

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

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

Wenjun Zhou^{1,2}, Nan Zhang¹, Kaichi Huang³, Hongzhou Lin¹, Jie Tu¹, Chenqing Zheng¹, Pinjia Que^{4,5}, Chung-Yu Chiang⁶, Jonathan Martinez⁷, Halimubieke Naerhulan⁸, Tamás Székely^{1,4,8}, Zhengwang Zhang⁴, Yang Liu^{1,8*}

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

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

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
5 these genomic island. This conundrum can be solved through examination of the genomic
6 landscapes of species pairs that are still in the early stages of speciation. In this study,
7 population genomics analyses were undertaken using a wide range of sampling and whole-
8 genome resequencing data from 96 unrelated individuals of Kentish plover (*Charadrius*
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.
11 Genome-wide analyses for introgression indicate that ancient introgression had occurred in
12 Taiwan population, and recurrent gene flow is still ongoing in mainland coastal populations.
13 Furthermore, we identified a few genomic regions with significant levels of interspecific
14 differentiation and local recombination suppression, which contain several genes potentially
15 associated with disease resistance, coloration, and regulation of plumage moulting, thus
16 may be connected to the phenotypic and ecological divergence of the two nascent species.
17 Overall, our findings suggest that divergent selection in low recombination regions may be
18 the main force in shaping the genomic islands in two incipient shorebird species.

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

21

1 window proportion reaches 77.04% between coastal KPs and WFPs (Table S2). D_{FS} pattern
2 under this scenario shows a more distinctive shape of introgression between P2 (coastal
3 KPs) and P3 (WFPs), where both the total D value and the D values on low to intermediate-
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
6 flow, as the absolute values of total D s involved in the D_{FS} patterns are lower than 0.1.

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_{FS} pattern, where negative values appear on low-
10 frequency alleles and the same values become positive and rise from intermediate to high-
11 frequency alleles (Figure 3C). Consistent to what has been shown in previous sections, this
12 result shows no recent gene flow from P3 (mainland WFPs) to P2 (Taiwan KPs). However,
13 the absolute total D value under this tree structure is the highest among all three hypothesis
14 tests, and the whole D_{FS} shape is skewed towards high-frequency alleles. Combined with the
15 negative total D value in Figure 3A, this may be taken as another indication that there once
16 was extensive gene flow from mainland areas to the Taiwan KPs, but this event was cut off,
17 leaving only a limited number of introgressed alleles gradually being fixed in the island
18 population.

19

20 **Genomic islands between coastal KP and WFP**

21 Although recent gene flow is suggested by previous results and coastal KPs and WFPs share
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_{ST} 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
30 down to zero in some sliding windows.

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
10 coastal KPs compared to WFPs outside the contact zone (Figure 4F). However, the fact that
11 genes mentioned above locate within genomic regions with different selective strength
12 suggests that the existence of genomic islands may not only be due to phenotypic
13 differences. For example, even though the *ZBTB20* and *DRD3* genes locate within the 83-
14 84Mb genomic region which shows no difference in cross-population selection comparison
15 (Figure 4E and 4F), they have important immune function towards avian influenza which is
16 crucial regardless of the individuals' population and their phenotypic difference.

17 In this study, we show that most identified genes that locate within the high divergence
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
22 suggested that selection favours restricted recombination (Felsenstein 1981; Butlin 2005),
23 the existence of recombination variation increase the difficulty to conclude on whether
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
26 contact zone, potentially between Wenzhou and Ningde (26.66° N, 119.55° E - 27.99° N,
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.

1

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).

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

21

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

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

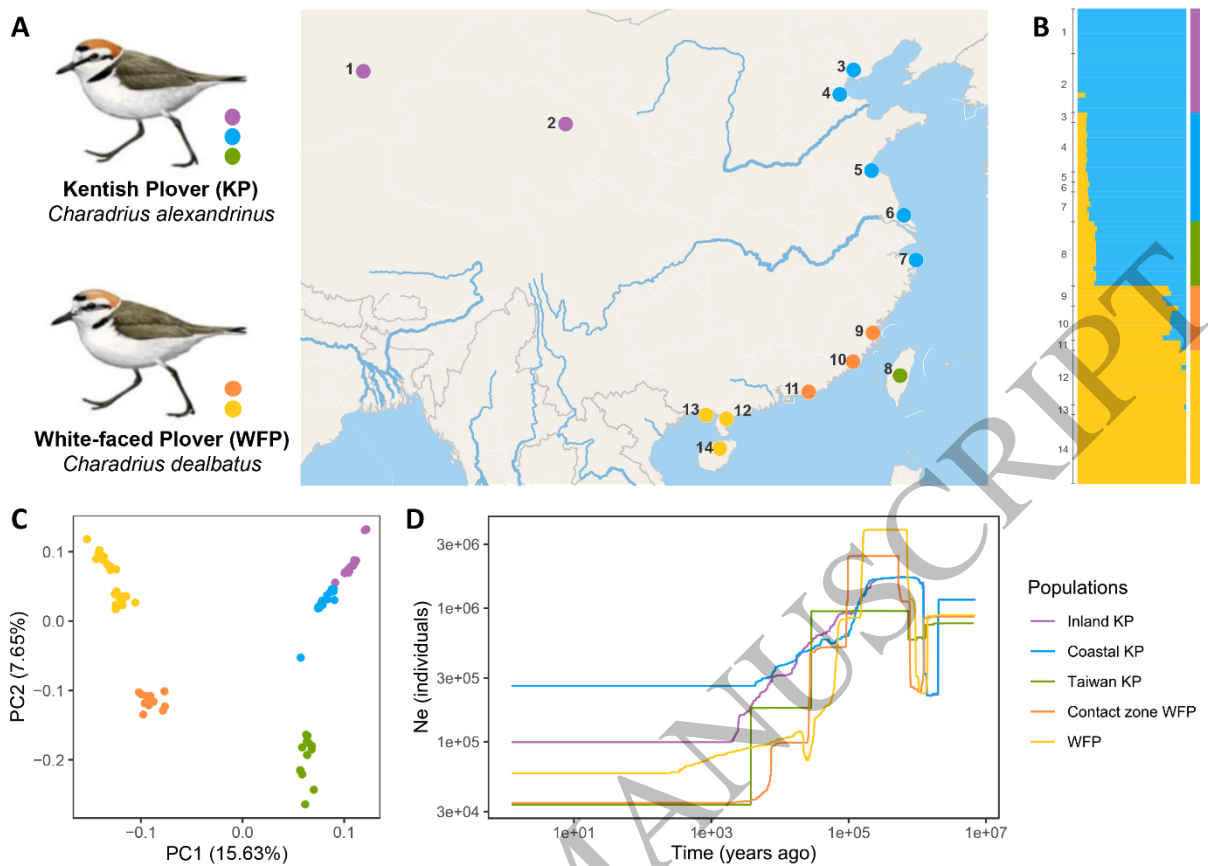
- 1 Rheindt FE, Székely T, Edwards SV, Lee PLM, Burke T, Kennerley PR, Bakewell DN, Alrashidi M,
2 Kosztolányi A, Weston MA, et al. 2011. Conflict between genetic and phenotypic differentiation: the
3 evolutionary history of a 'lost and rediscovered' shorebird. *PLoS ONE* 6: e26995.
- 4 Ruan J. 2020. 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-](https://www.authorea.com/users/660071/articles/663768-colder-ambient-temperatures-constrain-female-mate-preference-for-ornamental-traits?commit=9039c45438470953e5a76c2f6291624500e396a8)
16 [female-mate-preference-for-ornamental-](https://www.authorea.com/users/660071/articles/663768-colder-ambient-temperatures-constrain-female-mate-preference-for-ornamental-traits?commit=9039c45438470953e5a76c2f6291624500e396a8)
17 [traits?commit=9039c45438470953e5a76c2f6291624500e396a8](https://www.authorea.com/users/660071/articles/663768-colder-ambient-temperatures-constrain-female-mate-preference-for-ornamental-traits?commit=9039c45438470953e5a76c2f6291624500e396a8)
- 18 Szpiech ZA, Hernandez RD. 2014. Selscan: An efficient multithreaded program to perform EHH-based
19 scans for positive selection. *Molecular Biology and Evolution* 31:2824-2827.
- 20 Szpiech ZA, Novak TE, Bailey NP, Stevison LS. 2021. Application of a novel haplotype-based scan for
21 local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution Letters* 5:408-421.
- 22 Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time
23 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 26:21-33.
- 27 Wang S, Coop G. 2022. A complex evolutionary history of genetic barriers to gene flow in hybridizing
28 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
31 small genomic region underpins differentiation in multiple color traits between two warbler species.
32 *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

11

ACCEPTED MANUSCRIPT



1
2
3
4

Figure 1
243x178 mm (x DPI)

ACCEPTED MANUSCRIPT

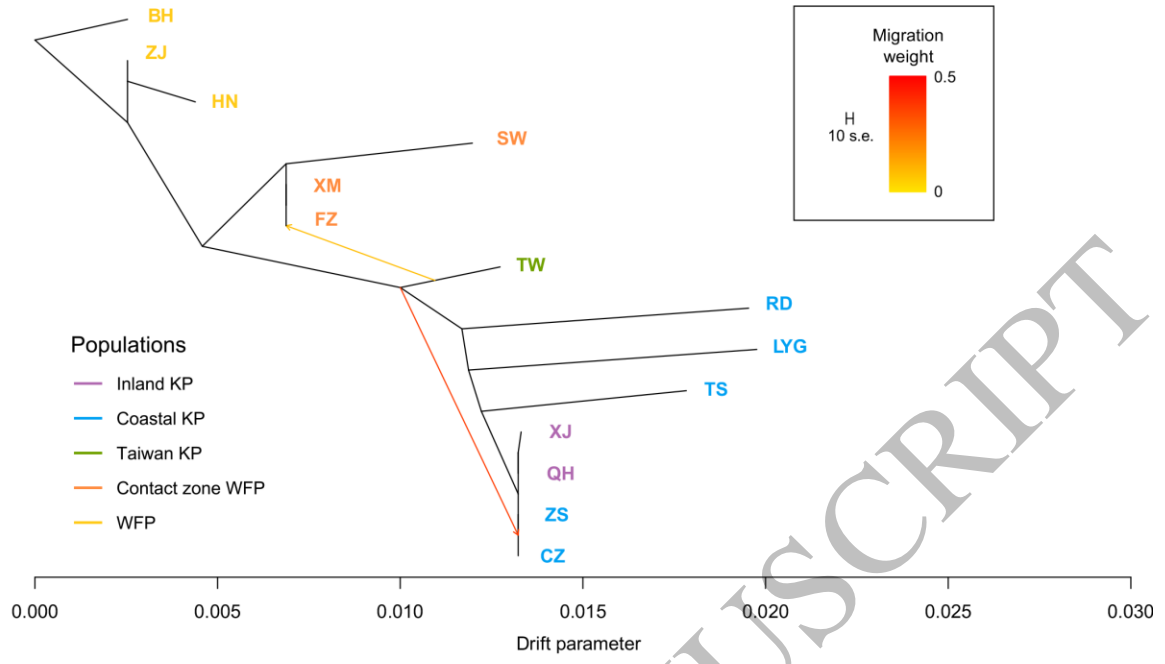


Figure 2

Figure 2
188x244 mm (x DPI)

1
2
3
4

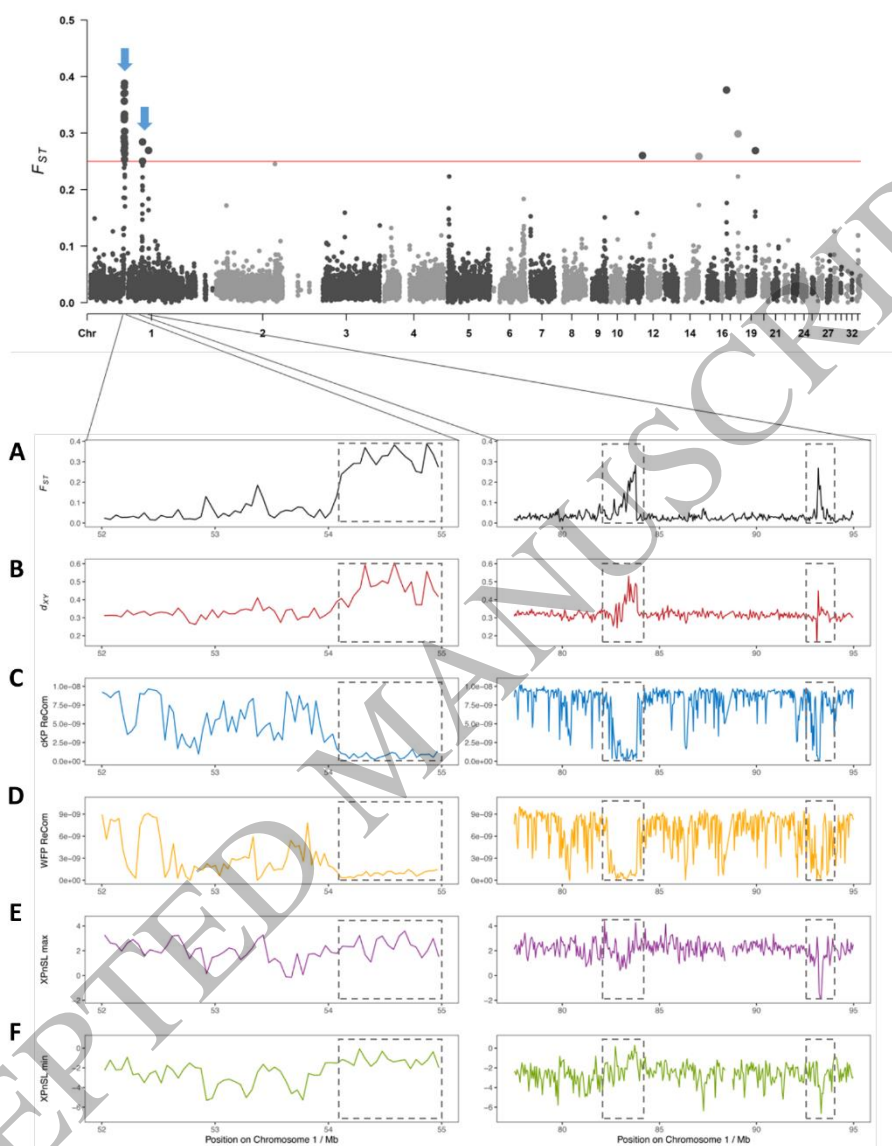


Figure 4

Figure 4
150x244 mm (x DPI)

1
2
3
4

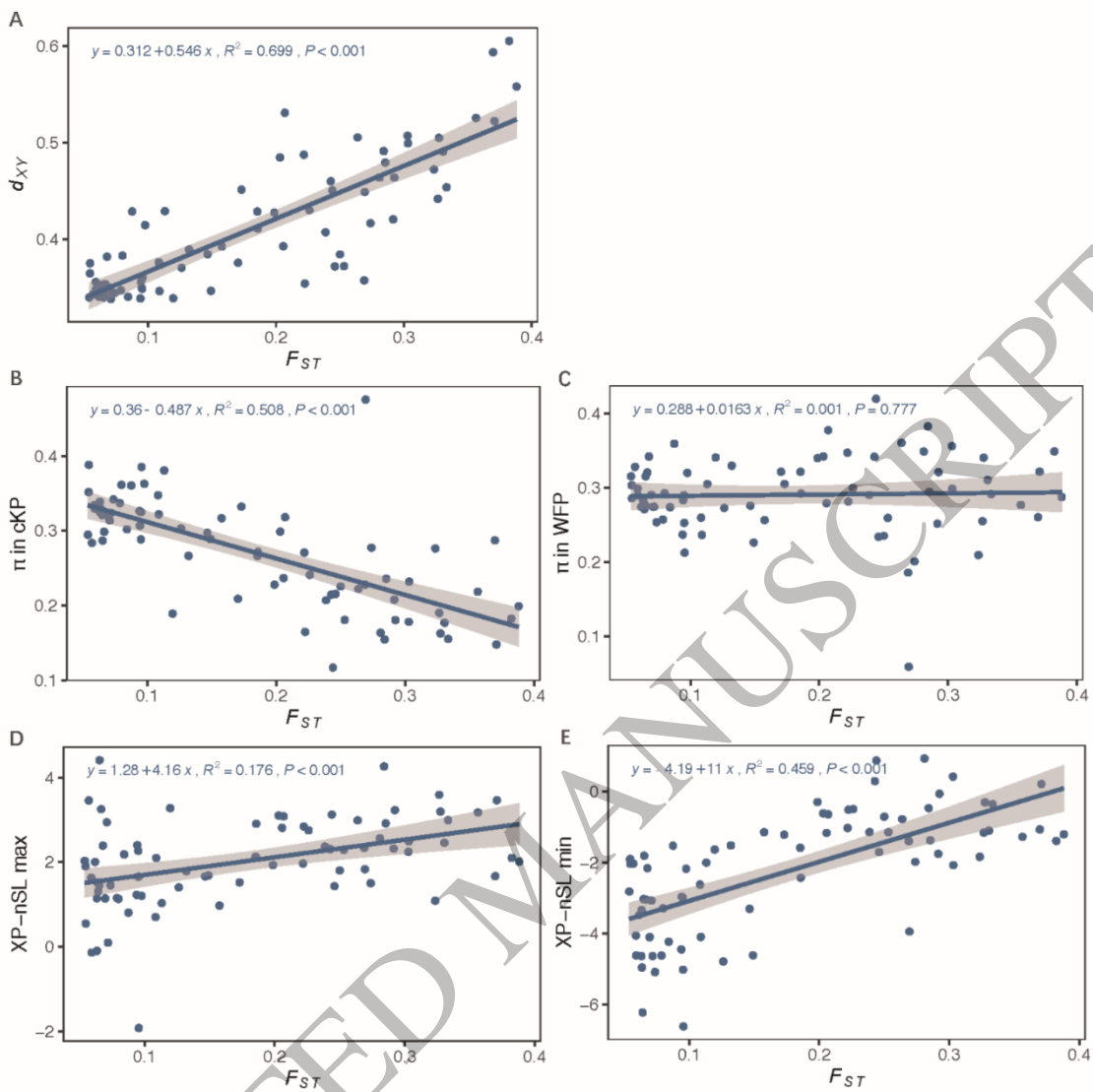


Figure 5

Figure 5
217x244 mm (x DPI)

1
2
3