter D, Gow JL. Speciation in reverse:
564 morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus*
565 *aculeatus*) species pair. Mol Ecol. 2006;15:343–55.

566 12. Webb WC, Marzluff JM, Omland KE. Random interbreeding between cryptic lineages of the
567 common raven: evidence for speciation in reverse. Mol Ecol. 2011;20:2390–402.

568 13. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. Trends Genet. 2012;28:342–
569 50.

570 14. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide
571 evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 2013;23:1817–28.

572 15. Nosil P, Sandoval CP. Ecological niche dimensionality and the evolutionary diversification of stick
573 insects. PLoS One. 2008;3:e1907.

574 16. Merrill RM, Gompert Z, Dembeck LM, Kronforst MR, McMillan WO, Jiggins CD. Mate preference
575 across the speciation continuum in a clade of mimetic butterflies. Evolution. 2011;65:1489–500.

576 17. McLean CA, Stuart-Fox D. Geographic variation in animal colour polymorphisms and its role in
577 speciation. Biol Rev. 2014;89:860–73.

578 18. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, et al. The genomic landscape underlying
579 phenotypic integrity in the face of gene flow in crows. Science. 2014;344:1410–4.

580 19. Turner TL, Hahn MW, Nuzhdin S V. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol.
581 2005;3:1572–8.

582 20. Harr B. Genomic islands of differentiation between house mouse subspecies. Genome Res.
583 2006;16:730–7.

584 21. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence.
585 Mol Ecol. 2009;18:375–402.

586 22. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of
587 species divergence in *Ficedula* flycatchers. Nature. 2012;491:756–60.

588 23. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands
589 of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philos
590 Trans R Soc B Biol Sci. 2012;367:343–53.

591 24. Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, et al. Genome-wide patterns
592 of divergence during speciation: the lake whitefish case study. Philos Trans R Soc B Biol Sci.
593 2012;367:354–63.

594 25. Roux C, Tsagkogeorga G, Bierne N, Galtier N. Crossing the species barrier: genomic hotspots of
595 introgression between two highly divergent *Ciona intestinalis* species. Mol Biol Evol. 2013;30:1574–
596 87.

597 26. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, et al. Genomic islands of
598 speciation separate cichlid ecomorphs in an East African crater lake. Science. 2015;350:1493–8.

599 27. Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds L, Ellegren H. Demographic
600 divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-
601 sequencing data. PLoS Genet. 2013;9:e1003942.

602 28. Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al. Plumage genes and
603 little else distinguish the genomes of hybridizing warblers. Curr Biol. 2016;26:2313–8.

604 29. Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation
605 guidelines. Trends Ecol Evol. 2001;16:613–22.

606 30. Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC. Locally adapted traits
607 maintained in the face of high gene flow. Ecol Lett. 2015;18:37–47.

608 31. Beaumont MA. Approximate Bayesian computation in evolution and ecology. Annu Rev Ecol Evol
609 Syst. 2010;41:379–406.

610 32. Li H, Durbin R. Inference of human population history from individual whole-genome sequences.
611 Nature. 2011;475:493–6.

612 33. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol
613 Evol. 2014;29:51–63.

614 34. Kosztolányi A, Székely T, Cuthill IC, Yilmaz KT, Berberoglu S. Ecological constraints on breeding
615 system evolution: the influence of habitat on brood desertion in Kentish plover. J Anim Ecol.
616 2006;75:257–65.

617 35. Kosztolányi A, Javed S, Küpper C, Cuthill IC, Al Shamsi A, Székely T. Breeding ecology of Kentish
618 plover *Charadrius alexandrinus* in an extremely hot environment. Bird Study. 2009;56:244–52.

619 36. Székely T. Sexual conflict between parents: offspring desertion and asymmetrical parental care. In:
620 Rice WR, Gavrilets S, editors. The Genetics and Biology of Sexual Conflict. Cold Spring Harbor
621 Laboratory Press; 2014. p. 245–63.

622 37. Bulla M, Valcu M, Dokter AM, Dondua AG, Kosztolányi A, Rutten AL, et al. Unexpected diversity in
623 socially synchronized rhythms of shorebirds. Nature. 2016;540:109–13.

624 38. Eberhart-Phillips LJ, Küpper C, Miller TEX, Cruz-López M, Maher KH, dos Remedios N, et al. Sex-
625 specific early survival drives adult sex ratio bias in snowy plovers and impacts mating system and
626 population growth. Proc Natl Acad Sci. 2017;114:E5474–E5481.

627 39. Maher KH, Eberhart-Phillips LJ, Kosztolányi A, Remedios N dos, Carmona-Isunza MC, Cruz-López M,
628 et al. High fidelity: extra-pair fertilisations in eight *Charadrius* plover species are not associated with
629 parental relatedness or social mating system. J Avian Biol. 2017;48:910–920.

630 40. del Hoyo J, Elliott A, Sargatal J. Handbook of the Birds of the World - Volume 3: Hoatzin to Auks.
631 Lynx Edicions; 1996.

632 41. Küpper C, Edwards S V., Kosztolányi A, Alrashidi M, Burke T, Herrmann P, et al. High gene flow on
633 a continental scale in the polyandrous Kentish plover *Charadrius alexandrinus*. Mol Ecol.
634 2012;21:5864–79.

635 42. Rheindt FE, Székely T, Edwards S V., Lee PLM, Burke T, Kennerley RP, et al. Conflict between genetic
636 and phenotypic differentiation: the evolutionary history of a 'lost and rediscovered' shorebird. PLoS
637 One. 2011;6:e26995.

638 43. Wang X, Que P, Heckel G, Hu J, Zhang X, Chiang C-Y, et al. Unveiling divergence and demographic
639 histories of two *Charadrius* plovers along the Chinese coast. Submitted.

640 44. Kennerley PR, Bakewell DN, Round PD. Rediscovery of a long-lost *Charadrius* plover from South-
641 East Asia. Forktail. 2008;24:63–79.

642 45. Székely T, Argüelles-Ticó A, Kosztolányi A, Küpper C. Practical guide for investigating breeding
643 ecology of Kentish plover *Charadrius alexandrinus*. Univ of Bath. 2011.

644 46. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-
645 efficient short-read *de novo* assembler. Gigascience. 2012;1:18.

646 47. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform.
647 Bioinformatics. 2009;25:1754–60.

648 48. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome
649 assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.

650 49. Xu Z, Wang H. LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.
651 Nucleic Acids Res. 2007;35:W265–8.

652 50. Kent WJ. BLAT - The BLAST-like alignment tool. Genome Res. 2002;12:656–64.

653 51. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988–95.

654 52. Salamov AA, Solovyev V V. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res.
655 2000;10:516–22.

656 53. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel.
657 Bioinformatics. 2003;19:ii215-225.

658 54. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
659 Bioinformatics. 2009;25:1105–11.

660 55. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly
661 and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among
662 isoforms. Nat Biotechnol. 2010;28:511–5.

663 56. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure
664 annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol.
665 2008;9:R7.

666 57. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in
667 genomic sequence. Nucleic Acids Res. 1997;25:955–64.

668 58. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-
669 coding RNAs in complete genomes. Nucleic Acids Res. 2005;33:D121–4.

670 59. Zhang G, Li B, Li C, Gilbert MTP, Jarvis ED, Wang J, et al. Comparative genomic data of the Avian
671 Phylogenomics Project. Gigascience. 2014;3:26.

672 60. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early
673 branches in the tree of life of modern birds. Science. 2014;346:1320–31.

674 61. Li H, Coghlan A, Ruan J, Coin LJ, Hériché J-K, Osmotherly L, et al. TreeFam: a curated database of
675 phylogenetic trees of animal gene families. Nucleic Acids Res. 2006;34:D572–80.

676 62. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
677 Acids Res. 2004;32:1792–7.

678 63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
679 phylogenies. Bioinformatics. 2014;30:1312–3.

680 64. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

681 65. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation
682 discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

683 66. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and
684 VCFtools. Bioinformatics. 2011;27:2156–8.

685 67. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient
686 Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–
687 542.

688 68. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large
689 phylogenetic trees. In: Gateway Computing Environments Workshop (GCE), 2010. New Orleans, LA;
690 2010. p. 1–8.

691 69. Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: more models, new heuristics and parallel
692 computing. Nat Methods. 2012;9:772.

693 70. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in
694 phylogenetic analysis. Syst Biol. 1993;42:182–92.

695 71. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
696 individuals. Genome Res. 2009;19:1655–64.

697 72. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal dynamics of avian
698 populations during pleistocene revealed by whole-genome sequences. Curr Biol. 2015;25:1375–80.

699 73. Pavlidis P, Laurent S, Stephan W. msABC: a modification of Hudson's ms to facilitate multi-locus
700 ABC analysis. Mol Ecol Resour. 2010;10:723–7.

701 74. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
702 Genetics. 1989;123:585–95.

703 75. Csilléry K, François O, Blum MGB. abc: an R package for approximate Bayesian computation (ABC).
704 Methods Ecol Evol. 2012;3:475–9.

705 76. Blum MGB, François O. Non-linear regression models for Approximate Bayesian Computation. Stat
706 Comput. 2010;20:63–73.

707 77. Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F, et al. Genome-wide synteny
708 through highly sensitive sequence alignment: *Satsuma*. Bioinformatics. 2010;26:1145–51.

709 78. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient swiss army knife
710 for population genomic analyses in R. Mol Biol Evol. 2014;31:1929–36.

711 79. R Core Team. R: a language and environment for statistical computing. R Found Stat Comput
712 Vienna, Austria. 2016.

713 80. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol.
714 1990;215:403–10.

715 81. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput
716 functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36:3420–35.

717 82. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein
718 function classification. Bioinformatics. 2014;30:1236–40.

719 83. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of
720 pig genomes provide insight into porcine demography and evolution. Nature. 2012;491:393–8.

721 84. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes
722 in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol Ecol.
723 2016;25:1058–72.

724 85. Blundell GM, Ben-David M, Groves P, Bowyer RT, Geffen E. Characteristics of sex-biased dispersal
725 and gene flow in coastal river otters: Implications for natural recolonization of extirpated populations.
726 Mol Ecol. 2002;11:289–303.

727 86. Verkuil YI, Piersma T, Jukema J, Hooijmeijer JCEW, Zwarts L, Baker AJ. The interplay between habitat
728 availability and population differentiation: a case study on genetic and morphological structure in an
729 inland wader (Charadriiformes). Biol J Linn Soc. 2012;106:641–56.

730 87. Eberhart-Phillips LJ, Hoffman JI, Brede EG, Zefania S, Kamrad MJ, Székely T, et al. Contrasting
731 genetic diversity and population structure among three sympatric Madagascan shorebirds: parallels
732 with rarity, endemism, and dispersal. Ecol Evol. 2015;5:997–1010.

733 88. D'Urban Jackson J, dos Remedios N, Maher KH, Zefania S, Haig S, Oyler-McCance S, et al. Polygamy
734 slows down population divergence in shorebirds. Evolution. 2017;71:1313–26.

735 89. Wright S. Evolution in Mendelian populations. Genetics. 1931;16:97–159.

736 90. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987;236:787–
737 92.

738 91. Rogers SM, Bernatchez L. The genetic architecture of ecological speciation and the association with
739 signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. Mol Biol

740 Evol. 2007;24:1423–38.

741 92. Barrett RDH, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev
742 Genet. 2011;12:767–80.

743 93. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis
744 of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188:379–97.

745 94. Haldane JBS. Sex ratio and unisexual sterility in hybrid animals. J Genet. 1922;12:101–9.

746 95. Mořkovský L, Janoušek V, Reif J, Rídl J, Pačes J, Choleva L, et al. Genomic islands of differentiation
747 in two songbird species reveal candidate genes for hybrid female sterility. Mol Ecol. 2018;27:949–58.

748 96. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to
749 reduced diversity, not reduced gene flow. Mol Ecol. 2014;23:3133–57.

750 97. Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology:
751 advantages, limitations and practical recommendations. Mol Ecol. 2017;26:5369–406.

752 98. Ma Z, Melville DS, Liu J, Chen Y, Yang H, Ren W, et al. Rethinking China's new great wall. Science.
753 2014;346:912–4.

754 99. Que P, Chang Y, Eberhart-Phillips L, Liu Y, Székely T, Zhang Z. Low nest survival of a breeding
755 shorebird in Bohai Bay, China. J Ornithol. 2015;156:297–307.

756

757 **Data Availability**

758 This information will become available upon the acceptance of the manuscript.

759

760 **Competing interests**

761 The authors declare that they have no competing interests.

762

763 **Author Contributions**

764 YL and TS conceived of the study. XJW and YL designed the study. PQ, QH, BW, ZWZ and YL collected
765 the samples. XW, NZ and KHM, CZ, SL, BW, DC, XY analysed data. KHM, AOU and YL wrote the
766 manuscript with contributions from all authors. All authors read and approved the final version of the
767 manuscript.

768

769

770 **Supplementary Material**

771 Supplementary information, Supplementary Figures 1-10 and Supplementary Tables S1-S22 are

772 available online.

773

774 **Table 1.** Posterior median, mean, mode and range of 95% highest probability distribution (HPD) of
775 demographic parameters. K represents Kentish Plover *Charadrius alexandrinus* and W represents
776 White-faced Plover *C. dealbatus*. A represents Ancestral population. Ne: recent effective population
777 size; T: population split time; M: migration rate per generation.

	$T/10^4$	$Ne_K/10^4$	$Ne_W/10^4$	$Ne_A/10^4$	$2NM_{W->K}$	$2NM_{K->W}$
Median	60.60	29.68	11.45	72.15	7.85	3.21
Mean	61.20	137.09	20.02	98.71	7.62	3.11
Mode	54.10	16.01	5.62	26.74	7.31	3.27
2.5% HPD	27.70	5.22	3.33	13.35	3.07	1.84
97.5% HPD	97.40	1262.12	99.91	416.42	11.32	4.03

778

779 **Figure 1.** Sampling locations of two plover species, Kentish plover *Charadrius alexandrinus* and white-
780 faced plover *C. dealbatus.* The red triangle represents the location where one individual of Kentish
781 plover for *de novo* sequencing was collected.

782 **Figure 2.** Illustration of the models simulated in ABC analysis. Eight models in two groups were
783 simulated. Effective population sizes in group A were constant. Effective population sizes in group B
784 were based on Ne changes in PSMC. Judged from PSMC, the divergence time was not early than the
785 beginning of population declines 1 million years ago. To simplify the models, population size shifts
786 and changes of gene flow were set to the same time point (*T1*). Prior ranges are available in Table
787 S17.

788 **Figure 3.** Population genetic structure and historical demography. *C. alexandrinus* marked in blue and
789 *C. dealbatus* in yellow. a) Genetic clustering inferred with ADMIXTURE when K=2. b) Phylogenetic
790 relationship between the *C. dealbatus* (WFP) and different populations of *C. alexandrinus* (KP) using
791 Bayesian and Maximum Likelihood methods based on mitochondrial genome sequences (c.a. 15kb).
792 Posterior probabilities (pp) and bootstrap supports are indicated at each node. White-faced plover and
793 Kentish plover form two independent evolutionary lineages. c) Demographic history of the Kentish
794 plover, blue line, and white-faced plover, yellow line reconstructed from the reference and population
795 resequencing genomes. The line represents the estimated effective population size (*N*e), and the 100
796 thin blue curves represent the PSMC estimates for 100 sequences randomly resampled from the
797 original sequence. Generation time (*g*) = 2.5 years, and neutral mutation rate per generation (μ) = 8.11
798 × 10^{-10}. The Last Interglacial period (LIG, from approx. 130 to 116 ka) is marked by a grey block.

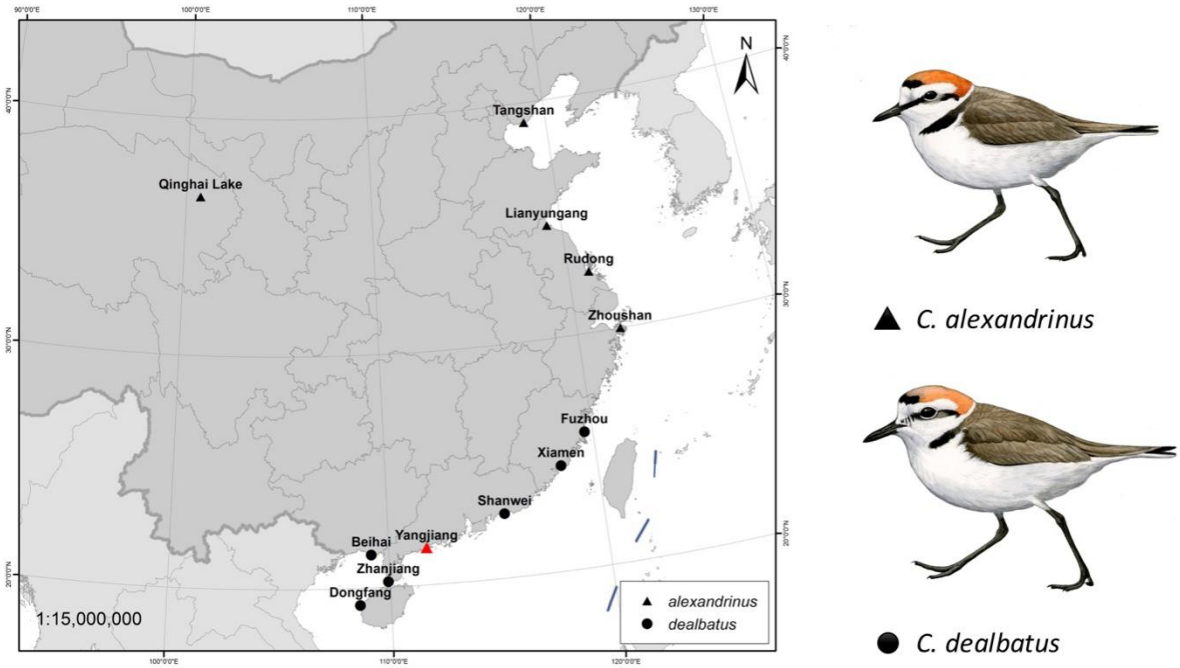
799 **Figure 4.** Genome wide landscape of F_{ST}, π, and *Tajima's D* for 50 kb sliding window. Different
800 autosomes are marked with alternating light and dark colors. **a**) Genome wide landscape for
801 autosomes. 95th percentile outliers are plotted for F_{ST} and 95th/5th percentile outliers are plotted for
802 *Tajima's D* in grey, as calculated per gene. 99.9th percentile outliers for F_{ST} and 99.9th/0.1th percentile
803 outliers for *Tajima's D* are plotted in black per gene and labelled with gene symbols. **b**) Genome wide
804 landscape for Z chromosome. 95th percentile outliers are plotted for F_{ST} and 95th/5th percentile outliers
805 are plotted for *Tajima's D* in grey, as calculated per gene. 99th percentile outliers for F_{ST} and 99th/1st
806 percentile outliers for *Tajima's D* are plotted in black per gene and labelled with gene symbols.

807

808

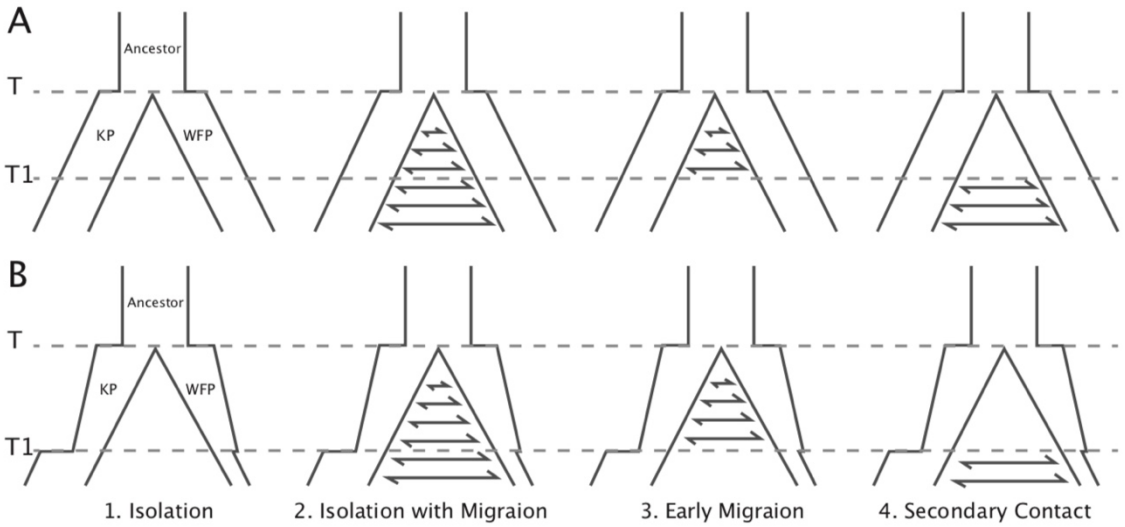
809 **Figure 1**

810



811
812
813

814 **Figure 2**



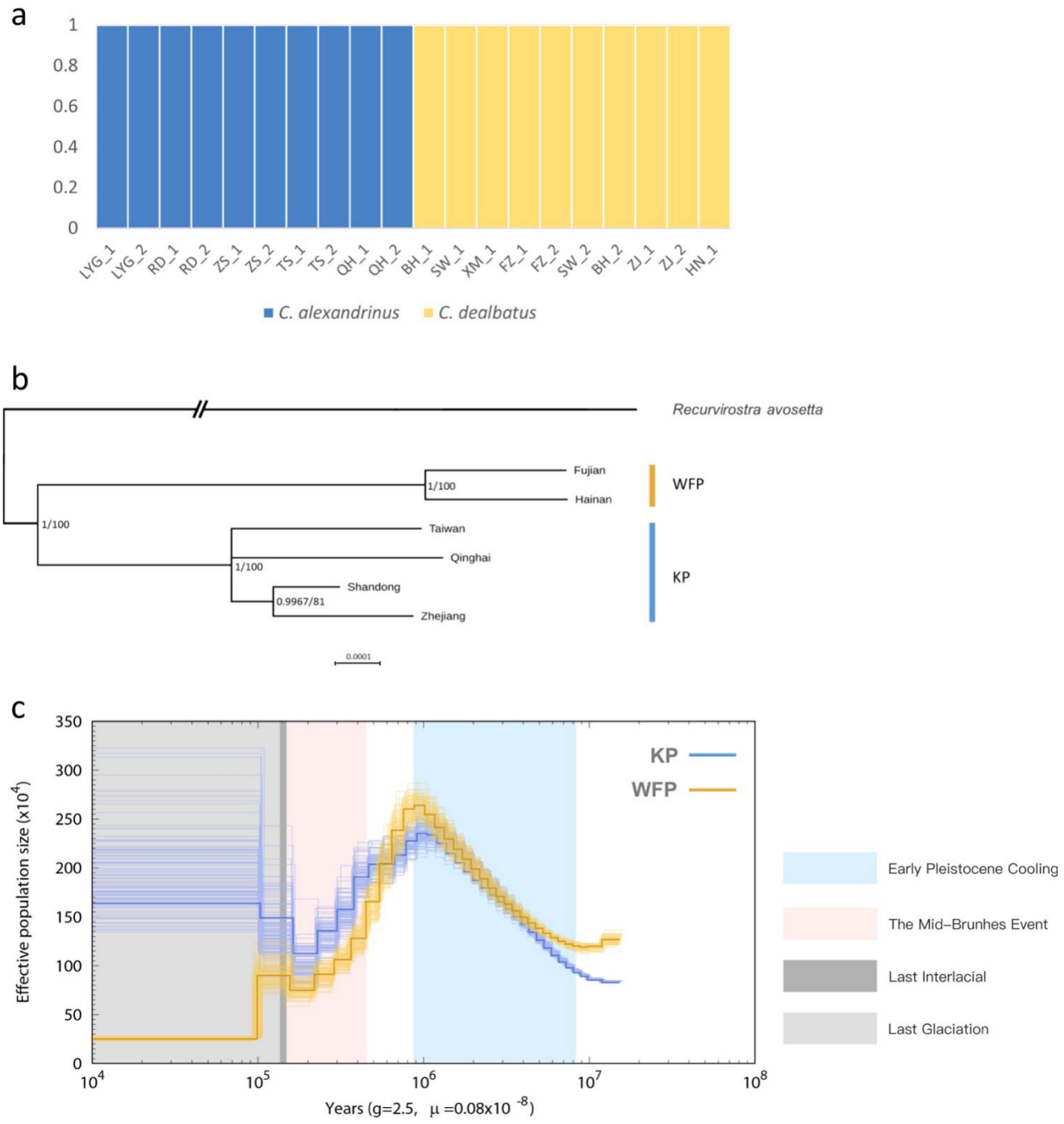
815
816
817
818

819

820 **Figure 3**

821

822



823

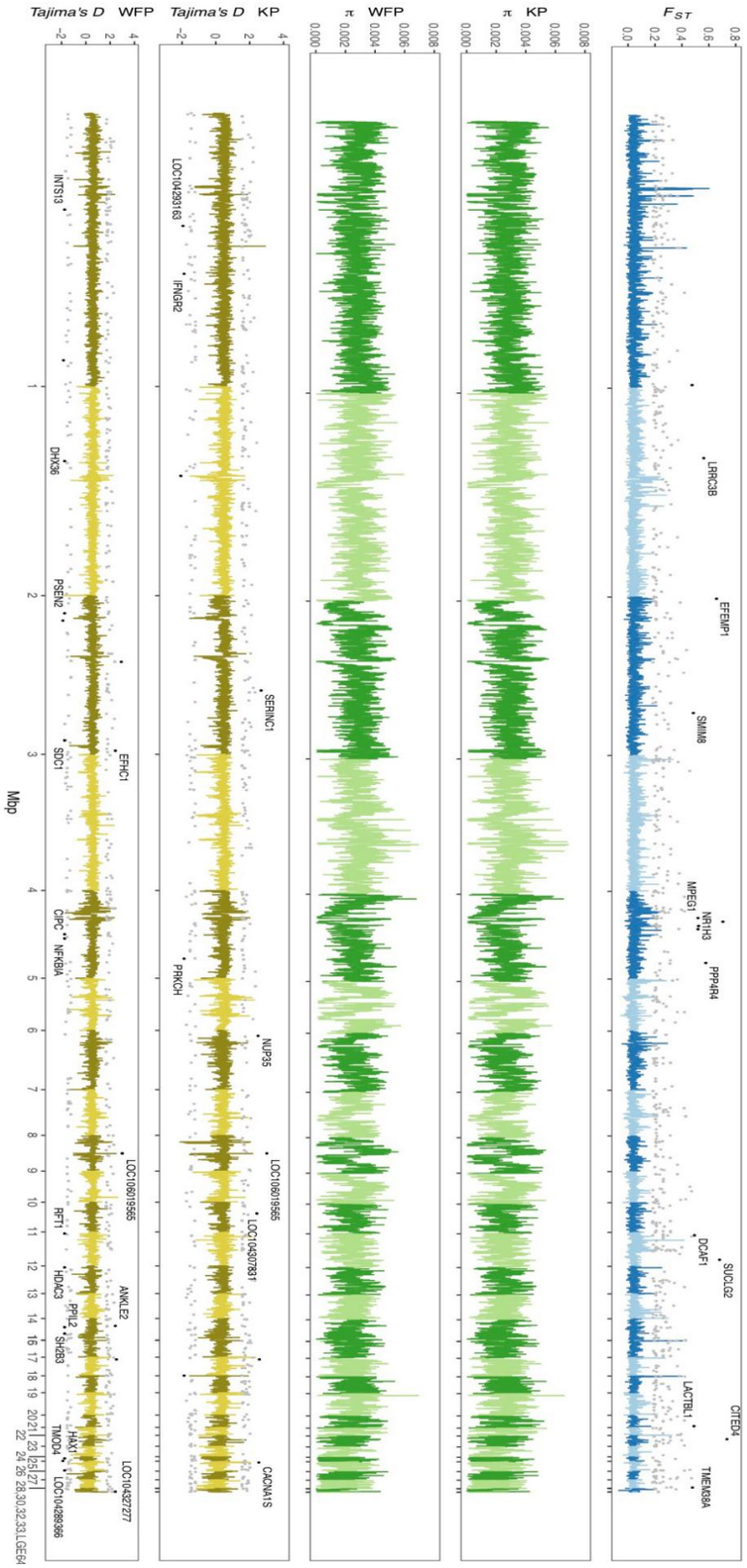
824

825

826

827 **Figure 4a**

828



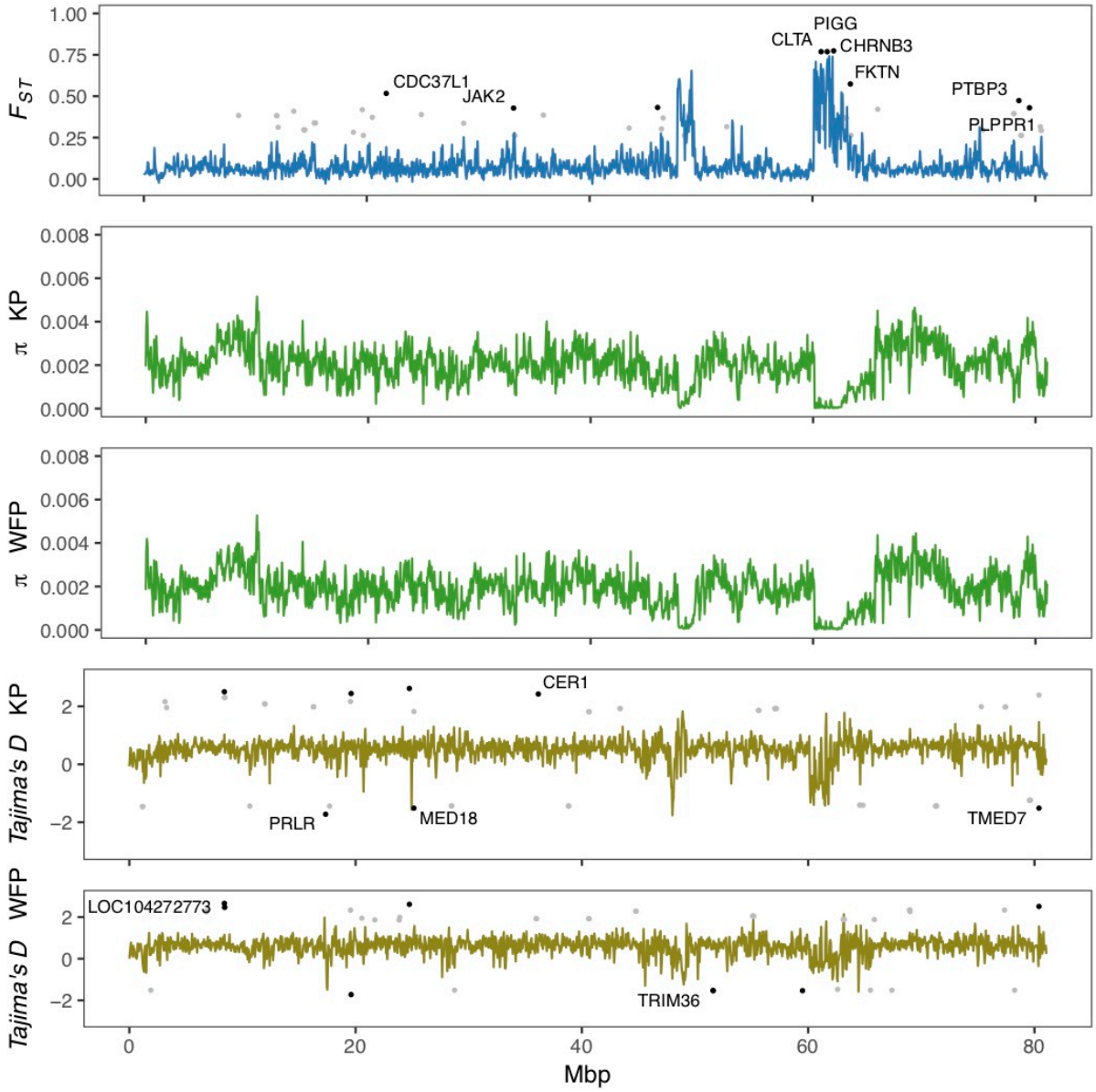
829

830

831

832 **Figure 4b**

833



834