1

2	Divergent selection in low recombination regions shapes the genomic islands in two	
3	incipient shorebird species	

4

Wenjun Zhou<sup>1,2</sup>, Nan Zhang<sup>1</sup>, Kaichi Huang<sup>3</sup>, Hongzhou Lin<sup>1</sup>, Jie Tu<sup>1</sup>, Chenqing Zheng<sup>1</sup>, Pinjia
Que<sup>4,5</sup>, Chung-Yu Chiang<sup>6</sup>, Jonathan Martinez<sup>7</sup>, Halimubieke Naerhulan<sup>8</sup>, Tamás Székely<sup>1,4,8</sup>,
Zhengwang Zhang<sup>4</sup>, Yang Liu<sup>1,8\*</sup>

- 1. State Key Laboratory of Biocontrol, School of Ecology/School of Life Sciences, Sun 8 9 Yat-sen University, Guangzhou, China 10 2. Department of Zoology, University of Cambridge, Cambridge, UK 3. Department of Botany and Biodiversity Research Centre, University of British 11 12 Columbia, Vancouver, BC, Canada 4. Ministry of Education Key Laboratory for Biodiversity Sciences and Ecological 13 Engineering, College of Life Sciences, Beijing Normal University, Beijing, China 14 5. Sichuan Key Laboratory of Conservation Biology for Endangered Wildlife, Chengdu 15 Research Base of Giant Panda Breeding, Sichuan Academy of Giant Panda, Chengdu, 16 China 17 18 6. Department of Environmental Science, Tunhai University, Taichun, Taiwan, China 7. 14, bis rue des Temples 45240 La Ferté Saint Aubin, France 19
- 8. Milner Centre for Evolution, Department of Biology and Biochemistry, University of
   Bath, Bath, UK
- 22

23 \*Corresponding author: E-mail: liuy353@mail.sysu.edu.cn

24

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com 1

#### 1 Abstract

2 Divergent selection in the face of gene flow is usually associated with a heterogeneous 3 genomic landscape of divergence in nascent species pairs. However, multiple factors, such 4 as divergent selection and local recombination rate variation, can influence the formation of these genomic island. This conundrum can be solved through examination of the genomic 5 landscapes of species pairs that are still in the early stages of speciation. In this study, 6 7 population genomics analyses were undertaken using a wide range of sampling and wholegenome resequencing data from 96 unrelated individuals of Kentish plover (Charadrius 8 9 alexandrinus) and white-faced plover (C. dealbatus). We suggest that the two species exhibit 10 varying levels of population admixture along the Chinese coast and on Taiwan Island. Genome-wide analyses for introgression indicate that ancient introgression had occurred in 11 Taiwan population, and recurrent gene flow is still ongoing in mainland coastal populations. 12 Furthermore, we identified a few genomic regions with significant levels of interspecific 13 differentiation and local recombination suppression, which contain several genes potentially 14 associated with disease resistance, coloration, and regulation of plumage moulting, thus 15 may be connected to the phenotypic and ecological divergence of the two nascent species. 16 Overall, our findings suggest that divergent selection in low recombination regions may be 17 the main force in shaping the genomic islands in two incipient shorebird species. 18

Keywords: population genomics, incipient speciation, genomic islands, gene flow, contact
zone

#### 1 Introduction

Heterogeneous genomic landscapes are often observed during the formation of 2 3 reproductive isolation (Feder et al. 2012). Under the speciation with gene flow model, this type of genomic landscape can occur during early (Belleghem et al. 2020; Choi et al. 2020), 4 intermediate (Liu et al. 2020) or even late (Ravinet et al. 2018) stage of divergence. The 5 6 divergent genomic regions between populations, so-called "genomic islands", have often 7 been involved in the formation of phenotypic differences (Wang et al. 2020) and species 8 barriers (Poelstra et al. 2014; Wang and Coop 2022). However, these regions may not be necessarily associated with reproductive isolation (Clarkson et al. 2014) and can sometimes 9 be caused by hitchhiking effect around regions that have been divergently selected (Feder 10 et al. 2012). This pattern is expected to be more pronounced in genomic regions with an 11 elevated local recombination rate (Cruickshank and Hahn 2014; Burri et al. 2015). In some 12 cases, the force of direct selection and recombination would be entangled due to the 13 reinforced effects of selection within low recombination regions (Martin et al. 2019; Nguyen 14 15 et al. 2022; Schluter and Rieseberg 2022). The fact that speciation is affected by multiple, complex, sometimes interacting factors signifies the importance of choosing a system that 16 allows the detection and quantification of these intrinsic forces. 17

The Kentish Plover (Charadrius alexandrinus, hereafter "KP") and White-faced Plover (C. 18 dealbatus, hereafter "WFP") are two common shorebird species in East Asia. They form a 19 20 system including three taxa, the nominate subspecies of KP (C. a. alexandrinus) found in 21 inland saline lakes and north-eastern coastal areas of China, the WFP (C. dealbatus) spanning coastal areas from Southeast China to northern Viet Nam and a resident breeding 22 population on the Taiwan Island which is sometimes assigned to the C. a. nihonensis of KP 23 (Küpper et al. 2012). Previous studies have shown that KP and WFP are incipiently diverging 24 25 species under secondary contact (Sadanandan et al. 2019; Wang, Que, et al. 2019; Wang, Maher, et al. 2019). The two species show phenotypic differences in multiple aspects. Based 26 on previous studies, male KPs in mainland China show differences not only in plumage 27 28 colour such as darker lores and neck rings compared to WFPs (Figure 1A), but also in other 29 morphometric traits (Wang, Que, et al. 2019), as well as behaviours and habitat preferences (Rheindt et al. 2011). The breeding ranges of the two species are parapatric and overlap in a 30

contact zone along the eastern coastline of China. This contact zone is believed to stretch 1 2 from the southern part of Zhejiang to the northern part of Fujian, along the northwest coast 3 of the Taiwan Strait (Sadanandan et al. 2019; Wang, Que, et al. 2019). Microsatellites 4 analyses support that the Taiwan population is closely related to *alexandrinus* (Wang, Que, et al. 2019), which would be consistent with a scenario where there has been strong 5 population admixture of *alexandrinus* and *dealbatus* near the Taiwan Strait. However, the 6 hybridization pattern in the hypothetical contact zone and the extent to which genomic 7 factors mentioned above influence the speciation of KP and WFP remain ambiguous. 8

9

Here, we investigated the hybridization history and genomic divergence patterns of KP and 10 11 WFP across mainland China and Taiwan Island by addressing the following questions: 1) What is the extent and timing of introgression among the above populations? 2) What 12 genomic regions of divergence are maintained in the face of gene flow? and 3) How are the 13 characteristics of these genomic regions associated with the formation of species barriers 14 between the two species? Based on a dataset of 96 resequenced individual genomes 15 representing majority of different geographical settings, i.e. inland, coastal and islands, our 16 population genomic results suggest two distinct genetic groups, KPs and WFPs, with five 17 18 subgroups with different levels genetic admixture (inland KPs, coastal KPs and Taiwan KPs, WFPs in the contact zone and WFPs outside the contact zone). Using this resulting 19 geographical setting, we carried out demographic history reconstruction. We confirm that 20 Taiwan Island is an ancient hybrid zone of mainland coastal KPs and WFPs, and gene flow in 21 this area has been cut off. Mainland coastal KPs and WFPs share a low autosomal 22 divergence level, but a few specific genomic regions of higher divergence are found on 23 Chromosome 1. Furthermore, our findings suggest that some genomic regions affected by 24 25 selective sweep with decreased local recombination rate can be maintained during speciation with gene flow between incipient species. Overall, our study provides insight into 26 a system where both ancient and ongoing gene flow affect speciation and suggests that 27 28 even in early stage of speciation, regions containing genes with strong selective advantage 29 in low recombination regions can be maintained and may be involved as early species barriers. 30

#### 1 Results

### 2 Population structure and divergence history of KP and WFP

Principal component analysis (PCA) was performed to 96 individuals with no positive kinship
value with each other indicated by KING software (Table S1). Clusters form among five main
groups of KP and WFP (Figure 1C). To be noted, WFPs from the contact zone appear to be
closer to KPs on the PC1 axis compared to those WFPs outside the contact zone. The five
main groups of KP and WFP also cluster in the NJ tree in a similar pattern, supporting this
population structure (Fig. S1).

Admixture can be observed among populations of KP and WFP when K = 2 (Figure 1B) as
suggested by the CV value (Fig. S2A). When the geographical location shifts from north to
south, the admixture level in coastal and island KP populations shows a slight increase. In
WFPs, most admixed individuals are from the contact zone areas. Inland KPs and WFPs from
outside the contact zone show relatively lower admixture level compared to other
populations. When K = 3, all KP individuals can be distinguished by shared genetic materials
(Fig. S2B).

Demographic history inference showed at least one bottleneck in all five populations in a 16 time span of 200,000 years before present (Figure 1D). Overall, the two species have rather 17 different demographic history patterns, with WFP reaching a higher effective population 18 size than KP among mainland populations about 10,000 years ago, but their population sizes 19 dropped more drastically than mainland KPs in recent time, finally down to the lowest point 20 21 around 1,000 years ago. The effective population size of inland KPs decreased to around 22 100,000 about 20,000 years ago, as the one of coastal KPs decreased to 260,000 about 23 40,000 years ago and stayed stable afterwards. Both populations near the contact zone 24 experienced two bottlenecks respectively in the last 20,000 years.

<sup>25</sup>Nucleotide diversity ( $\pi$ ) of the five main populations shows no obvious divergence (Fig. S3), <sup>26</sup>with Taiwan Island population having the lowest diversity level, which is in agreement with <sup>27</sup>their small effective population size indicated by demographic history inference. The five <sup>28</sup>populations share similar patterns in the distribution of Tajima's *D* (Fig. S4). Combining with <sup>29</sup>results of their demographic history (Figure 1D) this distribution pattern of Tajima's *D* may <sup>30</sup>suggest recent decrease in population size in all five populations. To define the best sliding window size for following population statistics calculation, linkage disequilibrium ( $r^2$ , LD) decay analyses were performed for each of these populations. LD decays rapidly in these five populations, with  $r^2$  reaches the lowest value within about 50kb in distance (Fig. S5). Among these populations, two contact zone populations have greater background  $r^2$  levels than other three populations, and  $r^2$  levels of three KP populations drop faster than those of WFP populations.

7

# 8 Time and extent of introgression within two contact zone populations

9 In order to infer whether there is any migration event among the existing populations, 10 TreeMix analyses were run with migration edges set from zero to seven to investigate possible gene flow pattern. For each migration edge value, the same analysis was repeated 11 ten times. OptM result suggests m = 2 as the best migration edges (Fig. S6A). Under this 12 scenario (Figure 2A), the one with higher weight points from the branch node of Taiwan KPs 13 to one of the coastal KP populations (Cangzhou, sampling site marked as "4" in Figure 1A), 14 and the other is from Taiwan KPs to one of the contact zone WFP populations (Fuzhou, 15 16 sampling site marked as "9" in Figure 1A).

17

In addition, three ABBA-BABA tests were performed to evaluate the proportion and timing 18 of introgression between KP and WFP. Under the assumption of a given tree structure, the 19 proportion of introgression is evaluated with proportion of sliding windows with both D and 20  $f_{dM}$  higher than zero, and the timing of introgression can be reflected by the  $D_{FS}$  pattern, as 21 recent introgression involves positive *D* values in low frequency alleles (Martin and Amos 22 23 2020). Under the tree structure ((Taiwan KPs, coastal KPs), WFPs) with three KP individuals 24 from Azores Islands as outgroup, 25.60% windows have both D and  $f_{dM}$  higher than zero (Table S2), showing a limited proportion of introgression. The same tree structure shows a 25 26 negative total D value (red dashed line) in the  $D_{FS}$  shape, with positive and high-weighted D 27 on low-frequency alleles and negative D from intermediate to high-frequency alleles (Figure 3A), which indicates very recent but limited gene flow between P2 (coastal KPs) and P3 28 (WFPs), as well as possible ancient gene flow between P1 (Taiwan KPs) and P3 (WFPs). 29 30 When replacing the Taiwan population with inland KPs in the same tree test, introgressed

1 window proportion reaches 77.04% between coastal KPs and WFPs (Table S2). D<sub>FS</sub> pattern 2 under this scenario shows a more distinctive shape of introgression between P2 (coastal KPs) and P3 (WFPs), where both the total D value and the D values on low to intermediate-3 4 frequency alleles are positive (Figure 3B), which supports recent gene flow between the two 5 coastal populations. To be noted, both situations above show signs of relatively limited gene flow, as the absolute values of total Ds involved in the D<sub>FS</sub> patterns are lower than 0.1. 6 7 Under the tree structure ((inland KPs, Taiwan KPs), WFPs) with Azores KPs as outgroup, 8 85.48% of total windows show both positive values in D and  $f_{dM}$  (Table S2). A higher positive 9 total D value can be found in the D<sub>FS</sub> pattern, where negative values appear on lowfrequency alleles and the same values become positive and rise from intermediate to high-10 frequency alleles (Figure 3C). Consistent to what has been shown in previous sections, this 11 result shows no recent gene flow from P3 (mainland WFPs) to P2 (Taiwan KPs). However, 12 the absolute total *D* value under this tree structure is the highest among all three hypothesis 13 14 tests, and the whole *D<sub>FS</sub>* shape is skewed towards high-frequency alleles. Combined with the negative total D value in Figure 3A, this may be taken as another indication that there once 15 was extensive gene flow from mainland areas to the Taiwan KPs, but this event was cut off, 16 leaving only a limited number of introgressed alleles gradually being fixed in the island 17 population. 18

19

### 20 Genomic islands between coastal KP and WFP

Although recent gene flow is suggested by previous results and coastal KPs and WFPs share 21 22 a relatively low divergence level (autosomal average  $F_{ST}$  = 0.028, autosomal average  $d_{XY}$  = 23 0.31), a heterogeneous pattern with two major genomic islands on Chromosome 1 can be 24 observed in F<sub>ST</sub> landscape between coastal KP and WFP populations (Figure 4). When the 25 threshold of high divergence regions was set to 0.25 according to Hartl and Clark (Hartl et al. 26 1997), two mega base level islands of divergence can be found on Chromosome 1 (Figure 4). 27  $d_{XY}$  shows similar rising pattern in these regions, suggesting high level of both relative and 28 absolute genetic divergence (Figure 4A and 4B). Local recombination rate of both 29 populations decreases severely within these regions (Figure 4C and 4D), nearly dropping down to zero in some sliding windows. 30

2 Cross-population selective sweep scans were run within each of these genomic regions to 3 compare the strength of selective sweep. XP-nSL, a modified version of nSL for crosspopulation comparisons was used because of its sensitivity to recent soft sweep (Ferrer-4 5 Admetlla et al. 2014). Normalised XP-nSL patterns within these regions suggest that the two 6 species have gone through positive selection in different strength. Selective sweep in 7 coastal KPs is indicated by a slight peak in the XP-nSL max value in the 54-55Mb region (Figure 4E, p = 0.0046 when comparing with the previous nearby region of the same length), 8 9 while those in WFP are indicated by a valley in the XP-nSL min value in the 93-94Mb region (Figure 4F). It is then suggested that the 54-55Mb region on Chromosome 1 is more 10 influenced by selective sweeps in coastal KPs, while the 93-94Mb region is more influenced 11 by selective sweeps in WFPs, and the 83-84Mb region shows no signal of stronger selective 12 sweep in both populations. 13

Unary linear regression results show significant positive correlation between  $F_{ST}$  and  $d_{XY}$  in 14 windows with top 5%  $F_{ST}$  and  $d_{XY}$  (Figure 5A). In the same windows, a significant negative 15 correlation is seen between  $F_{ST}$  and  $\pi$  of coastal KPs, but not  $F_{ST}$  and  $\pi$  of WFPs out of 16 contact zone (Figure 5B and 5C), suggesting the reduction of diversity in coastal KPs, not 17 18 those in the WFPs out of contact zone, as the main cause for their divergence. For XP-nSL max and XP-nSL min values, we expect an opposite correlation with F<sub>ST</sub>. As peak in XP-nSL 19 max means stronger selection in coastal KPs, a significant positive correlation between this 20 value and F<sub>ST</sub> indicates more contribution of selection in this population to the divergence 21 22 level. On the contrary, because valley in XP-nSL min means stronger selection in WFPs out of contact zone, the correlation between this value and  $F_{ST}$  should be significantly negative if 23 24 the selection in this population contributes more to the divergence level. As a result, both 25 XP-nSL max and XP-nSL min show significant positive correlations with  $F_{ST}$  in these regions 26 (Figure 5D and 5E). The above results indicate that selective sweeps in the coastal KPs had 27 more contribution to the divergence level in genomic islands between the two populations. Forty genes were found in regions of F<sub>ST</sub> higher than 0.25 (Table S3) using the annotation of 28 29 Gallus gallus. Among these genes, 32 are located on Chromosome 1, two of which (ZBTB20 30 and *IGF1*) are enriched in the term "positive regulation of glycolytic process" (GO: 0045821) 31 with p = 0.019 in Gene Ontology (GO) enrichment analysis. In KEGG pathway enrichment

1 analysis, two of the forty genes, *PAH* and *THL*, were enriched onto the term "Folate

2 biosynthesis" (gga00790), with p = 0.063.

3 One of these genes that locates in the 54-55Mb region that is selected in coastal KP population (Figure 4F), *IGF1*, is the coding gene for insulin-like growth factor I (IGF-1), which 4 5 is an evolutionarily conserved peptide related to growth regulation (Baker et al. 1993), body 6 size (Poole et al. 2011), neural development (D'Ercole et al. 2002) and life span (Shimokawa 7 et al. 2002; Duran-Ortiz et al. 2021) in a large range of organisms. Although it has been 8 suggested that IGF-1 levels did not show relatedness to migration strategies in some 9 passerine birds (Lodjak et al. 2018), this peptide has been proved to play a role in the expression of plumage traits and moulting in the passerine species Panurus biarmicus (Mahr 10 et al. 2020; Lendvai et al. 2021). This gene may play a role in difference in sexually selected 11 traits in these two species, as it has been observed that large males in warmer environment 12 may display more ornamentation during breeding seasons (Song et al. 2023). 13 Besides, five genes were found in the 83-84Mb region where selective strength shows no 14 different patterns between coastal KPs and WFPs outside the contact zone (Table S3). 15

16 Among these five genes, *ZBTB20* and *DRD3* are both relevant to the development and

17 function of nervous system and have been proved to be relevant to the immunity to avian

18 influenza in chickens (Drobik-Czwarno et al. 2018).

19

# 20 Discussion

Based on large range sampling and high coverage sequencing, we used a variety of genomic 21 scanning techniques in this study to confirm the existence of certain genomic island regions 22 23 in an incipient species pair of plovers. These findings support the argument that selection in 24 genomic regions with reduced recombination rate can act as a driving force for early 25 speciation process, and that these regions can be maintained in the face of ongoing gene 26 flow. We confirm that the Taiwan Island serves as an ancient hybrid zone for coastal KPs 27 and WFPs, where introgression between these two species leave obvious and detectable 28 traces on their genomes. Populations on either side of the Taiwan Strait display significant 29 differences in the timing and extent of gene flow, with historical introgression on Taiwan 30 Island and contemporary gene flow occurred along the coast of mainland China.

1

#### 2 Introgression near the Taiwan Strait

3 Previous studies have shown that introgression can be found on specific loci between KP 4 and WFP populations (Wang, Que, et al. 2019). Here, we show that there are interspecific 5 gene flows from KP to WFP, among which the highest weighted migration appears between 6 Taiwan KPs and WFPs (Figure 2A). The D<sub>FS</sub> analysis further suggested that there is little 7 evidence of ongoing gene flow between Taiwan KPs and mainland populations. According to the original paper of the D<sub>FS</sub> method, recent gene flow that occurs within 0.1 generation ago 8 would show a higher positive D value on low-frequency alleles in the allele frequency 9 spectrum (Martin and Amos 2020). Our result of contact zone WFPs shows recent but 10 relatively low level of introgression between coastal KPs and WFPs in contrast with that 11 between Taiwan KPs and WFPs. This further confirms the hypothesis that coastal KPs and 12 13 WFPs have created admixtures under secondary contact after their speciation (Wang, Que, et al. 2019), and that this gene flow is still ongoing. Further studies that focus on 14 populations near Fujian Province may provide more evidence on the specific areas of their 15 hybridization and narrow down the geographical boundary of the contact zone. 16 17 Hybrid zone study has been an effective way of investigating incipient speciation process (Brelsford and Irwin 2009; Pinto et al. 2019). In this study, we show that even with a short 18 geographical distance, the genomic pattern of two contact zone populations can be 19 different due to multiple genetic and environmental factors. One explanation is that the 20 21 Taiwan Island was once connected with coastal areas in Chinese mainland such as Fujian 22 with recurring land bridges (Voris 2000). During this period, coastal KPs that adopt long 23 distance migration and WFPs that inhabit the contact zone areas may have created 24 admixture on Taiwan Island. As the historical sea level rose, this connection between the

25 visland and mainland was cut off, so was the gene flow between them.

For Taiwan KP population, indirect evidence for the cut-off of their gene flow with mainland populations can also be seen from the introgression pattern. In the *D<sub>FS</sub>* result, the cut-off of recent gene flow is shown by the lack of positive *D* values on low-frequency alleles (Figure 3C). Unlike what happened to the WFPs in the mainland contact zone, the Taiwan KPs have been more deeply impacted by historical geographic events. The disappearance of land

bridges in the Taiwan Strait has cut off, or at the very least, significantly reduced its gene 1 2 flow with mainland populations, which in some ways has contributed to its current isolation. 3 Combined with their relatively lower population diversity (Fig. S3) and smaller population 4 size (Figure 1D), we suggest that further studies into the examination of genetic loads and 5 inbreeding are significant to evaluate conservation status of this unique island population. 6 Despite different timing of introgression in Taiwan and mainland coastal KPs, the admixture 7 level in these populations is nearly identical (Figure 1B). This may indicate genome 8 stabilization in the evolutionary history of these populations. For Taiwan KPs, the 9 introgressed genes and loci may have been kept at a stable level after the cut-off of gene flow. But for the mainland coastal KPs among which the gene flow with WFPs is still 10 ongoing, the specific mechanism that maintains the similar admixture level in the face of 11 recent ongoing gene flow is still unknown, and it can be taken as an indication of some post-12 zygotic reproductive isolation between the two species. Future studies on hybrid zone 13 14 individuals may shed light on the complex speciation history of this system. There has also been effort to sort out relationship between the Taiwan KPs and KP 15 populations in Japan (Sadanandan et al. 2019) because they were ascribed to the same 16 subspecies *nihonensis* of KP. Future studies are encouraged to test the ancestry and 17 18 colonization history of the Taiwan and Japan populations of KP, especially their tendency to create genetic admixture with WFP populations, which would provide a better 19 understanding of the colonization routes and speciation process of the two closely related 20 species along East-Asia coastlines and islands along the Pacific-Asian continental shelf. 21

22

### 23 Divergence of KP and WFP during incipient speciation

The KP and WFP have been shown to have low average divergence level based on
microsatellite study (Wang, Que, et al. 2019), but rather obvious differences on phenotypic
level (Rheindt et al. 2011). The low background *F*<sub>ST</sub> level shown in this study supports that
the divergence level between KP and WFP is small on a genomic aspect (Figure 4).
Theoretically, this divergence can be homogenized by the ongoing gene flow after their
secondary contact through hybrid zone near Fujian Province. However, with differences in
multiple behaviours (particularly migration strategies), these two birds may have formed

divergent mating preferences which facilitate their speciation despite the homogenizing
 effect of gene flow. Future studies on ecological traits such as assortative mating may
 provide a deeper insight into the driving force of their speciation.

4 Our results of genomic islands are consistent with the finding that regions of higher 5 divergence often locate in low-recombination areas (Cuevas et al. 2022). In our result, we 6 showed that most of the high divergence genomic regions span across a length of Mb 7 (Figure 4). To examine whether structural variation such as inversions caused this pattern, a 8 local PCA was performed following the previously described method (Huang et al. 2020). However, we found no obvious evidence of inversions in two genomic islands (Fig. S7). On 9 one hand, this could be explained by positive selection around these genomic regions 10 strengthened by low recombination rate, as higher linkage disequilibrium within these 11 regions is expected to give rise to stronger hitchhiking effect. This is further supported by 12 the LD map on Chromosome 1 (Fig. S8), where the two island regions show high level of LD 13 both with themselves and with each other. We suggest that the two regions may be 14 15 distributed close to each other on the genome but being mapped to separate positions during the conversion from contigs to Gallus gallus chromosomes. Another possibility is that 16 introgressed alleles in low recombination regions can be purged more easily by selection 17 against them due to the existence of incompatibilities (Moran et al. 2021), causing a similar 18 pattern of increased divergence between populations without positive selection. To confirm 19 20 this, further investigations into the fitness and genome architecture of hybrid individuals are needed to gain a clearer view of the intrinsic mechanism underlying hybridization. 21

Another factor that may influence the inference of selection is the heterogenous 22 23 distribution of N<sub>e</sub> across the genome. N<sub>e</sub> heterogeneity has been shown in multiple systems (Jiménez-Mena et al. 2016), and it is suggested to have impact on statistics that are largely 24 25 affected by population diversity such as F<sub>ST</sub> (Wolf and Ellegren 2017). In our result, significant negative correlation between  $\pi$  and  $F_{ST}$  (Figure 5B), reduced recombination rate 26 27 (Figure 4C and 4D), and a high level of LD (Fig. S8) can be observed within high divergence 28 regions on Chromosome 1, indicating potential effects of heterogeneous  $N_{\rm e}$  on the selection inferences. In order to confirm the extent of this influence, we also evaluated the 29 correlation between  $F_{ST}$  and  $d_{XY}$ , an absolute divergence indicator that is not affected by the 30

1 distribution of intra-population diversities, within the high divergence regions (Figure 5A).

2 The positive correlation between these two measurements suggests that the high *F*<sub>ST</sub> values

3 within high divergence island regions are not mainly elevated by heterogeneous  $N_{\rm e}$ 

4 distribution. In addition, nSL was chosen for inference of selective sweeps because of its

5 relatively higher robustness when dealing with local recombination rate variation (Ferrer-

6 Admetlla et al. 2014) as a supplement of the above divergence measurements.

7 Some populations show extremely large drift parameters in the TreeMix plot (for example,

8 the SW, BH, RD, LYG and TS populations) (Figure 2A). These drift parameters are

9 represented by longer branches in the tree, indicating the amount of genetic drift along a

10 population (Wang et al. 2016). In our result, populations with longer branches are all those

11 with low (5×) sequencing depth. This issue was fixed when rerunning TreeMix with

12 individuals of a standardized 15× sequencing depth (Fig. S6B). Thus, we suggest that the

13 possible reason for these extreme branch lengths may be the mixing of samples sequenced

14 by multiple depths, and its subsequent effect on the representativeness for a real

15 population, as lower sequencing depth would only cover a limited number of usable sites.

16

# 17 Species barrier between KP and WFP

Despite the recent ongoing gene flow between them, the coastal KPs and WFPs show 18 heterogeneous yet obvious divergence pattern on their genomes (Figure 4). Combined with 19 the results from linear regression analyses (Figure 5), it can be concluded that some 20 21 genomic islands on Chromosome 1 are maintained with the existence of ongoing gene flow. 22 The contrasting patterns between  $\pi$  and  $F_{ST}$  in coastal KPs and WFPs outside the contact 23 zone suggest that these genomic islands may be driven by selection in different directions in 24 these two populations (Figure 5B and 5C). If the selection in one population is relevant to 25 the formation of genomic islands, a significant negative correlation between its  $\pi$  and  $F_{ST}$ would be expected, as divergent selection in one population would reduce its diversity, 26 subsequently increase  $F_{ST}$  value which is calculated based on population diversities. 27 However, in these island regions, only  $\pi$  in coastal KPs shows this pattern (Figure 5B), which 28 29 could be indicating that it is the stronger selection in coastal KPs, but not in WFPs outside

1 the contact zone, that contributes to these islands of divergence. This conclusion can be 2 further confirmed by the correlations between XP-nSL max and XP-nSL min and  $F_{ST}$  in the 3 two populations (Figure 5D and 5E), where positive correlation between XP-nSL max and  $F_{ST}$ 4 shows that relatively stronger positive selection in coastal KPs is the main cause of the island 5 regions, compared to those in WFPs outside the contact zone which is assumed to be 6 negative if it contributes to the island regions.

7 A few genes found within these regions have been suggested to be relevant to plumage 8 phenotypes and natural immunity to avian influenza (Table S3). For example, the *IGF1* gene 9 relevant to plumage phenotypes locates in the 54-55Mb region that is positively selected in coastal KPs compared to WFPs outside the contact zone (Figure 4F). However, the fact that 10 genes mentioned above locate within genomic regions with different selective strength 11 12 suggests that the existence of genomic islands may not only be due to phenotypic differences. For example, even though the ZBTB20 and DRD3 genes locate within the 83-13 14 84Mb genomic region which shows no difference in cross-population selection comparison (Figure 4E and 4F), they have important immune function towards avian influenza which is 15 crucial regardless of the individuals' population and their phenotypic difference. 16

In this study, we show that most identified genes that locate within the high divergence 17 18 regions show no obvious enrichment to specific functional term, even though one of them 19 (IGF1) does have potential relevance to plumage phenotypic differences. However, the 20 relationship between loci that are targeted by selection and the formation of reproductive 21 isolation still remains ambiguous (Jiggins and Martin 2017; Schluter and Rieseberg 2022). As suggested that selection favours restricted recombination (Felsenstein 1981; Butlin 2005), 22 the existence of recombination variation increase the difficulty to conclude on whether 23 24 there is clear causal relationship between the genes found within high divergence regions 25 and the species barriers between KP and WFP. Future studies on the populations in the contact zone, potentially between Wenzhou and Ningde (26.66° N, 119.55° E - 27.99° N, 26 27 120.70° E) in Fujian Province according to sampling sites in our study, that particularly focus 28 on determining whether assortative mating occurs, may yield crucial evidence on the 29 preservation of reproductive barriers between the two species. Additionally, the proportion 30 of F1 offspring that survive in contact zone populations may shed light on the two species' 31 post-zygotic separation.

## 2 Conclusions

3 Our results confirm the existence of introgression between KP and WFP in Chinese coastal 4 areas, and that the Taiwan Island is an ancient hybrid zone of the two species. The gene 5 flow between Taiwan KPs and mainland WFPs were cut off, but the gene flow between 6 mainland coastal populations is still ongoing and weaker in strength. KP and WFP share only 7 a small average difference on their autosomal genomic landscape, with large proportion of 8 genomic regions that are affected by introgression. Our results suggest that the KP and WFP are under recurrent gene flow from a contact zone along the coastline near Fujian Province. 9 10 Even with a short divergence time and secondary contact thereafter, the two species have 11 maintained a few genomic islands on Chromosome 1 in the face of gene flow, and genes found in these regions are related to plumage and immune phenotypes. We suggest that 12 13 genomic islands can be maintained between incipient species with the joint force of 14 divergent selection and reduced local recombination rate.

15

#### 16 Materials and Methods

17 Sampling and sequencing

A total of 80 individuals were caught by traps during breeding seasons from 2014 to 2019 (Figure 1A., Supplementary Table 1), combining with 22 individuals from previous study (Wang, Que, et al. 2019). Blood samples were stored in ethanol immediately after collection and were preserved at -40°C until sequencing. All sampling procedures were carried out under the permission and guidance from the School of Life Sciences, Sun Yat-sen University.

Sample from one female White-faced Plover was selected for PacBio sequencing to
assemble a reference genome. High-quality genomic DNA was extracted using the DNeasy
Blood and Tissue kit (Qiagen, GmbH, Hilden, Germany) according to the manufacturer's
protocol. The SMRTbell 20Kb library was constructed using the SMRTbell Template Prep Kit
1.0 (Pacific Biosciences). The genomic library was sequenced on the Pacific Biosciences

(PacBio) Sequel single molecule real-time (SMRT) platform (PacBio RSII) in the Genome
Center of Novo Genomics (Novogene, Beijing, China). The estimated coverage was 120×.
Other 79 individual samples were sequenced using Illumina HiSeq X. Due to the sample
quality and duration of sampling time at different locations, sequencing depth varies from
5× to 30× (Table S1). The final dataset contains 101 whole-genome resequenced individuals
with different sequencing depth to cover as many sampling locations as possible.

7

### 8 Reference genome assembly

9 We assembled PacBio long-reads using Wtdbg2 (Ruan 2020) to generate the draft genome 10 with augments: -K 2000 --edge-min 4 -p 17 -S 4 -L 5000 --tidy-reads 8000. Next the draft 11 genome was polished using Minimap2 (Li 2018), Samtools (Li et al. 2009) and BWA (Li and 12 Durbin 2009) as Wtdbg2 recommended with the default parameters by using NGS data. The 13 reference genome assembled by using PacBio long reads was approximately 1.23Gb in 14 length and contains 814 contigs (contig N50 of 24.48 Mb). The completeness of the genome 15 evaluated by BUSCO(Waterhouse et al. 2018; Manni et al. 2021) is 94.9%.

16

# 17 Read mapping and variant calling

We aligned resequencing data of individuals of Kentish plover and white-faced plovers to 18 the WFP reference genome using BWA (Li and Durbin 2009) and found all coverage and 19 mapping rate was above 95%. We further used Trimmomatic (Bolger et al. 2014) for 20 trimming out adaptors. BWA (Li and Durbin 2009) was used to align raw sequences to the 21 22 WFP reference genome. Samtools (Li et al. 2009) was used to convert sam files into bam 23 format and remove duplicates. We used the HaplotypeCaller and GenotypeGVCFs tools 24 implemented in GATK (McKenna et al. 2010) to generate call haplotypes and do genotyping 25 for all individuals.

## 1 Data filtering

- 2 After obtaining a raw SNP dataset, two filter steps were performed aiming to control data
- 3 quality: a) We filtered out sites using GATK (4.1.4.1) with augments: QD < 2.0 || FS > 60.0 ||
- 4 MQ <40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 3.0 and kept sites
- 5 with "PASS" flag in the dataset. b) We used VCFtools 0.1.17 (Danecek et al. 2011) to filter
- 6 out sites with augments: --minDP 3 --maf 0.05 --min-meanDP 3 --max-missing 0.5.
- 7 In order to exclude system bias caused by a dataset of mixed sexes, the WFP reference
- 8 genome was aligned to the RefSeq of *Gallus gallus* (GRCg6a) using Chromosemble
- 9 implemented in Satsuma2 (Grabherr et al. 2010). The coordinates of filtered SNPs in VCF file
- 10 were then transformed from contig-level to chromosome-level based on Satsuma2 output
- 11 using custom Python scripts. In most of the following analyses, SNPs on both Z and W
- 12 chromosomes were excluded in this way and were included only when individuals used for
- 13 analysis were the same sex. Considering the potential discordance of sex chromosomes
- 14 between these two species, genomic landscapes and other statistics are not calculated for Z
- 15 and W chromosomes either.
- 16 Kinship analysis was performed using software KING 2.2.6 (Manichaikul et al. 2010) with 17 default augments. Five individuals (see Table S1) that show a positive kinship value with 18 others within the same population were then excluded from following analyses.
- 19

# 20 Population structure

Principal component analysis was performed with PLINK v1.90b6.9 (Purcell et al. 2007), and
 result was visualized using R package ggplot2 (Wickham 2011). The first two principal
 components were used as PC1 and PC2, respectively.

ADMIXTURE 1.3.0 (Alexander et al. 2009) were run multiple times with K from 1 to 6. The best result was selected based on the lowest cross-validation (CV) error generated by ADMIXTURE. Results were then visualized with R package POPHELPER (Francis 2017).

## 2 Demographic history inference

- 3 Thirty-eight female individuals with sequencing depth higher than 15× (Table S1) were
- 4 selected for demographic history inference using Stairway Plot 2 (Liu and Fu 2020). BAM
- 5 files representing each population obtained after alignment were used for folded SFS

6 calculation using ANGSD v0.937 (Korneliussen et al. 2014). The output SFS files were used as

7 input files for Stairway Plot 2.

8

# 9 Phylogenetic relationship analysis

10 A distance matrix of 96 unrelated individuals was constructed with VCF2dis

11 (https://github.com/BGI-shenzhen/VCF2Dis). An NJ tree was then built with the generated

12 matrix using PHYLIP (Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6.

13 Distributed by the author. Department of Genome Sciences, University of Washington,

14 Seattle).

TreeMix (Pickrell and Pritchard 2012) was used to infer migrations between populations
from different sampling locations. A set of simulations were run to define the best migration
edges in TreeMix analysis. Migration edges were set as 0~7, each with ten parallel runs.
Outputs were analyzed using online version of OptM (Fitak 2021). The number of migration
edges was set as 2 based on previous result to best describe migration pattern between
populations.

21

# 22 **Population statistics calculation and genomic regions scanning**

- 23 We used python scripts from <u>https://github.com/simonhmartin/genomics\_general</u> as used
- in previous study (Martin et al. 2020) to calculate  $F_{ST}$ ,  $d_{XY}$  and  $\pi$  values in 50kb non-
- 25 overlapping sliding windows. High divergence genomic regions with top 5% level of both  $F_{ST}$
- and  $d_{XY}$  values on Chromosome 1 are selected for unary linear regression analyses.

1 For detecting signals of natural selection on the genomes, selective sweep scanning was

2 performed on each population using the software selscan (Szpiech and Hernandez 2014).

3 Cross population analyses were performed, then normalized in 50kb non-overlapping sliding

4 windows using norm, a program for downstream analysis of selscan output. We chose XP-

5 nSL (Szpiech et al. 2021), an extension of nSL (Ferrer-Admetlla et al. 2014) for quantifying

6 selective strength for its robustness for soft sweeps (Ferrer-Admetlla et al. 2014).

7 Visualization was obtained using R package ggplot2 and CMplot (Yin et al. 2021).

8 ReLERNN (Adrion et al. 2020) was used to estimate the recombination landscape of coastal

9 KP and WFP populations. Sliding window size was suggested by the software for each

10 chromosome, and each recombination rate estimation was assigned to the middle position

11 of that sliding window when plotting.

12

## 13 Introgression analyses

14 Under the assumption that coastal KPs and WFPs have the potential to create admixture

15 and hybridize both near Fujian coastal areas and in Taiwan Island, we performed ABBA-

16 BABA tests for tree structures as followed: (((Taiwan KPs, coastal KPs), WFPs), Outgroup),

17 (((inland KPs, coastal KPs), WFPs), Outgroup) and (((inland KPs, Taiwan KPs), WFPs),

18 Outgroup). According to their phylogenetic relationship and previous population history

19 inference, WFPs from both within and out of the contact zone were set as P3 together. Both

20 values in sliding window and total *D* values were calculated to better indicate the

21 introgression patterns in the two contact zone populations. A sliding window is considered

as an "introgressed" window if its D and  $f_{dM}$  values are both positive.

23 Three KP individuals collected in Azores Islands with no kinship to each other were

24 Sequenced as described above and used as outgroup, as previous studies using

25 microsatellites and mitochondrial DNA showed that Azores KP population exhibited high

26 level of divergence and low level of gene flow to its Eurasian continental counterparts

27 (Küpper et al. 2012; AlmalkI et al. 2016). D and  $f_{dM}$  values in 50kb non-overlapping windows

28 were calculated using python scripts from

29 <u>https://github.com/simonhmartin/genomics\_general</u>. *D<sub>FS</sub>* was obtained using python and R

scripts from <u>https://github.com/simonhmartin/dfs</u> as described in previous study (Martin
 and Amos 2020).

- 3
- 4 DATA AVAILABILITY: The filtered VCF containing 104 individuals (including three individuals
- 5 representing the Azore Islands outgroup) has been uploaded to Dryad (DOI:
- 6 10.5061/dryad.4f4qrfjjp). Custom scripts will be made available by the authors on request.
- 7
- ACKNOWLEDGEMENTS: We thank Xin Lin and Zitan Song for sample collection during
  breeding seasons from 2014 to 2021. This study is kindly funded by the DFGP Project of
  Fauna of Guangdong-202115 from Science and Technology Planning Projects of Guangdong
  Province (2021B121210002 to YL). The Interdisciplinary Innovation Team of the Chinese
  Academy of Sciences (CAS) "Light of West China" Program (xbzg-zdsys-202207).
- 13
- 14 **AUTHOR CONTRIBUTIONS:** Sample collection: N.Z., P.Q., C.Y.C., J.M., H.N. Genomic
- 15 sequencing: C.Z. Reference genome assembly: J.T. Data preprocessing: H.L. and W.Z.
- 16 Analyses and visualizations: W.Z., N.Z. and K.H. Manuscript writing: W.Z. Manuscript editing:
- 17 all authors.
- 18

# 19 Reference

- Adrion JR, Galloway JG, Kern AD. 2020. Predicting the landscape of recombination using deep learning. *Molecular Biology and Evolution* 37:1790–1808.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*:254–260.
- 24 AlmalkI M, Kupán K, Carmona-Isunza MC, López P, Veiga A, Kosztolányi A, Székely T, Küpper C. 2016.
- 25 Morphological and Genetic Differentiation Among Kentish Plover *Charadrius alexandrinus*
- 26 Populations in Macaronesia. *Ardeola* 64: 3–16.

- 1 Baker J, Liu J-P, Robertson EJ, Efstratiadis A. 1993. Role of insulin-like growth factors in embryonic
- 2 and postnatal growth. *Cell* 75:73–82.
- 3 Belleghem SMV, Cole JM, Montejo-Kovacevich G, Bacquet CN, McMillan WO, Papa R, Counterman
- 4 BA. 2020. Selection and gene flow define polygenic barriers between incipient butterfly species.
- 5 :2020.04.09.034470. Available from:
- 6 https://www.biorxiv.org/content/10.1101/2020.04.09.034470v1

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data.
 *Bioinformatics* 30:2114–2120.

- 9 Brelsford A, Irwin DE. 2009. Incipient speciation despite little assortative mating: the yellow-rumped
  10 warbler hybrid zone. *Evolution* 63:3050–3060.
- 11 Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ,
- 12 et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic
- 13 landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome*
- 14 *Research* 25:1656–1665.
- 15 Butlin RK. 2005. Recombination and speciation. *Molecular Ecology* 14:2621–2635.
- 16 Choi JY, Purugganan M, Stacy EA. 2020. Divergent selection and primary gene flow shape incipient
- 17 speciation of a riparian tree on Hawaii Island. *Molecular Biology and Evolution* 37:695–710.
- 18 Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, Webster M, Antão
- 19 T, MacInnis B, et al. 2014. Adaptive introgression between Anopheles sibling species eliminates a
- 20 major genomic island but not reproductive isolation. *Nature Communications* 5:4248.
- Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23:3133–3157.
- Cuevas A, Eroukhmanoff F, Ravinet M, Sætre GP, Runemark A. 2022. Predictors of genomic
   differentiation within a hybrid taxon. *PLoS Genetics* 18:1–30.
- 25 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth
- 26 GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- D'Ercole JA, Ye P, O'Kusky JR. 2002. Mutant mouse models of insulin-like growth factor actions in the central nervous system. *Neuropeptides* 36:209–220.
- 29 Drobik-Czwarno W, Wolc A, Fulton JE, Arango J, Jankowski T, O'Sullivan NP, Dekkers JCM. 2018.
- Identifying the genetic basis for resistance to avian influenza in commercial egg layer chickens.
   Animal 12:1363–1371.
- 32 Duran-Ortiz S, List EO, Basu R, Kopchick JJ. 2021. Extending lifespan by modulating the growth 33 hormone/insulin-like growth factor-1 axis: coming of age. *Pituitary* 24:438–456.
- Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* 28:
   342–350.
- 36 Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? 37 Evolution 35:124–138
- 37 *Evolution* 35:124–138.

- 1 Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard
- 2 selective sweeps using haplotype structure. *Molecular Biology and Evolution* 31:1275–1291.
- Fitak RR. 2021. OptM: Estimating the optimal number of migration edges on population trees using
   Treemix. *Biology Methods and Protocols* 6:1–6.
- Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population
  structure. *Molecular Ecology Resources* 17:27–32.
- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F, Lindblad-Toh K. 2010. Genome wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26:1145–1151.
- 9 Hartl DL, Clark AG. 1997. Principles of population genetics. Sinauer associates Sunderland
- Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversions
   contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology* 29:2535–2549.
- Jiggins CD, Martin SH. 2017. Glittering gold and the quest for Isla de Muerta. *J Evol Biol* 30:1509–
  1511.
- 14 Jiménez-Mena B, Hospital F, Bataillon T. 2016. Heterogeneity in effective population size and its
- 15 implications in conservation genetics and animal breeding. *Conservation Genetics Resources* 8:35–
- 16 41.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing
  Data. *BMC Bioinformatics* 15:356.
- 19 Küpper C, Edwards S V., Kosztolányi A, Alrashidi M, Burke T, Herrmann P, Argüelles-Tico A, Amat JA,
- 20 Amezian M, Rocha A, et al. 2012. High gene flow on a continental scale in the polyandrous Kentish
- 21 plover Charadrius alexandrinus. *Molecular Ecology* 21:5864–5879.
- 22 Lendvai ÁZ, Tóth Z, Mahr K, Osváth G, Vogel-Kindgen S, Gander BA. 2021. Effects of experimental

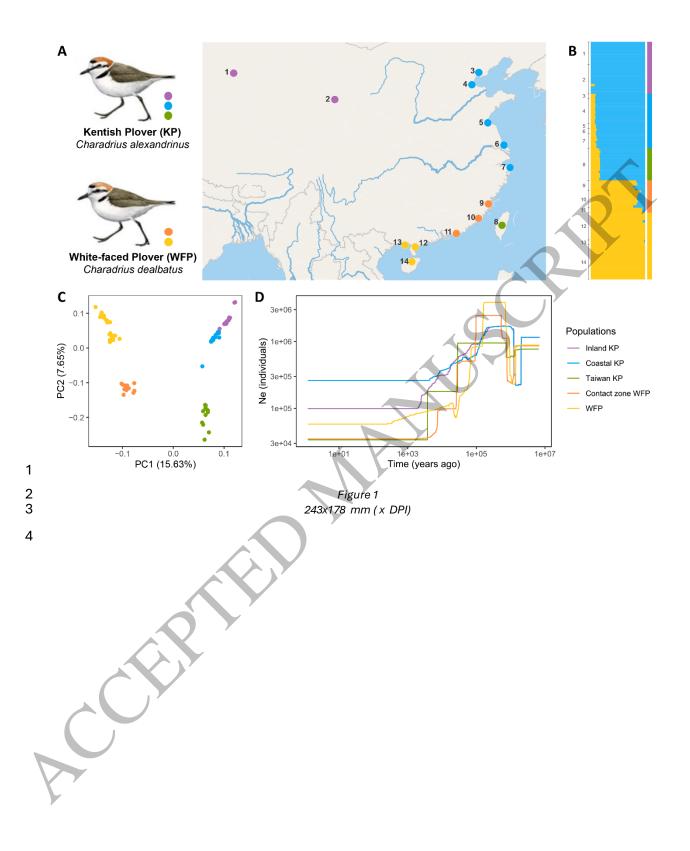
23 increase in insulin-like growth factor 1 on feather growth rate, moult intensity and feather quality in

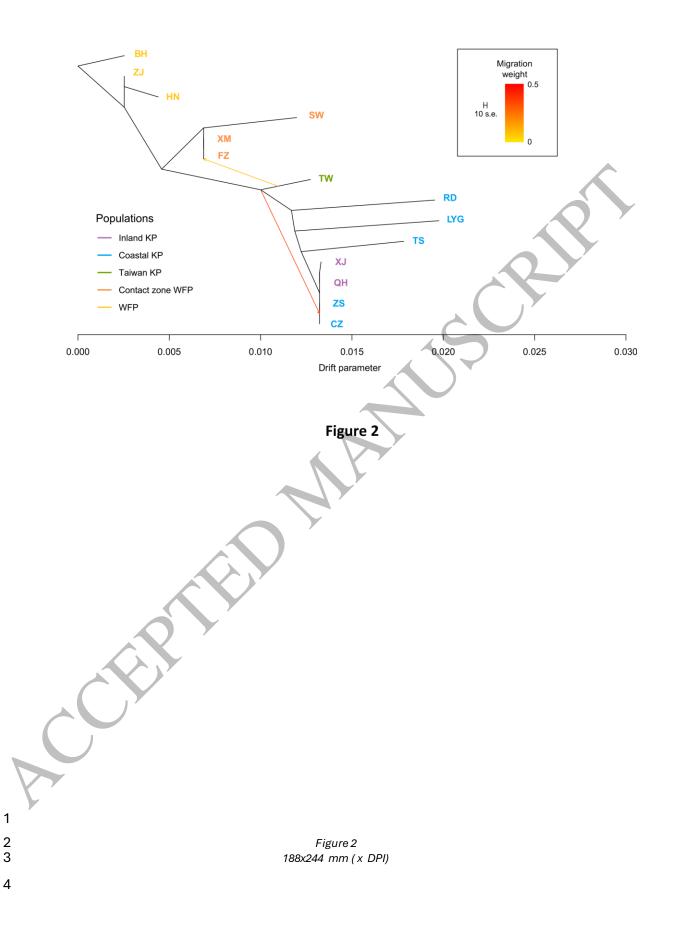
- 24 a passerine bird. *Journal of Experimental Biology* 224:jeb242481.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. *Bioinformatics* 34:3094–3100.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
   *Bioinformatics* 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Liu X, Fu Y-X. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency
   spectra. *Genome Biology* 21:280.
- Liu X, Glémin S, Karrenberg S. 2020. Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (Silene). *Molecular Ecology* 29:3511–3525.
- Lodjak J, Mänd R, Mägi M. 2018. Insulin-like growth factor 1 and life-history evolution of passerine
   birds. *Functional Ecology* 32:313–323.

- 1 Mahr K, Vincze O, Tóth Z, Hoi H, Lendvai ÁZ. 2020. Insulin-like growth factor 1 is related to the
- 2 expression of plumage traits in a passerine species. *Behavioral Ecology and Sociobiology* 74:39.
- 3 Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship 4 inference in genome-wide association studies. *Bioinformatics* 26:2867–2873.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and
   Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of
- 7 Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* 38:4647–4654.
- 8 Martin SH, Amos W. 2020. Signatures of introgression across the allele frequency spectrum.
- 9 *Molecular Biology and Evolution*:1–24.
- 10 Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to 11 introgression across butterfly genomes. *PLoS Biology* 17: e2006288.
- 12 Martin SH, Singh KS, Gordon IJ, Omufwoko KS, Collins S, Warren IA, Munby H, Brattström O, Traut W,
- Martins DJ, et al. 2020. Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *PLoS* Biology 19: c2000010
- 14 *Biology* 18: e3000610.
- 15 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,
- 16 Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: A MapReduce framework for analyzing
- 17 next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- 18 Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic
- 19 consequences of hybridization. *eLife* 10: e69016.
- 20 Nguyen ANT, Woods LC, Gorrell R, Ramanan S, Kwok T, McDonald MJ. 2022. Recombination resolves
- 21 the cost of horizontal gene transfer in experimental populations of *Helicobacter pylori*. *Proceedings*
- of the National Academy of Sciences the United States of America 119: e2119010119.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* 8: e1002967
- Pinto BJ, Titus-McQuillan J, Daza JD, Gamble T. 2019. Persistence of a geographically-stable hybrid
   zone in Puerto Rican dwarf geckos. *Journal of Heredity* 110: 523-534.
- 27 Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M,
- Grabherr MG, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene
   flow in crows. *Science* 344:1410-1414.
- 30 Poole EM, Tworoger SS, Hankinson SE, Schernhammer ES, Pollak MN, Baer HJ. 2011. Body size in
- 31 early life and adult levels of insulin-like growth factor 1 and insulin-like growth factor binding
- 32 protein 3. *American Journal of Epidemiology* 174:642-651.
- 33 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker
- 34 PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based
- 35 linkage analyses. *American Journal of Human Genetics* 81:559–575.
- Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018. The genomic landscape at
- a late stage of stickleback speciation: High genomic divergence interspersed by small localized
- 38 regions of introgression. *PLoS Genetics* 14: e1007358.

- 1 Rheindt FE, Székely T, Edwards SV, Lee PLM, Burke T, Kennerley PR, Bakewell DN, Alrashidi M,
- 2 Kosztolányi A, Weston MA, et al. 2011. Conflict between genetic and phenotypic differentiation: the
- 3 evolutionary history of a 'lost and rediscovered' shorebird. *PLoS ONE* 6: e26995.
- 4 Ruan J. 2020. Fast and accurate long-read assembly with wtdbg2. *Nature Methods* 17:155-158.
- 5 Sadanandan KR, Küpper C, Low GW, Yao C Te, Li Y, Xu T, Rheindt FE, Wu S. 2019. Population
- 6 divergence and gene flow in two East Asian shorebirds on the verge of speciation. *Scientific Reports*
- 7 9: 8546.
- 8 Schluter D, Rieseberg LH. 2022. Three problems in the genetics of speciation by selection.
- 9 Proceedings of the National Academy of Sciences of the United States of America 119: e2122153119.
- 10 Shimokawa I, Higami Y, Utsuyama M, Tuchiya T, Komatsu T, Chiba T, Yamaza H. 2002. Life span
- 11 extension by reduction in growth hormone-insulin-like growth factor-1 axis in a transgenic rat
- 12 model. *The American Journal of Pathology* 160:2259-2265.
- 13 Song Z, Que P, Jones W, Huang C, Halimubieke N, Ding P, Zhang Z, Liu Y. 2023. Colder ambient
- 14 temperatures constrain female mate preference for ornamental traits. Preprints Available from:
- 15 https://www.authorea.com/users/660071/articles/663768-colder-ambient-temperatures-constrain-
- 16 female-mate-preference-for-ornamental-
- 17 traits?commit=9039c45438470953e5a76c2f6291624500e396a8
- 18 Szpiech ZA, Hernandez RD. 2014. Selscan: An efficient multithreaded program to perform EHH-based
- scans for positive selection. *Molecular Biology and Evolution* 31:2824-2827.
- Szpiech ZA, Novak TE, Bailey NP, Stevison LS. 2021. Application of a novel haplotype-based scan for
   local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution Letters* 5:408-421.
- Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time
   durations. *Journal of Biogeography* 27:1153-1167.
- 24 Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, Fan RX, Yin TT, Zhu CL, Poyarkov AD, et al.
- 25 2016. Out of southern East Asia: The natural history of domestic dogs across the world. *Cell Research*26:21-33.
- Wang S, Coop G. 2022. A complex evolutionary history of genetic barriers to gene flow in hybridizing
   warblers. :2022.11.14.516535. Available from:
- 29 https://www.biorxiv.org/content/10.1101/2022.11.14.516535v1
- 30 Wang S, Rohwer S, de Zwaan DR, Toews DPL, Lovette IJ, Mackenzie J, Irwin D. 2020. Selection on a
- small genomic region underpins differentiation in multiple color traits between two warbler species.
   *Evolution Letters* 4:502-515.
- 33 Wang X, Maher KH, Zhang N, Que P, Zheng C, Liu S, Wang B, Huang Q, Chen D, Yang X, et al. 2019.
- 34 Demographic histories and genome-wide patterns of divergence in incipient species of shorebirds.
- 35 Frontiers in Genetics 10: 919.
- 36 Wang X, Que P, Heckel G, Hu J, Zhang X, Chiang CY, Zhang N, Huang Q, Liu S, Martinez J, et al. 2019.
- 37 Genetic, phenotypic and ecological differentiation suggests incipient speciation in two *Charadrius*
- 38 plovers along the Chinese coast. *BMC Evolutionary Biology* 19: 135.

- 1 Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV,
- 2 Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and
- 3 phylogenomics. *Molecular Biology and Evolution* 35:543-548.
- 4 Wickham H. 2011. Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* 3:180-185.
- 5 Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation.
  6 Nature Reviews Genetics 18:87-100.
- 7 Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, et al. 2021. rMVP: A memory-
- 8 efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study.
- 9 *Genomics, Proteomics & Bioinformatics* 19: 619-628.
- 10





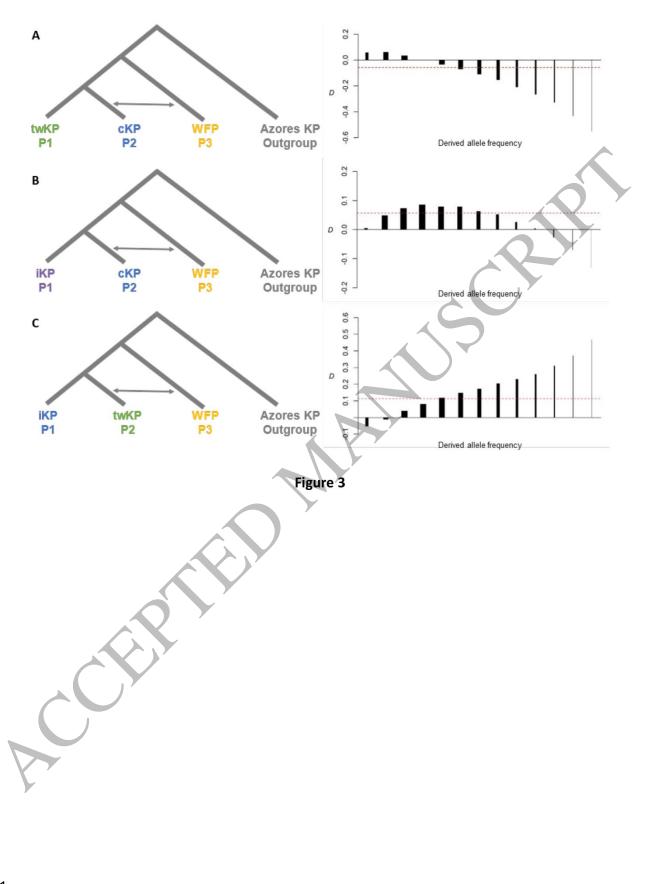


Figure 3 169x244 mm ( x DPI)

