

1 Correlated genomic landscapes in birds

2

3 ORIGINAL ARTICLE

4

5 **Correlated patterns of genetic diversity and differentiation across an avian family**

6

7 **Benjamin M. Van Doren^{1,2*} (BVD), Leonardo Campagna^{1,2} (LC), Barbara Helm³ (BH),**

8 **Juan Carlos Illera⁴ (JCI), Irby J. Lovette^{1,2} (IJL), and Miriam Liedvogel⁵ (ML)**

9

10 ¹ Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

11 ² Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA

12 ³ Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow,

13 Glasgow G12 8QQ, UK

14 ⁴ Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, Campus of Mieres, Research

15 Building, 5th Floor, C/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain

16 ⁵ Max Planck Institute for Evolutionary Biology, AG Behavioural Genomics, August-

17 Thienemann-Str. 2, 24306 Plön, Germany

18 * Corresponding author: benjamin.vandoren@zoo.ox.ac.uk. Current address: Edward Grey

19 Institute, Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road,

20 Oxford, OX1 3PS, UK

21 Keywords: Differentiation peaks, *Ficedula*, Genetic drift, Genomic scan, Linked selection,

22 Pooled sequencing, *Saxicola*.

23 **ABSTRACT**

24

25 Comparative studies of closely related taxa can provide insights into the evolutionary forces that
26 shape genome evolution and the prevalence of convergent molecular evolution. We investigated
27 patterns of genetic diversity and differentiation in stonechats (genus *Saxicola*), a widely
28 distributed avian species complex with phenotypic variation in plumage, morphology, and
29 migratory behavior, to ask whether similar genomic regions have become differentiated in
30 independent, but closely related, taxa. We used whole-genome pooled sequencing of 262
31 individuals from 5 taxa and found that levels of genetic diversity and divergence are strongly
32 correlated among different stonechat taxa. We then asked if these patterns remain correlated at
33 deeper evolutionary scales and found that homologous genomic regions have become
34 differentiated in stonechats and the closely related *Ficedula* flycatchers. Such correlation across a
35 range of evolutionary divergence and among phylogenetically independent comparisons suggests
36 that similar processes may be driving the differentiation of these independently evolving lineages,
37 which in turn may be the result of intrinsic properties of particular genomic regions (e.g., areas of
38 low recombination). Consequently, studies employing genome scans to search for areas
39 important for reproductive isolation or adaptation should account for corresponding regions of
40 differentiation, as these regions may not necessarily represent speciation islands or evidence of
41 local adaptation.

42

43

44

45

46 INTRODUCTION

47
48 Evolutionary biologists seek to understand the genetic basis of speciation and the degree to which
49 the divergence of lineages may involve independent changes on similar loci (Seehausen et al.
50 2014). Genomic sequencing has made it possible to examine patterns of differentiation across the
51 genomes of organisms at different stages of divergence. Recent comparative studies of genome-
52 wide patterns of variation, or “genomic landscapes,” have identified areas of the genome that are
53 conspicuously differentiated relative to the genomic baseline among closely related taxa
54 (Ellegren et al. 2012; Ruegg et al. 2014; Burri et al. 2015; Wang et al. 2016). It remains uncertain
55 whether these regions are functionally important in speciation, and whether they typically arise
56 during speciation-with-gene-flow or as a consequence of selection in allopatry.

57 Most empirical studies have used F-statistics (Wright 1965) and other measures that
58 compare allele frequencies between two populations to infer the magnitude of differentiation
59 across the genome. These statistics are influenced by levels of within-population genetic variation
60 and are therefore classified as “relative” measures of divergence (Hedrick 2005). Genomic outlier
61 regions of high differentiation were first described as “islands of speciation” in the face of gene
62 flow (Turner et al. 2005) and hypothesized to harbor loci that were important for reproductive
63 isolation (Nosil et al. 2009; Feder et al. 2012). However, subsequent studies have identified
64 alternate mechanisms by which isolated genomic regions of elevated differentiation can be
65 generated in allopatry, and thus independently of gene flow (e.g., Noor and Bennett 2009; Turner
66 and Hahn 2010; White et al. 2010; Cruickshank and Hahn 2014). For example, post-speciation
67 selective sweeps or background selection, especially in regions of reduced recombination, can

68 drive the differentiation of these loci relative to the rest of the genome (Nachman and Payseur
69 2012; Cruickshank and Hahn 2014; Burri et al. 2015).

70 Selective sweeps bring beneficial alleles to high frequency in a population, greatly
71 reducing genetic diversity at linked sites via “hitchhiking” (Smith and Haigh 1974; Kaplan et al.
72 1989). The magnitude of the hitchhiking effect is influenced by recombination rate, in addition
73 to the strength of selection, with areas of low recombination experiencing greater linkage and a
74 commensurate reduction in diversity across larger sections of a chromosome (Begun and
75 Aquadro 1993; Charlesworth et al. 1997; Nielsen 2005). This reduction in local within-
76 population genetic diversity results in high differentiation as measured by F_{ST} (Charlesworth
77 1998; Keinan and Reich 2010; Cruickshank and Hahn 2014). Recurrent sweeps in similar
78 genomic regions across independent populations may be caused by selection for different
79 advantageous alleles at the same locus, or by selection on different but tightly linked loci.
80 Alternatively, they could result from the adaptive introgression of globally advantageous
81 mutations transmitted among populations by gene flow, followed by sweeps due to local
82 adaptation (see Roesti et al. 2014; Delmore et al. 2015). A similar pattern could also arise from
83 the increased establishment probability of beneficial mutations linked to selected sites in areas of
84 low recombination (Yeaman et al. 2016). These related processes can result in corresponding
85 areas of low genetic diversity and high differentiation in independent population comparisons.

86 Similarly, background (or purifying) selection purges deleterious alleles as they arise and
87 may also independently generate similar genomic landscapes of diversity and differentiation
88 across populations (Charlesworth 2013; Burri et al. 2015; Wang et al. 2016). Under this scenario,
89 a neutral variant that emerges in a population will subsequently disappear if it is linked to a
90 deleterious mutation, a process that reduces nucleotide diversity (Charlesworth et al. 1993;

91 Charlesworth et al. 1997; Stephan et al. 1998). As recombination rates decrease, linkage extends
92 over larger genetic distances and the probability of a neutral variant associating with a deleterious
93 mutation (and thus being purged) is higher. Therefore, areas of low recombination will generally
94 exhibit a greater reduction in genetic diversity due to background selection (Charlesworth et al.
95 1993; Nordborg et al. 1996). If highly conserved genomic regions (e.g., of great functional
96 importance) and/or areas of low recombination are similar across species, the effect of
97 background selection may cause, or contribute to, parallel genomic landscapes of diversity and
98 differentiation in comparisons of independently evolving lineages (Nordborg et al. 1996;
99 Andolfatto 2001; Cruickshank and Hahn 2014).

100 The effects of background selection and selective sweeps on linked neutral loci have
101 collectively been referred to as “linked selection” (Turner and Hahn 2010; Cutter and Payseur
102 2013; Cruickshank and Hahn 2014). The frequency with which parallel signatures of linked
103 selection occur in closely related taxa and the contributions of background selection and selective
104 sweeps in shaping genomic landscapes remain actively debated (Keinan and Reich 2010; Burri et
105 al. 2015). In addition, the degree to which this parallelism may extend beyond a few well-studied
106 species complexes is currently unknown.

107 Genome-wide scans of two independent groups of closely related bird species, *Ficedula*
108 flycatchers and *Phylloscopus* warblers, have identified conspicuous peaks of relative divergence
109 (i.e., genomic regions with very different allele frequencies) in pairwise comparisons of
110 congeners that coincide with “valleys” of absolute divergence (i.e., regions with few sequence
111 differences, not influenced by within-population genetic diversity) (Burri et al. 2015; Irwin et al.
112 2016). This inverse relationship is inconsistent with the speciation-with-gene-flow paradigm, in
113 which regions of high relative divergence are resistant to gene flow and therefore should show

114 *high* absolute divergence (Noor and Bennett 2009; Nachman and Payseur 2012; Cruickshank and
115 Hahn 2014). This suggests that post-speciation selection—not divergence-with-gene-flow—
116 generates differentiation peaks in these systems. Within their respective species complexes,
117 flycatchers and warblers show signatures of selection in similar genomic areas (Burri et al. 2015;
118 Irwin et al. 2016), but neither the specific type of selection, nor their contribution to the
119 speciation process, has been fully characterized. Furthermore, these studies primarily test the
120 correspondence of divergent regions using correlation-based methods, which can be strongly
121 affected by pseudoreplication due to linkage.

122 Here, we characterize the course of genome-wide molecular evolution in a well-studied
123 group of birds, the *Saxicola* stonechats (Urquhart 2002; Collar 2016a, 2016b). This genus began
124 diversifying during the late Miocene (8.2 million years ago; Illera et al. 2008) and currently
125 comprises 15 recognized species (51 named taxa including subspecies; Gill and Donsker 2016).
126 Some taxa are restricted to small islands, while others span continents, and they range from long
127 distance migrants to year-round residents (Baldwin et al. 2010). The well-documented
128 evolutionary diversity in this clade makes *Saxicola* a powerful system for studying independently
129 evolving lineages across a gradient of differentiation, phenotypic variation, and life histories. We
130 examine five stonechat taxa at disparate stages of divergence, including two that likely still
131 exchange genes and two that diverged ~3.7 million years ago (Illera et al. 2008). These taxa show
132 variation in morphology and behavior (e.g., body size and migratory direction), and we survey
133 both island and continental taxa, which are likely to have varied demographic histories.

134 Our primary research focus is to investigate the extent to which genome evolution is
135 correlated in independently evolving, but closely related, taxa. We ask: Have the same regions of
136 the genome become differentiated over time in independent stonechat lineages? If so, what role

137 has natural selection played in driving this correlated differentiation? We further ask if evolution
138 is correlated at a deeper scale, between stonechats and two *Ficedula* flycatchers (Sætre and
139 Sæther 2010). Both genera belong to the family Muscicapidae. We posit that any loci that are
140 differentiated in both genera are unlikely to arise from parallel ecological selection pressures and
141 instead stem from intrinsic properties of those genomic regions. Finally, we examine the effects
142 of life history and demography on the genome by comparing patterns of genetic diversity and
143 differentiation between continental and island taxa. We hypothesize that an island taxon will
144 show a weaker overall effect of selection on the genome, reflecting the theoretical prediction of
145 increased drift with a smaller effective population size. Our goal is to shed light on the processes
146 that influence the most conspicuous features—the high “peaks” and low “valleys”—of stonechat
147 genomic landscapes. The results underscore the degree to which broad patterns of genetic
148 diversity and differentiation are correlated across evolutionary time.

149

150 **METHODS**

151 **Study system and sampling**

152 We included five stonechat taxa in this study: *Saxicola rubicola rubicola* from Austria (European
153 stonechat); *S. rubicola hibernans* from Ireland (European stonechat); *S. torquatus axillaris* from
154 Kenya (African stonechat); *S. maurus maurus* from Kazakhstan (Siberian stonechat); and *S.*
155 *dacotiae dacotiae* from Fuerteventura Island, Spain (Canary Islands stonechat) (Gill and Donsker
156 2016). Using mitochondrial DNA, Illera et al. (2008) estimated that African stonechats diverged
157 from the remaining four taxa about 3.7 mya, Siberian stonechats subsequently split from the
158 remaining three about 2.5 mya, and Canary Islands stonechats diverged from European
159 stonechats about 1.6 mya. Illera and colleagues could not distinguish Austrian and Irish

160 stonechats using mitochondrial DNA. We expect that Canary Islands stonechats have diverged
161 from other taxa without gene flow because this taxon occurs on an oceanic island, and the
162 present-day ranges of Kenyan and Siberian stonechats lead us to expect no ongoing gene flow
163 between these and the other taxa. Conversely, we expect that Austrian and Irish stonechats likely
164 still exchange genes because of their close geographic proximity, lack of mitochondrial
165 divergence, and evidence of breeding dispersal between the British Isles and continental Europe
166 (Helm et al. 2006).

167 Most of the 262 stonechats included in this study originated from a common-garden
168 experiment that Eberhard Gwinner initiated in 1981 at the Max-Planck Institute in Andechs,
169 Germany, except for Canary Islands stonechats, which were directly sampled in the wild (Table
170 S1). For the other species, most birds were taken into captivity as nestlings, except for Irish
171 stonechats (~50% captured on winter territories). The remaining sampled individuals were
172 offspring of these captive stonechats, hatched between 1988 and 2006. Despite the inclusion of
173 captive birds, relatedness within the pools was low (Table S1). Detailed descriptions of breeding
174 and raising conditions are published elsewhere (Gwinner et al. 1987; Helm 2003; Helm et al.
175 2009).

176 The inclusion of second-generation progeny in our study could potentially lower
177 measured levels of genetic diversity relative to a comparable sample of wild individuals.
178 However, we find average genetic diversity (π) to be highest in Siberian stonechats, the species
179 for which we incorporated the most captive-bred birds; this suggests that any putative bias is
180 small and potentially negligible for the purposes of this study.

181

182 **Draft reference genome**

183 We assembled the genome of a male Siberian stonechat (*S. maurus*) collected in Kazakhstan
184 (44.59° N, 76.609° E) and housed at the Burke Museum (UWBM# 46478). We generated one
185 fragment library with insert sizes of 180 base pairs (bp) and two mate-pair libraries (insert sizes:
186 3 and 8 kilobases), and we sequenced each of them on one Illumina HiSeq 2500 lane (obtaining
187 101-bp paired-end reads). We assembled the draft reference genome using the ALLPATHS-LG
188 algorithm (Gnerre et al. 2011) and used HaploMerger (Huang et al. 2012) to improve the
189 assembly by merging homologous scaffolds and removing those resulting from the erroneous
190 split of two haplotypes into separate scaffolds. The final Siberian stonechat assembly comprised
191 2,819 scaffolds, with a total scaffold length of 1.02 Gb and an N50 scaffold size of 10.0 Mb. Half
192 of the final assembly is represented in 24 scaffolds, and 75% in 65 scaffolds. Ambiguous bases
193 (N's) make up 4.4% of its total length.

194 We assembled scaffolds from our stonechat reference genome into draft chromosomes by
195 mapping them to the *Ficedula albicollis* genome assembly, version 1.5 (RefSeq accession
196 GCF_000247815.1) (Kawakami et al. 2014) and used SatsumaSynteny (Grabherr et al. 2010) to
197 align the *Saxicola* draft genome to the *F. albicollis* assembly. Because these species are
198 phylogenetically close and synteny is relatively conserved among birds (Ellegren 2010), this
199 method allowed us to position 97.1% of the stonechat reference genome in the presumed correct
200 order. Inversions and other chromosomal rearrangements occur in birds (Backström et al. 2008),
201 so it is possible that a small percentage of the genome may be ordered or oriented incorrectly.

202

203 **Resequencing of five stonechat taxa**

204 We extracted genomic DNA from stonechat blood or tissue samples using a salt extraction
205 protocol and selected 49-56 individuals ($n_{\text{total}} = 262$, including both males and females) from each

206 of the five stonechat taxa for sequencing (Table S1). We created five pooled libraries, one per
207 taxon, from equimolar aliquots of DNA using the Illumina TruSeq DNA kit and sequenced them
208 on an Illumina NextSeq 500 (Table S2).

209 We used BWA-MEM (Li 2013) to align sequences to our reference genome and
210 performed refinement and quality control steps with Picard
211 (<http://broadinstitute.github.io/picard/>) and the Genome Analysis Toolkit (GATK) (McKenna et
212 al. 2010), including filtering by mapping quality and removing duplicate sequences (Supporting
213 information). Sequences mapped to the draft reference genome at a mean per-pool coverage
214 between 13.8x and 26.1x, and mean mapping quality was between 45-46 for all taxa (Table S2).
215 Although this level of coverage is insufficient to sequence every individual at every locus, the
216 goal of our pooled sequencing strategy was to estimate population allele frequencies by sampling
217 a subset of chromosomes in a pool. Gautier et al. (2013) found that allele frequencies of
218 individual SNPs estimated with 10-50x pool coverage (pool size = 30) were strongly correlated
219 with estimates derived from separate individual-based ($n = 20$) sequencing at 1x-6x ($r = 0.93$) and
220 6-10x ($r = 0.94$) per individual. Additionally, the effects of pool-derived sampling error are
221 greatly reduced in window-based analyses where variation and differentiation are summarized
222 across groups of SNPs (Kofler et al. 2011a). Because we use a windowed approach and therefore
223 do not rely on the frequencies of individual SNPs, we are confident that we can accurately assess
224 and compare genome-wide patterns of genetic variation with this level of coverage.

225

226 **SNP-based phylogeny and inter-taxa divergence**

227 Although we used an existing mitochondrial phylogeny (e.g., Illera et al. 2008) as a basis for our
228 study approach and design, we also constructed a phylogenetic tree for the five focal stonechat

229 taxa using nuclear markers. This step was designed to confirm the mitochondrial findings and
230 serve as a basis for phylogeny-based inference. We used Pied and Collared Flycatchers (*Ficedula*
231 *hypoleuca* and *F. albicollis*) as outgroups. We constructed a maximum likelihood tree with
232 RAxML (Stamatakis 2014), using 16,876,859 fixed single-nucleotide polymorphisms (SNPs)
233 from across the nuclear genome, of which 330,592 were polymorphic among the stonechat
234 ingroup (Supporting information). We applied the Lewis correction, following the
235 recommendation of Stamatakis (2014), for ascertainment bias resulting from the exclusion of
236 invariant sites.

237 We generated mean genome-wide F_{ST} and d_{XY} pairwise distance matrices for the five
238 focal stonechat taxa and displayed them graphically using principal coordinate analyses
239 performed with the *ape* package in R (Paradis et al. 2004).

240

241 **Pooled population genomic analyses**

242 We analyzed sequence data with the software packages *npstat* (Ferretti et al. 2013) and
243 *Popoolation2* (Kofler et al. 2011b), designed specifically for the analysis of pooled sequencing
244 data. With *npstat* we calculated: (1) Tajima's D , to test for rare variants as a signal of directional
245 or purifying selection or large-scale demographic effects; (2) π , an estimate of genetic diversity,
246 which is derived from the number of pairwise sequence differences among members of a
247 population; and (3) Fay and Wu's H , a statistic related to Tajima's D but sensitive only to high
248 frequency derived alleles, thus influenced by positive selection but not by background selection
249 (Fay and Wu 2000). We polarized alleles using the Collared Flycatcher.

250 We then used *Popoolation2* to calculate pairwise F_{ST} among all pairs of taxa. We also
251 estimated d_{XY} (Nei and Li 1979; Cruickshank and Hahn 2014), a measure of absolute divergence,

252 as $A_X B_Y + A_Y B_X$, where A and B are the frequencies of the two alleles at a locus and X and Y
253 denote the two groups being compared.

254 For all analyses, we excluded bases within 5 bp of indels to reduce the probability of
255 including erroneous genotypes due to misalignments. We calculated all metrics for 50-kb non-
256 overlapping windows (Supporting information), within which we only considered sites with
257 minor allele counts ≥ 2 and coverage between half and three times that of the pool's average. We
258 only retained windows in which at least 40% of bases (i.e. 20 kb) satisfied this coverage criterion.
259 For d_{XY} , we calculated the windowed value by summing over the window and dividing by the
260 total number of sites with sufficient coverage (variable or not). We calculated standardized
261 nucleotide diversity for each taxon by dividing π by the maximum d_{XY} value from all pairwise
262 comparisons involving that taxon (following Irwin et al. 2016).

263

264 **Correlation analyses**

265 We first quantified the similarity of genome-wide patterns of genetic diversity and divergence
266 using Spearman rank correlations. Although the p-values of these tests are affected by
267 pseudoreplication due to the inclusion of genetically linked loci, they are nonetheless a valuable
268 summary of genome-wide similarity and provide a means to compare the results of the present
269 study with previous work.

270

271 **Identification of genomic outlier regions**

272 We identified regions of the stonechat genome showing consistently elevated or lowered values
273 of Tajima's D , π , Fay and Wu's H , F_{ST} , and d_{XY} , and therefore may be important in the
274 divergence of stonechat lineages. In particular, we wanted to determine whether any genome-

275 wide similarities revealed by the correlation analyses were driven by a relatively small number of
276 genomic regions. We applied a kernel-based smoothing algorithm across 50-kb windows (box
277 density with bandwidth of 30; see Supporting information) and compared this smoothed line with
278 25,000 smoothed lines obtained after permuting the order of the windows (see Ruegg et al. 2014).
279 We called outlier locations where the observed smoothed line was more extreme than the most
280 extreme smoothed value from the null (permutation) distribution. We merged outlier regions
281 separated by four windows or fewer (i.e. by <200 kb). Because the effective population size of
282 the Z chromosome is smaller than that of the autosomes, baseline levels of variation and
283 differentiation are different from those of autosomes (Charlesworth 2001). We therefore
284 permuted windows of the Z chromosome and autosomes separately (see Fig. S1, Supporting
285 information).

286

287 **Concordance of genomic outlier regions within *Saxicola***

288 Once outlier regions were identified, we assessed their overlap among stonechat taxa. For each
289 pairwise comparison, we counted the number of outlier regions that showed any degree of
290 overlap between the two datasets. By considering each region separately, we account for
291 autocorrelation of their constituent windows due to linkage. Although this approach addresses the
292 pseudoreplication that would have resulted from treating the multiple windows within the same
293 outlier region as independent observations, it is important to note that it does not address the
294 larger-scale possibility that multiple outlier *regions* could be clustered together, e.g. due to a very
295 large region of reduced recombination.

296 We then tested whether the observed number of overlapping regions was significantly
297 greater than expected under the null hypothesis of no association in outlier positions between

298 datasets, using a custom permutation test. While holding the number and size of outlier regions
299 constant, we randomly permuted their locations across the genome 1,000 times and measured the
300 degree of overlap under these simulated scenarios. The p-value of the test was the proportion of
301 simulations under which the number of overlapping outlier regions was equal to or greater than
302 the observed value; we thus accounted for the varying number and size of outlier regions in each
303 comparison. We applied a false discovery rate correction to each series of tests (Benjamini and
304 Hochberg 1995) and considered tests with corrected p-values less than 0.05 to be statistically
305 significant.

306 For each comparison, we calculate the proportion of outlier regions in one genomic
307 landscape also present in the other (and vice versa) and report the greater of these two values.
308 Thus, if landscape 1 shows 10 peaks and landscape 2 shows 50 peaks, and 9 out of 10 peaks in
309 landscape 1 are also present in landscape 2, our outlier similarity score will be $9/10 = 0.90$.

310 Previous studies have used scatterplots and correlation analyses as the primary manner of
311 assessing association between outlier regions in independent comparisons (e.g., Burri et al. 2015;
312 Irwin et al. 2016). However, these tests are affected by autocorrelation due to genetic linkage. By
313 considering each outlier region as a single unit, our permutation approach overcomes this issue
314 by treating each contiguous outlier region (instead of each 50-kb window) as an independent
315 observation.

316

317 **Correspondence of genomic landscapes between *Saxicola* and *Ficedula***

318 To test for conservation of genomic landscapes at a deeper level of divergence, we compared
319 stonechat genomic landscapes to those of the genus *Ficedula*. We calculated F_{ST} , d_{XY} , and π for
320 Pied Flycatchers (*F. hypoleuca*) and Collared Flycatchers (*F. albicollis*). We downloaded reads

321 from the Sequence Read Archive (project ERP007074; accession PRJEB7359;
322 <http://www.ncbi.nlm.nih.gov/sra>) for 10 individuals of each species (Smeds et al. 2015) (Table
323 S3), and processed reads for quality as described for the stonechat analysis (Supporting
324 information).

325 We filtered, trimmed and aligned *Ficedula* reads to the stonechat draft reference genome
326 so that we could directly compare the locations of outlier regions between genera, using the same
327 tools in GATK and Picard as for stonechat sequences (Supporting information). To calculate F_{ST} ,
328 we first generated a VCF file with UnifiedGenotyper from GATK and filtered raw variants with
329 the VariantFiltration tool (settings: $QD < 2.0 \parallel FS > 60.0 \parallel MQ < 40.0$). We then calculated F_{ST}
330 with VCFtools (Danecek et al. 2011) from the resulting SNPs across 50-kb non-overlapping
331 windows. We estimated d_{XY} from minor allele frequencies obtained in ANGSD (Korneliussen et
332 al. 2014), using a custom script to calculate 50 kb windowed averages. We only included sites
333 that had genotype calls for at least 5 out of 10 individuals per species and retained windows for
334 which at least 40% of bases satisfied this criterion.

335

336 **RESULTS**

337

338 **SNP-based phylogeny and inter-taxa divergence**

339 The Maximum Likelihood (ML) phylogeny built on fixed nuclear sites showed high support for
340 the placement of the Canary Islands stonechat as the sister taxon to the European stonechat
341 (Austria and Ireland) (Fig. 1 A). The clade comprising European, Canarian, and Kenyan
342 stonechats, to the exclusion of Siberian stonechats, was also strongly supported. This nuclear
343 phylogeny contradicted the existing mtDNA topology. We verified that this result was not an

344 artifact of sparse taxon sampling or choice of outgroup by constructing an ML tree with
345 cytochrome-*b* consensus sequences obtained from Austrian, Irish, Kenyan, and Siberian pools;
346 not enough mitochondrial sequence was recoverable for Canarian stonechats. Here, Kenyan
347 stonechats were placed as the sister lineage to the remaining stonechats, in agreement with past
348 mitochondrial studies (not shown).

349 We calculated mean genome-wide F_{ST} and d_{XY} to further examine relationships among
350 stonechat taxa. The first two principal coordinate axes calculated from a distance matrix of mean
351 pairwise F_{ST} (explaining a total of 87% of variance; 47% in first axis) revealed Siberian
352 stonechats to be approximately equidistant from the other taxa in terms of overall allele
353 frequency differentiation (Fig. 1 B). Stonechats from Austria and Ireland were extremely similar
354 (with only 7 fixed differences out of 10,164,331 sites with $F_{ST} > 0$, or 7×10^{-5} %), reflecting their
355 geographic proximity and common evolutionary history. In contrast, Kenyan and Canary Islands
356 stonechats were most different (1,251,605 fixed differences out of 12,401,462 sites with $F_{ST} > 0$,
357 or 10.09%). Overall, Canary Islands stonechats were strikingly dissimilar to even their closest
358 evolutionary relatives (Austria vs. Canary Islands: 782,967 fixed differences from 12,754,086
359 variable sites, or 6.14%). European stonechats were more similar genome-wide to Siberian and
360 Kenyan stonechats than to those from the Canary Islands, their sister lineage (Austria vs. Siberia:
361 244,623 fixed out of 15,168,199 variable, or 1.61%; Austria vs. Kenya: 640,425 fixed out of
362 12,032,148 variable, or 5.32%).

363 The principal coordinate analysis based on d_{XY} (68% of variance explained by first two
364 axes; 41% by the first) was similar to the one based on F_{ST} , except that Kenyan stonechats were
365 closer to European stonechats than to Siberian stonechats (Fig. 1 C). This is consistent with the

366 nuclear tree (Fig. 1 A). Again, Austria and Irish stonechats were nearly identical. Canary Islands
367 stonechats were distant from all stonechat taxa, but most similar to the European taxa.

368

369 **Shared regions of high differentiation show low genetic diversity, except in Canary Islands**
370 **stonechats**

371 Measures of divergence were strongly correlated among stonechats. D_{XY} showed strong
372 correlations across genomic windows (Fig. 2 A-B), and d_{XY} outlier regions were highly similar
373 (Fig. 2 D; Figs. S2 and S3, Supporting information); mean outlier similarity scores, averaged
374 across all comparisons, were 0.85 for low d_{XY} outliers and 0.79 for high d_{XY} outliers (Fig. S4,
375 Supporting information). F_{ST} was also significantly correlated in all comparisons, but the strength
376 of this correlation varied (Fig. 3 A-D). The association was greatest in comparisons including
377 Siberian stonechats (Fig. 3 A), but F_{ST} was also correlated among independent comparisons (i.e.
378 with no shared taxon) (Fig. 3 B-C). Overall F_{ST} outlier similarity was lower than d_{XY} for both
379 peaks and valleys (means of 0.31 and 0.24, respectively), indicating that approximately one-third
380 of F_{ST} peaks were shared (Fig. 3 E and Fig. S5, Supporting information). Across all comparisons,
381 windows with the lowest F_{ST} showed the most consistent associations. Of note, outlier regions
382 showed significant overlap in several comparisons where the four taxa being compared were all
383 different (Fig. 3 E), implicating common processes in independent stonechat lineages in the
384 generation of differentiation landscapes.

385 Generally, regions of high F_{ST} showed low genetic diversity, both within (π) and between
386 (d_{XY}) stonechat taxa. F_{ST} and d_{XY} were strongly negatively correlated, especially in comparisons
387 including Siberian stonechats (Fig. 4 A-B), and F_{ST} peaks overlapped strongly with d_{XY} valleys
388 (Fig. 4 A). F_{ST} valleys also overlapped with d_{XY} valleys in some comparisons. Regions of

389 reduced absolute divergence also showed reduced nucleotide diversity (Fig. S6, Supporting
390 information). Reductions in diversity occurred in the same genomic windows among stonechats,
391 even after standardizing for levels of between-population diversity (Figs. S7, S8 and S9,
392 Supporting information). Note that, due to the high similarity between Austrian and Irish
393 stonechats, we do not present comparisons of Irish stonechats with non-Austrian stonechats.

394 However, these patterns of genetic variation were often weaker, absent, or even reversed
395 for Canary Islands stonechats. Across the genome, F_{ST} and d_{XY} were *positively* correlated, despite
396 the lowest- F_{ST} windows showing high d_{XY} (Fig. 4 C). Canary Islands stonechats showed the
397 weakest associations in diversity correlations (Fig. S6 A,D and Fig. S7 B,D, Supporting
398 information). Standardized nucleotide diversity (π/d_{XY}) was negatively correlated between
399 Canary Islands and Siberian stonechats (Fig. S7 D, Supporting information), indicating that the
400 regions of the Siberian stonechat genome that showed the greatest diversity reductions were, in
401 fact, relatively more diverse in Canary Islands stonechats than the rest of the genome (Fig. S9,
402 Supporting information). No π/d_{XY} valleys regions were shared between Canarian stonechats and
403 other stonechat taxa (Fig. S7 E, Supporting information).

404

405 **Evidence of selection and effects of demography**

406 Among stonechats, genomic regions of high differentiation (F_{ST}), low absolute divergence (d_{XY}),
407 and low genetic diversity (π) coincided with significant decreases in Tajima's D and Fay and
408 Wu's H (Fig. 5 and Fig. 6). Fay and Wu's H showed strong associations with F_{ST} only in
409 comparisons including Siberian stonechats. Fay and Wu's H outlier regions were relatively
410 infrequent but coincided with low Tajima's D and π when they occurred, except in Canary Island
411 stonechats (Fig. S10, Supporting information). Tajima's D outlier regions were generally shared

412 across stonechats (Fig. S11, Supporting information), Some distinct low- H outlier regions
413 occurred in only one taxon (e.g., chromosomes 4A and 6) (Fig. S12, Supporting information).
414 Overall, in addition to lacking genetic diversity, outlier regions contained more low frequency
415 alleles than the rest of the genome, which is highly suggestive of a role of positive and/or
416 background selection in shaping differentiation patterns.

417 The genomic baseline value of Tajima's D can be biased downward by demographic
418 effects, particularly a population expansion. All stonechat taxa had median Tajima's D between -
419 0.5 and -1.1, with the exception of Canary Islands stonechats, at -2.8 (Fig. S13, Supporting
420 information). Negative values suggest that all five stonechat taxa have experienced past
421 demographic expansion events, with the signal especially strong in the insular Canary Islands
422 stonechats.

423

424 **Correspondence of genomic landscapes between *Saxicola* and *Ficedula***

425 Genome-wide patterns of genetic diversity and differentiation were also correlated between
426 stonechats and flycatchers. Absolute divergence was correlated between the two genera ($\rho =$
427 0.37-0.49, Fig. 2 C-D); d_{XY} outlier similarity was 0.48-0.52 between stonechats and flycatchers.
428 Flycatchers and stonechats shared a significant number of F_{ST} peaks and valleys, but only for a
429 subset of stonechat comparisons (Fig. 3 E). Genome-wide correlations of F_{ST} were significant but
430 weak (Fig. 3 F-G), and the strongest correlations occurred with Siberian stonechats. Finally,
431 within-population genetic diversity (π) was strongly correlated between stonechat and flycatcher
432 populations; some stonechat-flycatcher correlations were as strong or stronger than stonechat-
433 stonechat correlations (Fig. S7 E-G, Supporting information). Overall, these results suggest that
434 common processes are working independently and in parallel to influence genetic variation in

435 similar regions of the genome in both genera, although the association between genera is weaker
436 than within *Saxicola*.

437

438 **DISCUSSION**

439

440 We examined patterns of genetic diversity and differentiation in an avian radiation and identified
441 regions shared among stonechat taxa that were characterized by low within-population diversity,
442 low absolute inter-taxon divergence, and high (or, in some cases, low) differentiation. These
443 patterns are consistent with signatures of natural selection. We found that many stonechat outlier
444 regions also appeared in the closely related genus *Ficedula*. In this genus, genomic regions of
445 low genetic diversity and high differentiation are associated with infrequent recombination (Burri
446 et al. 2015), which suggests that one possible explanation for the parallel patterns of
447 differentiation in these genera is conserved (or convergently evolving) variation in recombination
448 rate (see Singhal et al. 2015). Overall, our results are consistent with linked selection (positive
449 selective sweeps and/or background selection) shaping large-scale patterns of genomic variation
450 in Muscicapid birds. The presence of Fay and Wu's H valleys in differentiation outlier regions
451 supports a role of positive selection in at least some cases. Despite a strong signal of similarity in
452 genomic landscapes, we also found evidence for substantial lineage-specific evolution: Siberian
453 stonechats appear to have experienced the strongest effects of selection, while drift may have
454 shaped Canary Islands stonechats' genomes.

455

456 **Discordance in nuclear and mitochondrial phylogenies**

457 The phylogenetic tree constructed with SNPs from across the nuclear genome (Fig. 1 A) was
458 highly supported at all nodes, yet it is not fully concordant with previous trees constructed from
459 mitochondrial DNA sequences (Illera et al. 2008; Woog et al. 2008; Zink et al. 2009). These
460 placed Kenyan stonechats (instead of Siberian, as in our reconstruction) as the sister lineage to
461 the remaining stonechats. Branch support for a sister relationship of Siberian stonechats and the
462 European/Canary Islands clade varied by study and tree-building algorithm. Mito-nuclear
463 discordance could be a sign of past admixture, sex biased gene flow, or other biological
464 phenomena (see Toews and Brelsford 2012). This well-resolved nuclear phylogeny serves as a
465 basis for testing broader questions about genome-scale differentiation in this complex: For
466 example, it helps explain why mean d_{XY} between European and Kenyan stonechats is relatively
467 low compared to Siberian stonechats (Fig. 1 C). Finally, although sparse taxon sampling (5 taxa)
468 and choice of outgroup could potentially introduce biases (e.g., Stervander et al. 2015), our
469 cytochrome *b*-only tree (not shown) was consistent with previous mitochondrial studies, which
470 achieved near-complete taxon sampling (e.g., Illera et al. 2008). The topology differences we find
471 between mitochondrial and nuclear-based phylogenies are therefore unlikely to be artifacts of
472 sampling. We note, however, that high bootstrap support does not always indicate a correct
473 species tree (e.g., Suh 2016); further investigation into the larger *Saxicola* clade (e.g., using gene
474 tree-based methods and demographic modeling; Nater et al. 2015) will be required to obtain a
475 better understanding of their phylogenetic affinities.

476

477 **Congruent genomic landscapes across a speciation continuum**

478 Patterns of within- and between-population genetic diversity in stonechats show high levels of
479 parallelism across multiple scales of evolutionary divergence. We found outliers in comparisons

480 of highly similar taxa in the same regions as comparisons at deeper levels of divergence. The
481 parallel reductions in d_{XY} are highly suggestive of selection before divergence (Cruickshank and
482 Hahn 2014), and analogous patterns in standardized nucleotide diversity (π/d_{XY}) indicate that
483 common selective forces have continued to reduce diversity on the branches leading to present-
484 day taxa (see Irwin et al. 2016).

485 Reductions in Fay and Wu's H in some outlier regions suggest that positive selection has
486 played a role in driving some of these regions of low genetic diversity and high differentiation.
487 Some H outliers are present in multiple taxa, while others occur in only one (as in *Ficedula*, Burri
488 et al. 2015), suggesting that localized selective sweeps may not have occurred in all groups, or
489 that sweeps occurred too far in the past for detection using this method.

490 Pairwise comparisons that include Siberian stonechats show the most conspicuous F_{ST}
491 peaks, which coincide with regions of low within-population diversity (π). Together, strongly
492 reduced within-population genetic diversity in specific genomic regions and corresponding peaks
493 of differentiation are consistent with Siberian stonechats experiencing the strongest effects of
494 selection in outlier regions. Most of the larger outlier regions also showed significant decreases in
495 Fay and Wu's H , suggesting that positive selective sweeps have contributed to this pattern. As
496 temperate zone breeders and obligate long-distance migrants, Siberian stonechats are expected to
497 generally show a faster pace of life, larger clutch sizes, and higher metabolic rates, along with a
498 range of specializations associated with a strongly migratory lifestyle (Wikelski et al. 2003;
499 Tieleman et al. 2009; Baldwin et al. 2010; Robinson et al. 2010). It is possible that a combination
500 of these factors has led to a strong footprint of selection on the Siberian stonechat genome.

501 Kenyan and Canarian stonechats showed the highest genome-wide F_{ST} . Notably,
502 however, we also observed conspicuous F_{ST} valleys in the same locations as the F_{ST} peaks of

503 other pairwise taxon comparisons (Fig. 5 and Fig. S5, Supporting information). In other words,
504 these taxa are differentiated across the vast majority of the genome, but they show low
505 differentiation (F_{ST}) in regions of low absolute divergence (d_{XY}). This pattern is not unique to this
506 comparison: Austrian and Irish stonechats, and occasionally others, show valleys in similar areas.
507 F_{ST} valleys may occur where d_{XY} (between-group variation) is reduced but π (within-group
508 variation) remains high, especially in Canary Islands stonechats, but more work is needed to
509 understand this phenomenon.

510

511 **Correlation of genomic variation between genera**

512 Genomic landscapes of genetic diversity and differentiation in stonechats are significantly
513 correlated with those in Pied and Collared Flycatchers. These results contrast with recent findings
514 in other passerine birds, for example greenish warblers (Irwin et al. 2016). Nucleotide diversity in
515 greenish warblers is only weakly correlated with that in outgroup comparisons (π : Pearson's $r =$
516 0.19). We found a stronger association in nucleotide diversity between stonechats and flycatchers
517 (π : Spearman's $\rho = 0.47-0.60$, excluding Canary Is.). *Saxicola* and *Ficedula* share certain aspects
518 of their life history (e.g., they are insectivores, and the flycatchers and most stonechats are
519 migratory), but the hypothesis that these parallel signatures of selection and differentiation are
520 due to shared ecological selection pressures on the same loci appears unlikely. Burri et al. (2015)
521 demonstrated a clear link between low recombination and areas of high differentiation in
522 *Ficedula*, which suggests that low recombination might also contribute to shared differentiation
523 outliers within *Saxicola*. Although initial evidence suggested that avian recombination landscapes
524 change drastically over time (Backström et al. 2010), recent work has shown that recombination
525 landscapes can be conserved in birds across millions of years of evolution (Singhal et al. 2015). It

526 is therefore possible that coincident areas of low recombination, in combination with linked
527 selection, may play a role in shaping the broad patterns of landscapes of genomic variation and
528 differentiation across both closely related and deeply diverged taxa. However, direct measures of
529 recombination rates in stonechats are needed to test this hypothesis. While recombination is
530 reduced in close proximity to avian centromeres (Backström et al. 2010), centromeres do not
531 explain the recombination deserts in the centers of acrocentric chromosomes (e.g. 4A, 9, 10, 11,
532 12, 13, and 18; Knief and Forstmeier 2015) (Kawakami et al. 2014; Burri et al. 2015). These
533 regions frequently show high differentiation among flycatchers (Burri et al. 2015) and stonechats.

534 Decreases in Fay and Wu's H in a subset of outlier regions and a subset of taxa suggest
535 that positive selection has also contributed to this convergent genomic evolution. Indeed, Irwin et
536 al. (2016) favor positive selection as the likely driver of differentiation landscapes in greenish
537 warblers, citing exceedingly low nucleotide diversity in differentiation peaks; in one comparison
538 in that study, regions with $F_{ST} > 0.9$ showed just 6.7% the nucleotide diversity of regions with F_{ST}
539 < 0.6 . We found diversity reductions in stonechats and flycatchers to be less severe: between
540 Austrian and Siberian stonechats, which show the greatest reduction in nucleotide diversity in F_{ST}
541 peaks, π in regions with F_{ST} above the 95% percentile was reduced to 30-34% of that of regions
542 with F_{ST} below the 50th percentile. In *Ficedula* flycatchers, this statistic was 43-50%. Therefore,
543 we consider background selection, in concert with reduced recombination, to be an additional
544 plausible driver of correlation in genomic landscapes.

545 Conserved variation in mutation rate is another possible driver of this correlation. Irwin et
546 al. (2016) found weak correlations in d_{XY} between greenish warblers and more distant
547 comparisons (Pearson's $r = 0.07-0.14$), which does not support this explanation. In contrast, we
548 found reasonably strong correlations in d_{XY} between stonechat and flycatcher genera (Spearman's

549 $\rho = 0.37-0.49$). Therefore, we cannot rule out a further contribution of conserved variation in
550 mutation rate to these patterns.

551

552 **Genomics and demography of Canary Islands stonechats**

553 Canary Islands stonechats' genomic landscapes differed from those of the other stonechats. The
554 valleys of standardized nucleotide diversity (π/d_{XY}) seen in other taxa were completely absent,
555 suggesting that selection has not reduced diversity across the genome in a heterogeneous way. In
556 fact, this ratio was elevated in the same regions in which it was reduced in the other taxa.
557 Tajima's D was highly negative genome-wide and showed lower variance than in other
558 stonechats. Combined, these results are most consistent with a demographic history that included
559 a severe population bottleneck (erasing existing patterns of variation), followed by a substantial
560 population expansion. Previous research has found evidence of founder effects and/or bottlenecks
561 in Canary Island birds (Barrientos et al. 2009; Barrientos et al. 2014; Spurgin et al. 2014). The
562 evidence for a bottleneck and expansion and the marked homogeneity of genetic diversity across
563 the genome in Canary Islands stonechats suggest that genetic drift has played a dominant role in
564 its divergence from other stonechats, possibly overpowering selection (see Hansson et al. 2014;
565 Spurgin et al. 2014; Gonzalez-Quevedo et al. 2015; Illera et al. 2016). The unusual pattern seen
566 in standardized nucleotide diversity may be explained by this prevalence of drift over selection.
567 Because selection has not reduced π in the outlier regions shared by other stonechats, this statistic
568 shows little variation across the genome of Canary Islands stonechats. This unusual pattern
569 therefore results from the lack of a reduction in within-population diversity (π) in areas where
570 between-population diversity (d_{XY}) is still reduced, presumably due to selection in the ancestral
571 stonechat.

572

573 Evidence of lineage-specific evolution

574 Despite striking similarities in the genomic landscapes of stonechats, we also find lineage-
575 specific evolution. At the broadest levels of our analysis, in which we compare genera, we
576 identified conspicuous differentiation peaks that appear in *Ficedula* but not *Saxicola* (e.g., on
577 chromosomes 3, 8, 10, 11, 12, 13, and 18; Fig. S14, Supporting information, shows chromosome
578 13), and vice versa (e.g., on chromosomes 6, 7, 17, and 20; Fig. S15, Supporting information
579 shows chromosome 20). These outlier regions should be further examined from a functional
580 perspective, as they appear to have resulted from evolutionary processes specific to a particular
581 lineage. The most conspicuous outlier regions shared between these systems should likewise be
582 examined (e.g., on chromosomes 1, 1A, 2, 3, 4, and 4A; Figs. S16 and S17, Supporting
583 information, show chromosomes 1A and 4A).

584

585 Conclusion

586 Few former studies (Burri et al. 2015; Lamichhaney et al. 2015; Irwin et al. 2016; Vijay et al.
587 2016) have examined genome-wide patterns of differentiation in more than two avian taxa, yet
588 comparative studies of closely related species have great potential to shed light on genome
589 evolution (Cutter and Payseur 2013). We find parallel patterns of selection in the stonechat
590 complex—likely occurring both before and after speciation—and evidence of demography
591 potentially overwhelming signatures of selection in one species. In addition, this study suggests
592 that parallel genomic processes are operating in independent evolutionary systems to drive the
593 differentiation of similar genomic regions across genera. We hypothesize that linked selection
594 coupled with areas of low recombination, which may be conserved across these taxa, have shaped

595 these broad patterns. Whether concordant outlier regions actually contribute to reproductive
596 isolation or are otherwise consequential in the speciation process is unknown. Therefore, we
597 recommend that outlier markers obtained through genome scans and their relevance to speciation
598 be interpreted with caution. Importantly, our comparative method also identified differentiation
599 outlier regions that are not widely shared; these may harbor loci important in lineage-specific
600 evolution and should be examined closely. As genomic comparisons among radiations
601 accumulate, we will be able to compare the congruence in genomic landscapes and potentially
602 reveal the phenomena that drive genomic differentiation over evolutionary time.

603

604 **ACKNOWLEDGEMENTS**

605

606 We thank the Max Planck Society, the Fuller Evolutionary Biology program at the Cornell Lab of
607 Ornithology, the Hunter R. Rawlings Presidential Research Scholars Program (Cornell), and the
608 Dextra Undergraduate Research Endowment Fund (Cornell) for funding. We thank the
609 University of Washington Burke Museum, Heiner Flinks and Bea Apfelbeck for providing
610 samples, and the Canarian Government and the Cabildo of Fuerteventura for permission to
611 sample Canary Islands stonechats. We thank Heinke Buhtz, Conny Burghardt, Sven Künzel,
612 Nicole Thomsen, Mayra Zamora Moreno, and Bronwyn Butcher for guidance in the lab; David
613 Toews, Scott Taylor, Jacob Berv, and Kira Delmore for valuable discussion; Kelly Zamudio,
614 Jennifer Walsh, Kristen Ruegg, Sonya Clegg, Eric Anderson, and four anonymous reviewers for
615 thoughtful feedback on this manuscript; and Susan and Daniel Van Doren for valued support.

616

REFERENCES

- 617
618
619 Andolfatto, P. (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in*
620 *Genetics & Development*, **11**, 635-641.
- 621 Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M. T. Webster,
622 T. Öst, M. Schneider, B. Kempnaers, and H. Ellegren (2010) The recombination
623 landscape of the zebra finch *Taeniopygia guttata* genome. *Genome research*, **20**, 485-495.
- 624 Backström, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer, A. Qvarnström, and H.
625 Ellegren (2008) A gene-based genetic linkage map of the Collared flycatcher (*Ficedula*
626 *albicollis*) reveals extensive synteny and gene-order conservation during 100 million
627 years of avian evolution. *Genetics*, **179**, 1479-1495.
- 628 Baldwin, M. W., H. Winkler, C. L. Organ, and B. Helm (2010) Wing pointedness associated with
629 migratory distance in common-garden and comparative studies of stonechats (*Saxicola*
630 *torquata*). *Journal of Evolutionary Biology*, **23**, 1050-1063.
- 631 Barrientos, R., L. Kvist, A. Barbosa, F. Valera, F. Khoury, S. Varela, and E. Moreno (2014)
632 Refugia, colonization and diversification of an arid-adapted bird: coincident patterns
633 between genetic data and ecological niche modelling. *Molecular Ecology*, **23**, 390-407.
- 634 Barrientos, R., L. Kvist, A. Barbosa, F. Valera, G. M. López-Iborra, and E. Moreno (2009)
635 Colonization patterns and genetic structure of peripheral populations of the trumpeter
636 finch (*Bucanetes githagineus*) from Northwest Africa, the Canary Islands and the Iberian
637 Peninsula. *Journal of Biogeography*, **36**, 210-219.
- 638 Begun, D. J., and C. F. Aquadro (1993) African and North American populations of *Drosophila*
639 *melanogaster* are very different at the DNA levels. *Nature*, **365**, 548-550.
- 640 Benjamini, Y., and Y. Hochberg (1995) Controlling the false discovery rate: a practical and
641 powerful approach to multiple testing. *Journal of the royal statistical society. Series B*
642 *(Methodological)*, 289-300.
- 643 Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S.
644 Bureš, L. Z. Garamszegi, S. Hogner, et al. (2015) Linked selection and recombination rate
645 variation drive the evolution of the genomic landscape of differentiation across the
646 speciation continuum of *Ficedula* flycatchers. *Genome research*, **25**, 1656-1665.
- 647 Charlesworth, B. (1998) Measures of divergence between populations and the effect of forces
648 that reduce variability. *Molecular Biology and Evolution*, **15**, 538-543.
- 649 Charlesworth, B. (2001) The effect of life-history and mode of inheritance on neutral genetic
650 variability. *Genetics Research*, **77**, 153-166.
- 651 Charlesworth, B. (2013) Background Selection 20 Years on: The Wilhelmine E. Key 2012
652 Invitational Lecture. *Journal of Heredity*, **104**, 161-171.
- 653 Charlesworth, B., M. T. Morgan, and D. Charlesworth (1993) The effect of deleterious mutations
654 on neutral molecular variation. *Genetics*, **134**, 1289-1303.
- 655 Charlesworth, B., M. Nordborg, and D. Charlesworth (1997) The effects of local selection,
656 balanced polymorphism and background selection on equilibrium patterns of genetic
657 diversity in subdivided populations. *Genetics Research*, **70**, 155-174.
- 658 Collar, N. (2016a) Common Stonechat (*Saxicola torquatus*) in Handbook of the Birds of the
659 World Alive (del Hoyo, J., A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Eds.).
660 Lynx Edicions, Barcelona.

- 661 Collar, N. (2016b) Fuerteventura Stonechat (*Saxicola dacotiae*) in Handbook of the Birds of the
662 World Alive (del Hoyo, J., A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Eds.).
663 Lynx Edicions, Barcelona.
- 664 Cruickshank, T. E., and M. W. Hahn (2014) Reanalysis suggests that genomic islands of
665 speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**,
666 3133-3157.
- 667 Cutter, A. D., and B. A. Payseur (2013) Genomic signatures of selection at linked sites: unifying
668 the disparity among species. *Nat Rev Genet*, **14**, 262-274.
- 669 Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G.
670 Lunter, G. T. Marth, S. T. Sherry, G. McVean, et al. (2011) The variant call format and
671 VCFtools. *Bioinformatics*, **27**, 2156-2158.
- 672 Delmore, K. E., S. Hübner, N. C. Kane, R. Schuster, R. L. Andrew, F. Câmara, R. Guigó, and D.
673 E. Irwin (2015) Genomic analysis of a migratory divide reveals candidate genes for
674 migration and implicates selective sweeps in generating islands of differentiation.
675 *Molecular Ecology*, **24**, 1873-1888.
- 676 Ellegren, H. (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology &*
677 *Evolution*, **25**, 283-291.
- 678 Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H.
679 Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf
680 (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*,
681 **491**, 756-760.
- 682 Fay, J. C., and C. I. Wu (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**,
683 1405-1413.
- 684 Feder, J. L., S. P. Egan, and P. Nosil (2012) The genomics of speciation-with-gene-flow. *Trends*
685 *in Genetics*, **28**, 342-350.
- 686 Ferretti, L., S. E. Ramos-Onsins, and M. Pérez-Enciso (2013) Population genomics from pool
687 sequencing. *Molecular Ecology*, **22**, 5561-5576.
- 688 Gautier, M., J. Foucaud, K. Gharbi, T. Cézard, M. Galan, A. Loiseau, M. Thomson, P. Pudlo, C.
689 Kerdelhué, and A. Estoup (2013) Estimation of population allele frequencies from next-
690 generation sequencing data: pool-versus individual-based genotyping. *Molecular*
691 *Ecology*, **22**, 3766-3779.
- 692 Gill, F., and D. Donsker, (Eds.) (2016) IOC World Bird List (v 6.1).
- 693 Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G.
694 Hall, T. P. Shea, S. Sykes, A. M. Berlin, et al. (2011) High-quality draft assemblies of
695 mammalian genomes from massively parallel sequence data. *Proceedings of the National*
696 *Academy of Sciences*, **108**, 1513-1518.
- 697 Gonzalez-Quevedo, C., L. G. Spurgin, J. C. Illera, and D. S. Richardson (2015) Drift, not
698 selection, shapes toll-like receptor variation among oceanic island populations. *Molecular*
699 *Ecology*, **24**, 5852-5863.
- 700 Grabherr, M. G., P. Russell, M. Meyer, E. Mauceli, J. Alföldi, F. Di Palma, and K. Lindblad-Toh
701 (2010) Genome-wide synteny through highly sensitive sequence alignment: Satsuma.
702 *Bioinformatics*, **26**, 1145-1151.
- 703 Gwinner, E., V. Neusser, E. Engl, and D. Schmidl (1987) Haltung, Zucht und Eiaufzucht
704 afrikanischer und europäischer Schwarzkehlchen *Saxicola torquata*. *Gefiederte Welt*, **111**,
705 118-120, 145-147.

- 706 Hansson, B., M. Ljungqvist, J.-C. Illera, and L. Kvist (2014) Pronounced Fixation, Strong
707 Population Differentiation and Complex Population History in the Canary Islands Blue
708 Tit Subspecies Complex. *PLoS ONE*, **9**, e90186.
- 709 Hedrick, P. W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633-1638.
- 710 Helm, B. (2003) Seasonal timing in different environments: comparative studies in Stonechats.
711 LMU Munich.
- 712 Helm, B., W. Fiedler, and J. Callion (2006) Movements of European Stonechats *Saxicola*
713 *torquata* according to ringing recoveries. *Ardea*, **94**, 33-44.
- 714 Helm, B., I. Schwabl, and E. Gwinner (2009) Circannual basis of geographically distinct bird
715 schedules. *Journal of Experimental Biology*, **212**, 1259-1269.
- 716 Huang, S., Z. Chen, G. Huang, T. Yu, P. Yang, J. Li, Y. Fu, S. Yuan, S. Chen, and A. Xu (2012)
717 HaploMerger: reconstructing allelic relationships for polymorphic diploid genome
718 assemblies. *Genome research*, **22**, 1581-1588.
- 719 Illera, J. C., D. S. Richardson, B. Helm, J. C. Atienza, and B. C. Emerson (2008) Phylogenetic
720 relationships, biogeography and speciation in the avian genus *Saxicola*. *Molecular*
721 *Phylogenetics and Evolution*, **48**, 1145-1154.
- 722 Illera, J. C., L. G. Spurgin, E. Rodríguez-Expósito, M. Nogales, and J. C. Rando (2016) What are
723 we learning about speciation and extinction from the Canary Islands? *Ardeola*, **63**.
- 724 Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens (2016) Recurrent selection
725 explains parallel evolution of genomic regions of high relative but low absolute
726 differentiation in a ring species. *Molecular Ecology*, **25**, 4488–4507.
- 727 Kaplan, N. L., R. R. Hudson, and C. H. Langley (1989) The "hitchhiking effect" revisited.
728 *Genetics*, **123**, 887-899.
- 729 Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, P. Olason, and
730 H. Ellegren (2014) A high-density linkage map enables a second-generation collared
731 flycatcher genome assembly and reveals the patterns of avian recombination rate variation
732 and chromosomal evolution. *Molecular Ecology*, **23**, 4035-4058.
- 733 Keinan, A., and D. Reich (2010) Human Population Differentiation Is Strongly Correlated with
734 Local Recombination Rate. *PLoS Genetics*, **6**, e1000886.
- 735 Knief, U., and W. Forstmeier (2015) Mapping centromeres of microchromosomes in the zebra
736 finch (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma*, 1-12.
- 737 Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol,
738 and C. Schlötterer (2011a) PoPoolation: A Toolbox for Population Genetic Analysis of
739 Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE*, **6**, e15925.
- 740 Kofler, R., R. V. Pandey, and C. Schlötterer (2011b) PoPoolation2: identifying differentiation
741 between populations using sequencing of pooled DNA samples (Pool-Seq).
742 *Bioinformatics*, **27**, 3435-3436.
- 743 Korneliussen, T. S., A. Albrechtsen, and R. Nielsen (2014) ANGSD: Analysis of Next
744 Generation Sequencing Data. *BMC Bioinformatics*, **15**, 1-13.
- 745 Lamichhaney, S., J. Berglund, M. S. Almen, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M.
746 Promerova, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, et al. (2015) Evolution of
747 Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371-375.
- 748 Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
749 *arXiv preprint arXiv:1303.3997*.

- 750 McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D.
751 Altshuler, S. Gabriel, and M. Daly (2010) The Genome Analysis Toolkit: a MapReduce
752 framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**,
753 1297-1303.
- 754 Nachman, M. W., and B. A. Payseur (2012) Recombination rate variation and speciation:
755 theoretical predictions and empirical results from rabbits and mice. *Philosophical*
756 *Transactions of the Royal Society of London B: Biological Sciences*, **367**, 409-421.
- 757 Nater, A., R. Burri, T. Kawakami, L. Smeds, and H. Ellegren (2015) Resolving Evolutionary
758 Relationships in Closely Related Species with Whole-Genome Sequencing Data.
759 *Systematic Biology*, **64**, 1000-1017.
- 760 Nei, M., and W. H. Li (1979) Mathematical model for studying genetic variation in terms of
761 restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269-
762 5273.
- 763 Nielsen, R. (2005) Molecular Signatures of Natural Selection. *Annual Review of Genetics*, **39**,
764 197-218.
- 765 Noor, M. A. F., and S. M. Bennett (2009) Islands of speciation or mirages in the desert?
766 Examining the role of restricted recombination in maintaining species. *Heredity*, **103**,
767 439-444.
- 768 Nordborg, M., B. Charlesworth, and D. Charlesworth (1996) The effect of recombination on
769 background selection. *Genetics Research*, **67**, 159-174.
- 770 Nosil, P., D. J. Funk, and D. Ortiz-Barrientos (2009) Divergent selection and heterogeneous
771 genomic divergence. *Molecular Ecology*, **18**, 375-402.
- 772 Paradis, E., J. Claude, and K. Strimmer (2004) APE: Analyses of Phylogenetics and Evolution in
773 R language. *Bioinformatics*, **20**, 289-290.
- 774 Robinson, W. D., M. Hau, K. C. Klasing, M. Wikelski, J. D. Brawn, S. H. Austin, C. E. Tarwater,
775 and R. E. Ricklefs (2010) Diversification of Life Histories in New World Birds. *The Auk*,
776 **127**, 253-262.
- 777 Roesti, M., S. Gavrillets, A. P. Hendry, W. Salzburger, and D. Berner (2014) The genomic
778 signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, **23**,
779 3944-3956.
- 780 Ruegg, K., E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith (2014) A role for migration-
781 linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, **23**,
782 4757-4769.
- 783 Sætre, G.-P., and S. A. Sæther (2010) Ecology and genetics of speciation in *Ficedula* flycatchers.
784 *Molecular Ecology*, **19**, 1091-1106.
- 785 Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L.
786 Peichel, G.-P. Saetre, C. Bank, A. Brannstrom, A. Brelsford, et al. (2014) Genomics and
787 the origin of species. *Nat Rev Genet*, **15**, 176-192.
- 788 Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li,
789 B. Raney, C. N. Balakrishnan, S. C. Griffith, et al. (2015) Stable recombination hotspots
790 in birds. *Science*, **350**, 928-932.
- 791 Smeds, L., V. Warmuth, P. Bolivar, S. Uebbing, R. Burri, A. Suh, A. Nater, S. Bures, L. Z.
792 Garamszegi, S. Hogner, J. Moreno, et al. (2015) Evolutionary analysis of the female-
793 specific avian W chromosome. *Nat Commun*, **6**.

- 794 Smith, J. M., and J. Haigh (1974) The hitch-hiking effect of a favourable gene. *Genetical*
795 *research*, **23**, 23-35.
- 796 Spurgin, L. G., J. C. Illera, T. H. Jorgensen, D. A. Dawson, and D. S. Richardson (2014) Genetic
797 and phenotypic divergence in an island bird: isolation by distance, by colonization or by
798 adaptation? *Molecular Ecology*, **23**, 1028-1039.
- 799 Stamatakis, A. (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of
800 Large Phylogenies. *Bioinformatics*.
- 801 Stephan, W., L. Xing, D. A. Kirby, and J. M. Braverman (1998) A test of the background
802 selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proceedings*
803 *of the National Academy of Sciences*, **95**, 5649-5654.
- 804 Stervander, M., J. C. Illera, L. Kvist, P. Barbosa, N. P. Keehnen, P. Pruißcher, S. Bensch, and B.
805 Hansson (2015) Disentangling the complex evolutionary history of the Western Palearctic
806 blue tits (*Cyanistes* spp.) – phylogenomic analyses suggest radiation by multiple
807 colonization events and subsequent isolation. *Molecular Ecology*, **24**, 2477-2494.
- 808 Suh, A. (2016) The phylogenomic forest of bird trees contains a hard polytomy at the root of
809 Neoaves. *Zoologica Scripta*, **45**, 50-62.
- 810 Tieleman, B. I., M. A. Versteegh, A. Fries, B. Helm, N. J. Dingemanse, H. L. Gibbs, and J. B.
811 Williams (2009) Genetic modulation of energy metabolism in birds through
812 mitochondrial function. *Proceedings of the Royal Society of London B: Biological*
813 *Sciences*, **276**, 1685-1693.
- 814 Toews, D. P. L., and A. Brelsford (2012) The biogeography of mitochondrial and nuclear
815 discordance in animals. *Molecular Ecology*, **21**, 3907-3930.
- 816 Turner, T. L., and M. W. Hahn (2010) Genomic islands of speciation or genomic islands and
817 speciation? *Molecular Ecology*, **19**, 848-850.
- 818 Turner, T. L., M. W. Hahn, and S. V. Nuzhdin (2005) Genomic Islands of Speciation in
819 *Anopheles gambiae*. *PLoS Biol*, **3**, e285.
- 820 Urquhart, E. D. (2002) *Stonechats. A guide to the genus Saxicola*. Christopher Helm.
- 821 Vijay, N., C. M. Bossu, J. W. Poelstra, M. H. Weissensteiner, A. Suh, A. P. Kryukov, and J. B.
822 W. Wolf (2016) Evolution of heterogeneous genome differentiation across multiple
823 contact zones in a crow species complex. *Nature Communications*, **7**, 13195.
- 824 Wang, J., N. R. Street, D. G. Scofield, and P. K. Ingvarsson (2016) Variation in linked selection
825 and recombination drive genomic divergence during allopatric speciation of European and
826 American aspens. *Molecular Biology and Evolution*.
- 827 White, B. J., C. Cheng, F. Simard, C. Costantini, and N. J. Besansky (2010) Genetic association
828 of physically unlinked islands of genomic divergence in incipient species of *Anopheles*
829 *gambiae*. *Molecular Ecology*, **19**, 925-939.
- 830 Wikelski, M., L. Spinney, W. Schelsky, A. Scheuerlein, and E. Gwinner (2003) Slow pace of life
831 in tropical sedentary birds: a common-garden experiment on four stonechat populations
832 from different latitudes. *Proceedings of the Royal Society of London B: Biological*
833 *Sciences*, **270**, 2383-2388.
- 834 Woog, F., M. Wink, E. Rastegar-Pouyani, J. Gonzalez, and B. Helm (2008) Distinct taxonomic
835 position of the Madagascar stonechat (*Saxicola torquatus sibilla*) revealed by nucleotide
836 sequences of mitochondrial DNA. *Journal of Ornithology*, **149**, 423-430.
- 837 Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to
838 systems of mating. *Evolution*, **19**, 395-420.

839 Yeaman, S., S. Aeschbacher, and R. Bürger (2016) The evolution of genomic islands by
840 increased establishment probability of linked alleles. *Molecular Ecology*, **25**, 2542-2558.
841 Zink, R. M., A. Pavlova, S. Drovetski, M. Wink, and S. Rohwer (2009) Taxonomic status and
842 evolutionary history of the *Saxicola torquata* complex. *Molecular Phylogenetics and*
843 *Evolution*, **52**, 769-773.

844

845

846 **DATA ACCESSIBILITY**

847 The Siberian stonechat genome assembly is archived at the European Nucleotide Archive with
848 accession number PRJEB19453. Pooled sequencing data from the five stonechat taxa are
849 archived with accession number PRJEB19452. Both datasets will be made public upon final
850 acceptance.

851

852 **AUTHOR CONTRIBUTIONS**

853 BVD, ML, LC, IJL, and BH designed the research. BH, ML, and JCI supplied samples. BVD and
854 LC performed all analyses and wrote the paper with input from all authors.

855 **FIGURE CAPTIONS**

856

857 Figure 1. (A) Maximum likelihood phylogenetic tree constructed with RAxML from fixed sites
858 across the stonechat nuclear genome, with two *Ficedula* species used as outgroups. Branch labels
859 denote bootstrap support from 100 rapid bootstrap iterations. This topology places Siberian
860 stonechat (*S. maurus*) as the sister lineage to the remaining taxa, in contrast to previous trees
861 based on mitochondrial DNA. Canary Islands stonechats are most closely related to European
862 stonechats. Illustrations (males shown) are reproduced with permission from Handbook of Birds
863 of the World Alive (Collar 2016a, 2016b). (B, C) Biplot of two principle coordinate axes derived
864 from analyses of: (B) mean F_{ST} and (C) mean d_{XY} . Axes are labeled with percent of variance
865 explained.

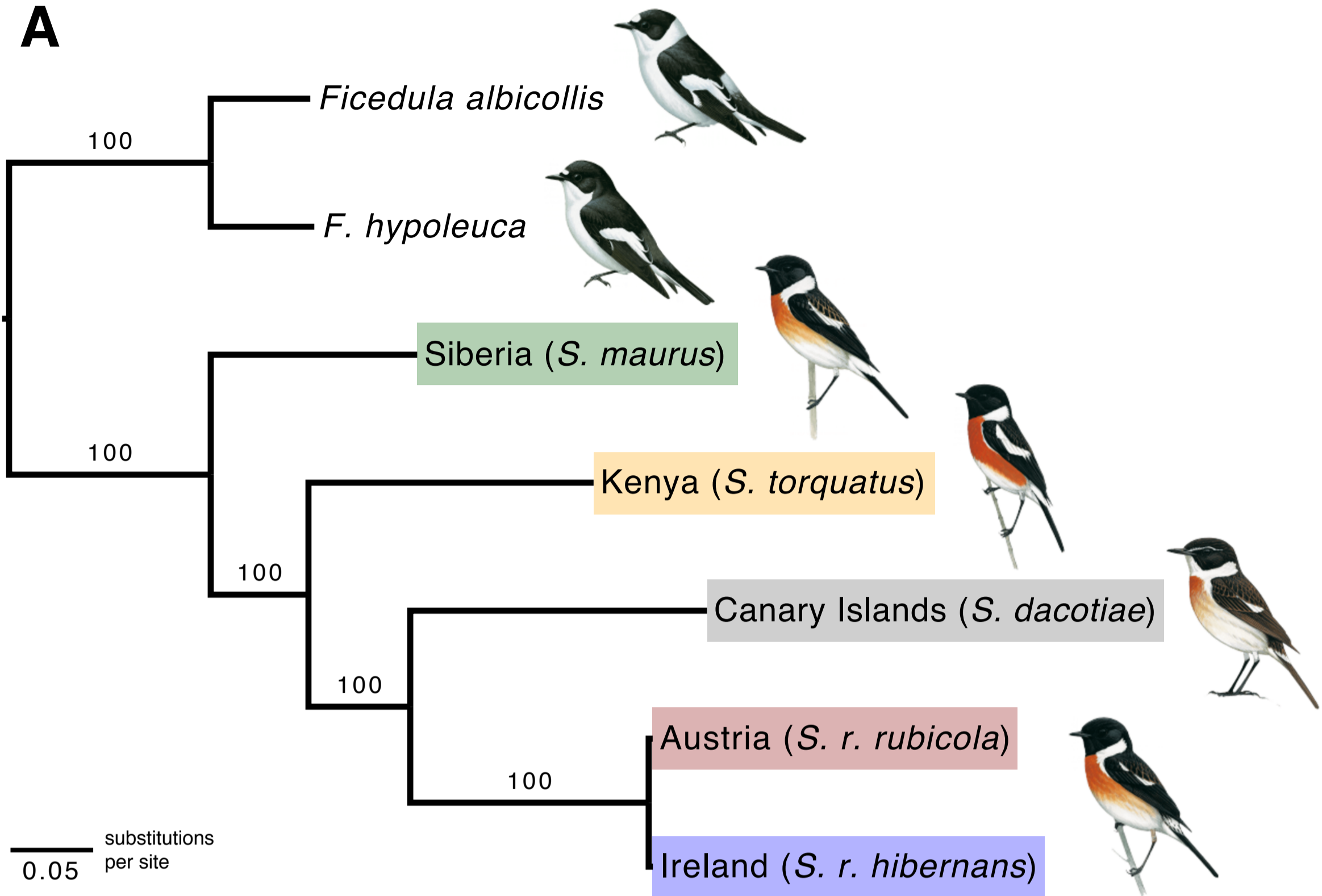
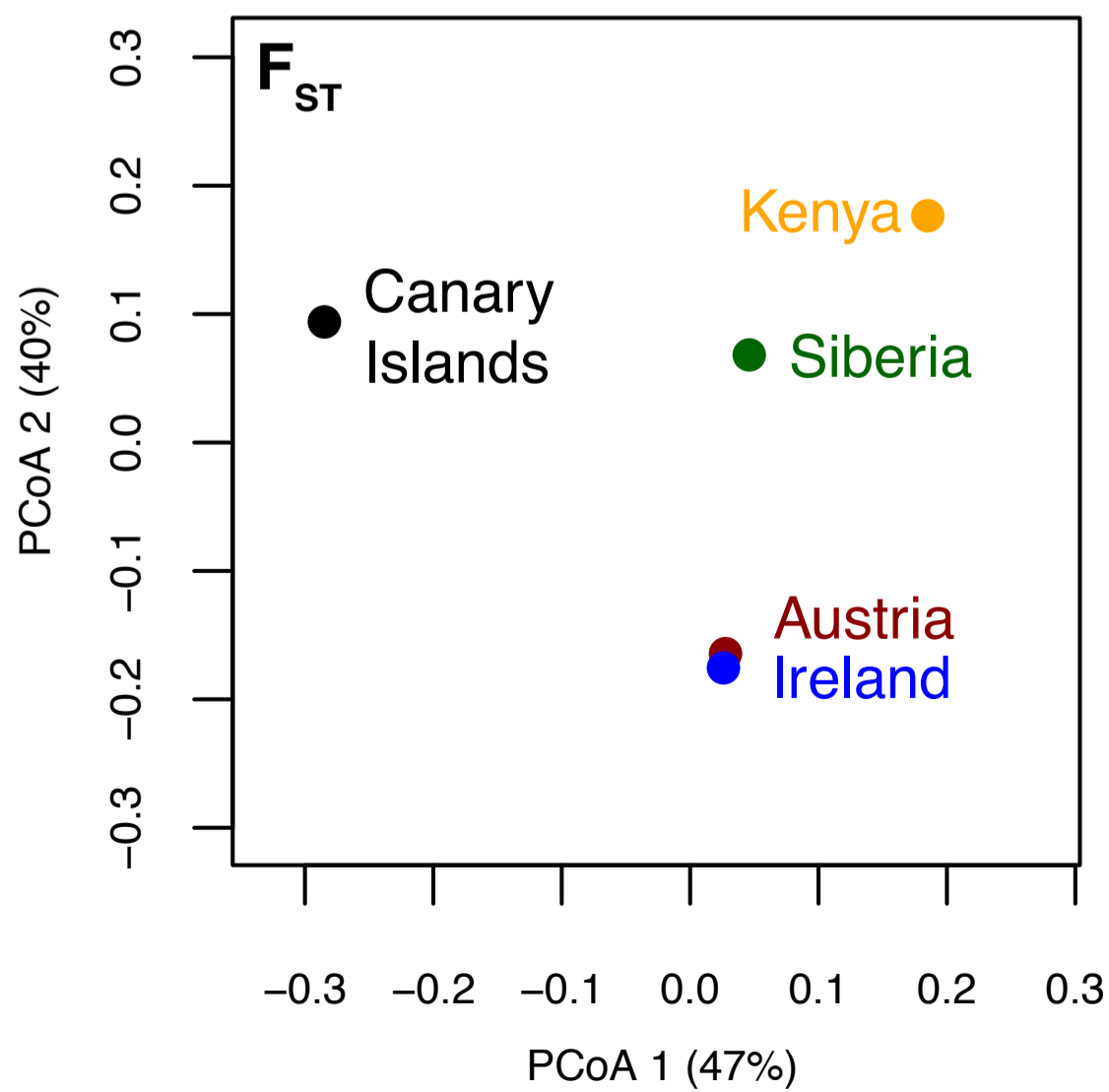
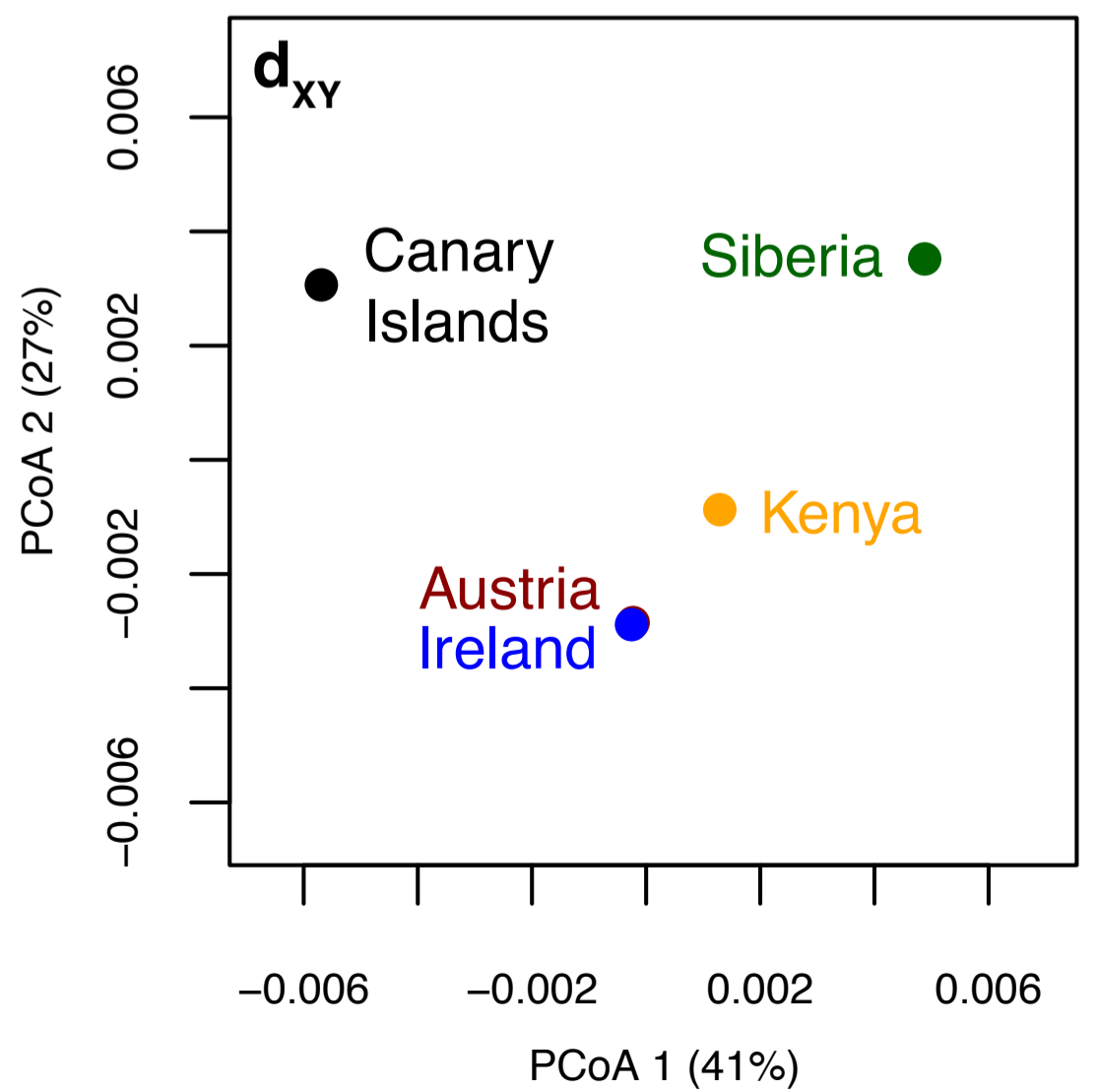
866 Figure 2. Correlation of d_{XY} among stonechats and flycatchers (*Ficedula hypoleuca* and
867 *albicollis*, or “Hyp.” and “Alb.”). (A,B,C) show scatterplots where each point represents one 50-
868 Kb genomic window. Orange lines are best-fit lines, and the Spearman’s rank correlation rho (ρ)
869 coefficient is given. (D) shows outlier similarity scores, which quantify the number of low- d_{XY}
870 “valleys” shared among different comparisons. Some comparisons including Irish stonechats are
871 not shown because of their similarity to Austrian stonechats. All tests were significant after
872 applying a false discovery rate correction. Cells with yellow backgrounds indicate that four
873 independent taxa are being compared. Letters in the upper right of cells show which cells
874 correspond to the scatterplots in sections A-C.
875

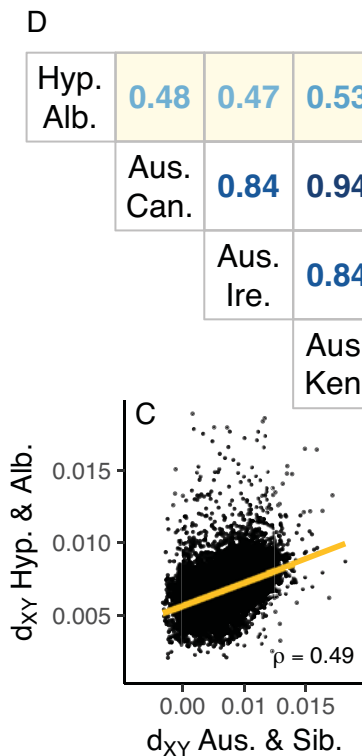
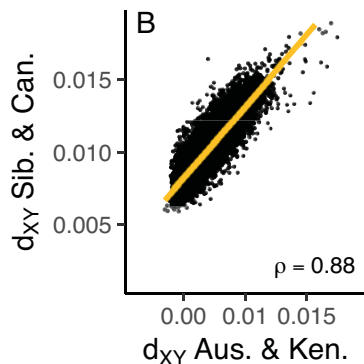
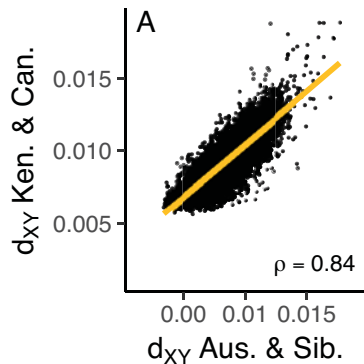
876 Figure 3. Correlation of F_{ST} among stonechats and flycatchers (*Ficedula hypoleuca* and
877 *albicollis*, or “Hyp.” and “Alb.”). (A,B,C,D,F,G) show scatterplots where each point represents
878 one 50-Kb genomic window. Orange lines are best-fit lines, and the Spearman’s rank correlation
879 rho (ρ) coefficient is also given. (E) shows outlier similarity scores, which quantify the number of
880 high- F_{ST} “peaks” shared among different comparisons (upper triangle of matrix) and the number
881 of low- F_{ST} “valleys” shared among different comparisons (lower triangle of matrix). Some
882 comparisons including Irish stonechats are not shown because of their similarity to Austrian
883 stonechats. Cells with an ‘X’ indicate tests that were not significant after applying a false
884 discovery rate correction. Cells with yellow backgrounds indicate that four independent taxa are
885 being compared. Letters in the upper right of cells show which cells correspond to the scatterplots
886 in the other sections.
887
888

889 Figure 4. Correlation of F_{ST} and d_{XY} among stonechats and flycatchers (*Ficedula hypoleuca* and
890 *albicollis*, or “Hyp.” and “Alb.”). (A) shows outlier similarity scores, which quantify the number
891 of low- d_{XY} “valleys” that coincide with either high- F_{ST} “peaks” (top row) or low- F_{ST} “valleys”
892 (bottom row). (B,C) show scatterplots where each point represents one 50-Kb genomic window.
893 Refer to Figs. 2-3 for details.
894
895

896 Figure 5. Genomic statistics calculated across stonechat and flycatcher chromosomes 1A and 4A.
897 Yellow and blue boxes indicated shared peaks and valleys, respectively. From top to bottom, the
898 statistics and box details are: F_{ST} among stonechats (peaks shared by 3 or more comparisons), F_{ST}
899 between flycatchers (peaks that also overlap with shared stonechats peaks), d_{XY} among
900 stonechats (valleys shared by 3 or more comparisons), d_{XY} among flycatchers (valleys that also
901 overlap with shared stonechats valleys), Tajima's D (valleys shared by 2 or more taxa),
902 nucleotide diversity (π) (valleys shared by 2 or more taxa), and Fay & Wu's H (valleys shared by
903 2 or more taxa).
904

905 Figure 6. Correlation of F_{ST} and d_{XY} with Tajima's D and Fay & Wu's H among stonechats and
906 flycatchers (*Ficedula hypoleuca* and *albicollis*, or "Hyp." and "Alb."). (A) shows outlier
907 similarity scores, which quantify the number of high- F_{ST} "peaks" that coincide with low Tajima's
908 D (top section) and low Fay & Wu's H (bottom section). Within each section, the top (No. 1) and
909 bottom (No. 2) rows show the results for each of the two taxa being compared. This is necessary
910 because Tajima's D and Fay & Wu's H are single-population statistics, while F_{ST} and d_{XY}
911 compare two populations. All comparisons were significant after applying a false discovery rate
912 correction. (B,C) show scatterplots where each point represents one 50-Kb genomic window.
913 Refer to Figs. 2-3 for details.

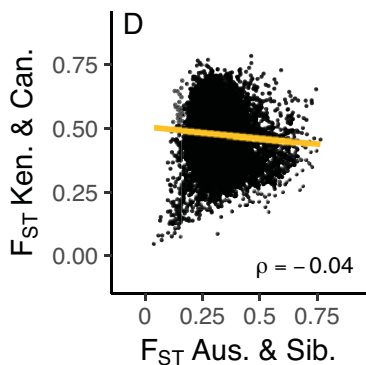
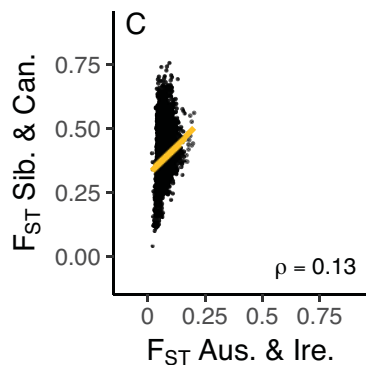
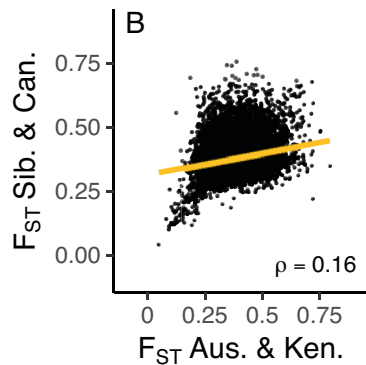
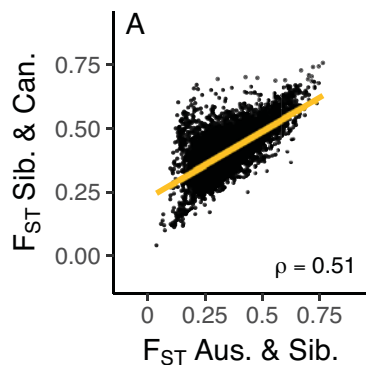
A**B****C**



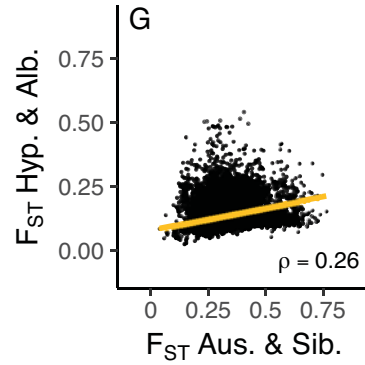
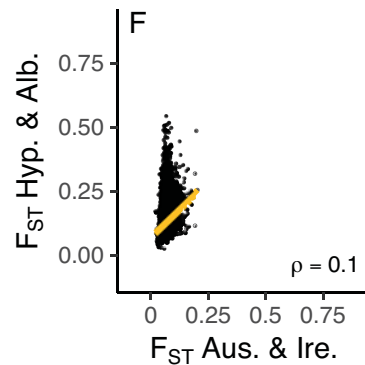
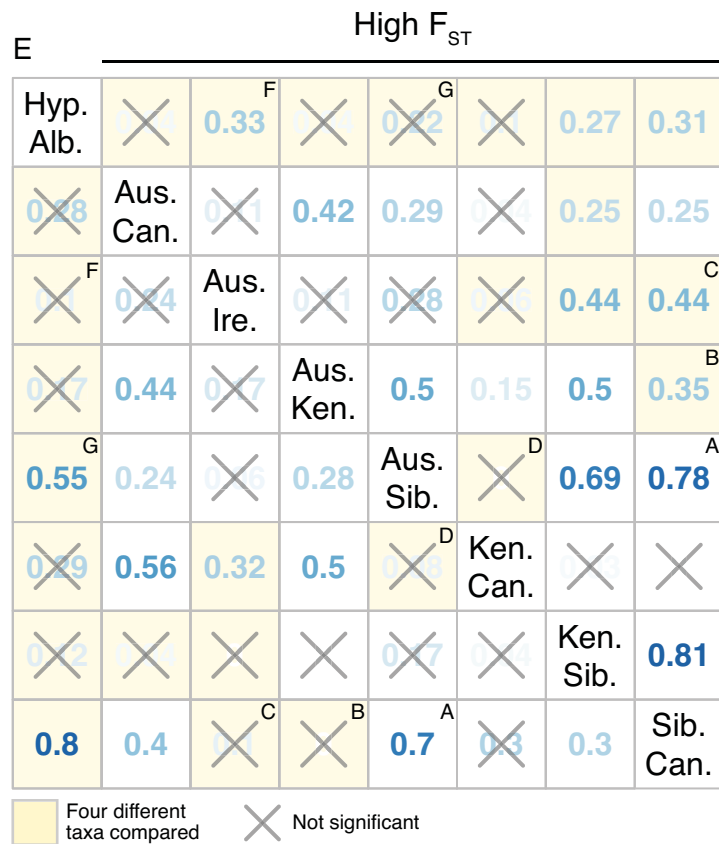
D Low d_{XY} Outlier Similarity

Hyp. Alb.	0.48	0.47	0.53	^C 0.53	0.52	0.55	0.53
Aus. Can.	0.84	0.94	0.86	0.86	0.86	0.84	0.86
Aus. Ire.	0.84	0.75	0.74	0.74	0.78	0.77	
Aus. Ken.	0.84	0.93	0.87	0.84	^B 0.84		
Aus. Sib.	^A 0.87	0.9	0.9				
Ken. Can.	0.87	0.87					
Ken. Sib.						0.95	
Sib. Can.							

Four different taxa compared

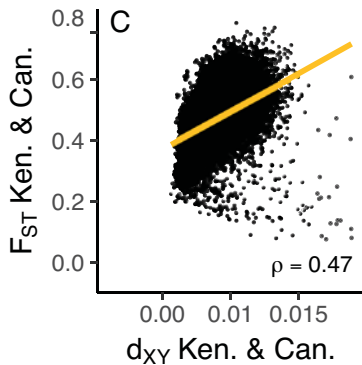
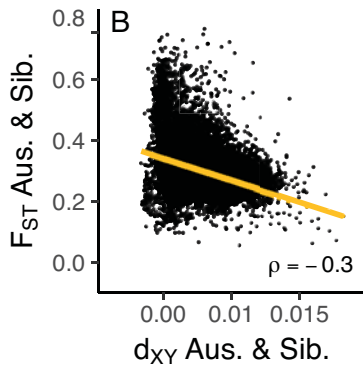


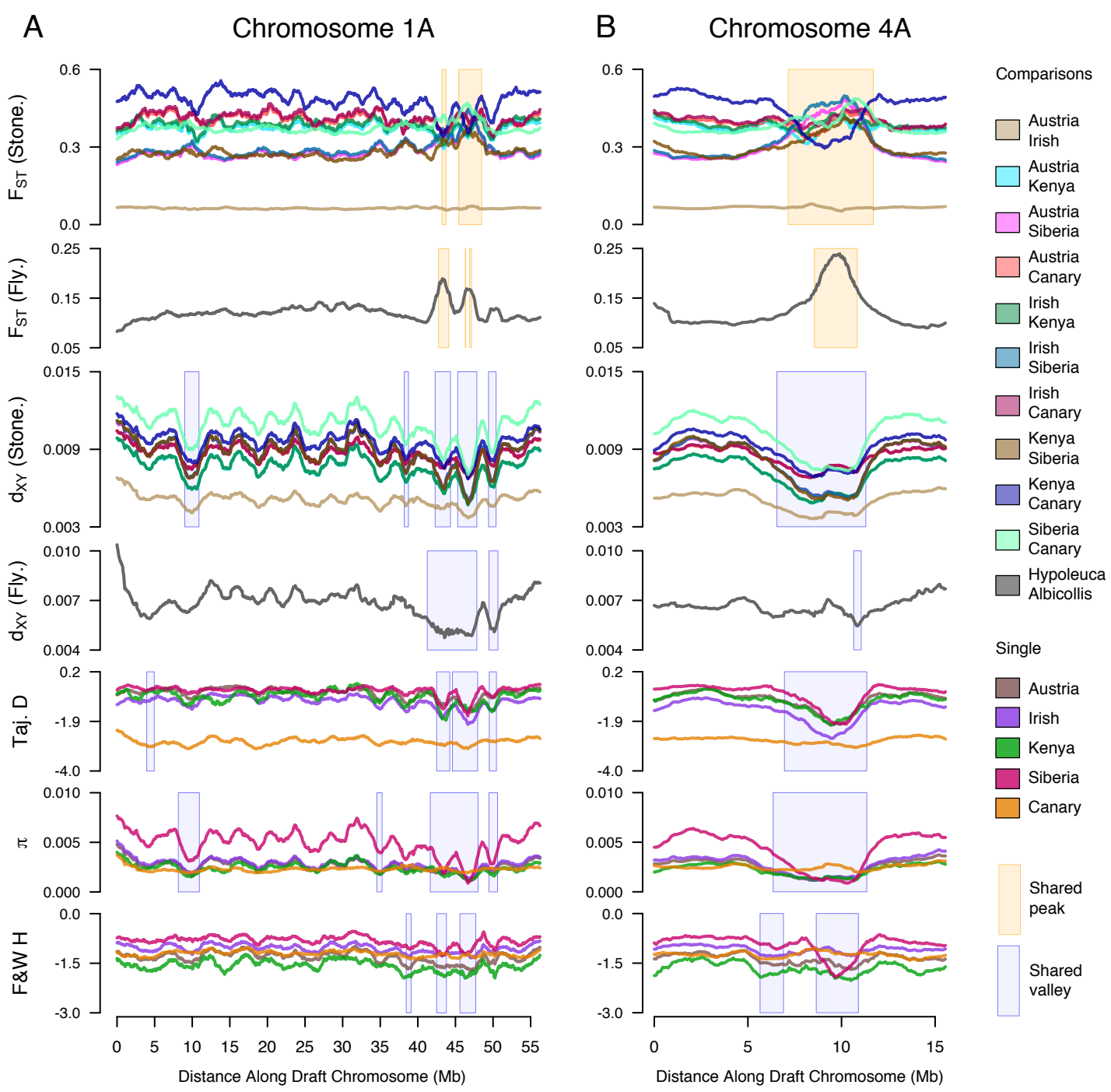
Low
 F_{ST}



A

	Hyp. Alb.	Aus. Can.	Aus. Ire.	Aus. Ken.	Aus. Sib.	Ken. Can.	Ken. Sib.	Sib. Can.
Low d_{XY} High F_{ST}	0.51	0.29	0.56	0.42	0.91 ^B	× ^C	0.92	0.93
Low d_{XY} Low F_{ST}	0.7	0.44	0.45	0.56	0.8 ^B	0.79 ^C	0.4	0.2
High d_{XY} High F_{ST}	0.4	×	0.2	×	× ^B	0.25 ^C	×	×



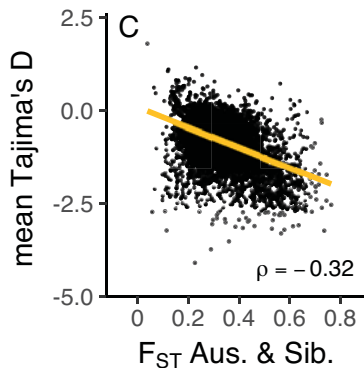
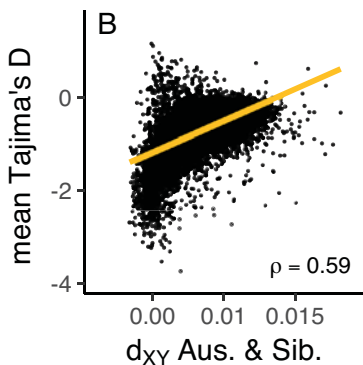


A

1:	Hyp.	Aus.	Aus.	Aus.	Aus.	Ken.	Ken.	Sib.
2:	Alb.	Can.	Ire.	Ken.	Sib.	Can.	Sib.	Can.

High F_{ST} Low D (1)	0.54	0.29	0.39	0.5	0.88 ^C	0.8	0.7	0.68
High F_{ST} Low D (2)	0.33	0.46	0.56	0.38	0.72 ^C	0.7	0.68	0.67

High F_{ST} Low F & W 's H (1)	0.38	0.26	0.2	0.53	0.5	0.48	0.45
High F_{ST} Low F & W 's H (2)	0.8	0.6	0.7	0.5	0.5	0.47	0.6



SUPPORTING INFORMATION

Correlated patterns of genetic diversity and differentiation across an avian family

Benjamin M. Van Doren, Leonardo Campagna, Barbara Helm, Juan Carlos Illera, Irby J. Lovette, and Miriam Liedvogel

ADDITIONAL METHODS

Draft Reference Genome

To extract DNA from Siberian stonechat (*S. maurus*) muscle tissue, we used the Gentra Puregene Tissue kit, following the protocol for fixed tissue. Gel electrophoresis revealed the DNA to be composed of highly intact molecules (all visible >10 kb). The ALLPATHS-LG algorithm (Gnerre et al. 2011) used 90.1% of the fragment library (total 506,475,396 reads), covering the genome at a mean depth of 46.9x. Combined, the two mate-pair libraries comprised 923,232,904 reads; ALLPATHS-LG used 20.7% of these reads, which covered the genome at 19.6x. This initial assembly required 147.50 hours on a 64-core computer with 512 GB of memory (1735.37 hours of CPU time). ALLPATHS-LG grouped 39,301 contigs into 4,396 scaffolds, with a total scaffold length of 1.027 Gb. The N50 scaffold size was 8.02 Mb, and 4.6% of bases were ambiguous (N's).

We then used HaploMerger (Huang et al. 2012) to improve the assembly by merging homologous contigs and removing those that had arisen from the erroneous split of two haplotypes. HaploMerger requires “soft-masking” repetitive elements in the genome, which we did with RepeatMasker version open-4.0.2 (Smit et al. 2013-2015). HaploMerger has been used to improve a number of genome assemblies in this manner (e.g., Derks et al. 2015; Davey et al. 2016). After running the original assembly through the HaploMerger pipeline using default settings (and manually breaking two scaffolds that HaploMerger indicated may have been misjoined), the final Siberian stonechat *de novo* assembly comprised 2,819 scaffolds, with a total scaffold length of 1.020 Gb; the N50 scaffold size increased to 10.0 Mb compared to the original assembly. We verified that the majority of removed scaffolds had fragment library coverage of less than 5x. HaploMerger therefore appears to have been successful in removing a large number of small scaffolds that likely represented duplicates (i.e., heterozygous regions). The 1,577 removed scaffolds spanned only 7.4 Mb (0.7% of the original assembly).

To assess completeness of the reference genome, we used NCBI command-line ‘blastn’ to search for 5561 ultraconserved elements identified by Faircloth et al. (2012) from an analysis of chicken, anole, and zebra finch. The final assembly contained 5486 (98.7%) of these ultraconserved elements. Because they are interspersed throughout the entire genome, this percentage can be considered an approximation for the completeness of the draft assembly; a value of 98.7% is evidence that the assembly covers nearly the entire Siberian stonechat genome.

Sampling

Most birds used in this study originated from the common-garden stonechat study that Eberhard Gwinner initiated in 1981 at the Max-Planck Institute in Andechs, Germany.

Specifically, parental populations originated from the following locations: Austrian stonechats from Lower Austria (48°14'N, 16°22'E); Irish stonechats from Iveragh Peninsula near Killarney, in the County of Kerry, Ireland (c. 52°N, 10°W); African stonechats from Lake Nakuru region, Kenya (0°14'S, 36°0'E), and Mount Meru region, Tanzania (3°50'S, 36°5'E); and Siberian stonechats from the vicinity of Naursum National Park (c. 51.5°N, 63°E), Kazakhstan. All blood samples for the Canary Islands stonechat were collected directly in the field between 2013 and 2016 at various locations on Fuerteventura, Canary Islands, Spain (Barranco de Mal Nombre, Fimbapaire, Norte de Fenimoy, Barranco de Jacomar, Barranco Gran Valle, Barranco de Los Canarios, Barranco de Vinamar).

Pooled sequencing

We selected between 49 and 56 individuals (including both males and females) from each stonechat taxon based on a careful assessment of DNA quantity via Trinean DropSense 96 multi-channel spectrophotometer (Trinean, Ghent, Belgium) and quality (check for integrity on a 2% agarose gel), and we created one library of pooled DNA for each taxon following the Illumina TruSeq DNA kit. Each library included an equimolar aliquot of DNA from each individual. We multiplexed 4 of the 5 five groups on one lane of an Illumina NextSeq sequencer (151-bp paired end reads), and ran the fifth with three unrelated samples on a second lane. Thus, each group was sequenced on approximately one-fourth of a lane.

We demultiplexed raw sequence data from the sequencer with the 'bcl2fastq' utility by Illumina, using default settings. This utility generates '.fastq' files after removing reads showing a 10% or greater error rate in the adapter sequence or more than 1 error in the barcode. It also masks adapter sequences extending into reads. We then used the program 'skewer' to conduct the following additional quality control measures on reads: trim the 3' end until quality ≥ 20 is reached; and remove reads with normalized error rate > 0.1 (default), indel error rate > 0.03 (default, based on comparison with known adapter sequence), mean base quality < 20 , or $> 15\%$ ambiguous bases (N's). Approximately 1% of the demultiplexed reads failed these criteria and were removed.

We used BWA-MEM (Li 2013) to align the pooled sequences to the reference genome, marking shorter split hits as secondary. We then converted the alignments to compressed BAM format using 'samtools view,' specifying a minimum mapping quality of 20. We sorted BAM files with 'samtools sort' and merged them across lanes using Picard's 'MergeSamFiles' (Picard: <http://broadinstitute.github.io/picard/>). Following this, we marked duplicate reads using Picard's MarkDuplicates utility; performed local realignment using the Genome Analysis Toolkit (GATK; RealignerTargetCreator and IndelRealigner) (McKenna et al. 2010; DePristo et al. 2011); and fixed mate information in Picard (FixMateInformation). Finally, we took the resulting 5 BAM files (one per taxon) and used 'samtools mpileup' to construct an mpileup file comparing the bases in overlapping reads at each position of the genome across populations.

Mapping quality for all stonechat taxa was high (Table S2). Mean mapping quality was lower for *Ficedula* species (*hypoleuca*: 35.64; *albicollis*: 35.82), but a high proportion of reads from these species were successfully mapped to the stonechat reference genome (*hypoleuca*: 0.93; *albicollis*: 0.95). This high mapping rate suggests that we are not introducing substantial bias by aligning flycatcher reads to the stonechat

reference genome. We detected bacterial DNA contamination in some of the stonechat pools; these sequences did not map to the reference and were thereafter ignored.

Mapping to *Ficedula* chromosomes

We assembled scaffolds from the stonechat assembly into draft chromosomes by mapping them to the *Ficedula albicollis* genome assembly, version 1.5 (RefSeq accession GCF_000247815.1; http://www.ncbi.nlm.nih.gov/assembly/GCF_000247815.1; [http://www.ncbi.nlm.nih.gov/genome/?term=txid59894\[orgn\]](http://www.ncbi.nlm.nih.gov/genome/?term=txid59894[orgn])) (Kawakami et al. 2014). We used SatsumaSynteny (Grabherr et al. 2010) to align the *Saxicola* draft genome to the *F. albicollis* assembly. This method unambiguously placed nearly all scaffolds of sufficient size (> 10 Kb) on a *F. albicollis* chromosome, with 85% of scaffolds mapping to single chromosomes across >70% of their extents. In the rare cases (~1%) where scaffolds mapped to more than one *Ficedula* chromosome across greater than 20% the scaffold length, we assigned the scaffold to the chromosome with the greatest amount of sequence aligned (always a majority of the scaffold). SatsumaSynteny thus allowed us to position scaffolds from the stonechat genome in the correct order and orientation along the chromosomes, assuming that synteny is conserved in these taxa (Ellegren 2013). This assumption appears robust given the high conservation of sequence within scaffolds. SatsumaSynteny successfully mapped 97.1% of the stonechat reference genome to a *Ficedula* chromosome. Most unmapped scaffolds had not passed the 10 Kb threshold.

Coverage heterogeneity

To rule out the possibility that variation in coverage could be driving differentiation patterns, we compared read depth in F_{ST} outlier regions to read depth outside of those regions. We selected the comparison of Irish and Siberian stonechats because this comparison showed arguably the most conspicuous F_{ST} peaks, and therefore any effect of coverage should be most pronounced. Because allele frequencies in adjacent 50 Kb windows are autocorrelated due to linkage and therefore contribute to pseudoreplication, we subsampled the genome at a ratio of 1:10. We used t -tests to test for differences inside and outside of outlier regions. Read depth was not significantly different within and outside of F_{ST} peaks for both taxa (Irish: $t = 0.78$, $df = 118.37$, $P = 0.44$; Siberian: $t = 0.89$, $df = 119.56$, $P = 0.39$). Specifically, for Irish stonechats, mean within-peak coverage was 26.01 and mean outside-of-peak coverage was 26.38. For Siberian stonechats these values were 15.08 and 15.23, respectively.

Phylogeny

We aligned raw reads from Pied and Collared Flycatcher re-sequencing data to the stonechat genome in order to call genotypes. We then selected 16,876,859 sites across the genome which satisfied the following criteria: minimum coverage of 5 in all populations; fixation of a single allele at the locus (allowing a maximum count of 1 of another allele because of the possibility for sequencing error); and variation in the fixed allele among the 7 taxa. Using these SNPs, we generated a phylogenetic tree with RAxML v. 8.2.6 (Stamatakis 2014) on CIPRES (<http://www.phylo.org>). We applied the Lewis correction, following the recommendation of Stamatakis (2014), for ascertainment bias resulting from the exclusion of constant sites and using 100 bootstrapped replicates to assess branch support.

Choice of window size and bandwidth size for genome-wide scans

We used a window size of 50 Kb for our genomic analyses because it provided us sufficiently fine resolution across the genome while still averaging over hundreds of SNPs per window. We felt it was important not to rely heavily on allele frequencies of individual SNPs because of the random variation in allele frequencies introduced by our pooled sequencing approach. We conducted a sensitivity analysis (not shown) and found that we identified fewer and larger outlier regions as we increased window size, but that the level of overlap detected between genomic landscapes did not systematically vary. We feel that this justifies a window size of 50 Kb because it allows us to capture relatively small outlier regions while keeping the number of regions to a manageable size for this whole-genome analysis of multiple taxa.

We selected a bandwidth of 30 because it allowed us to identify relatively small regions of differentiation while still providing a benefit by smoothing out apparent noise in the data. We conducted a sensitivity analysis (not shown) and found that the median size of outlier regions identified by our analysis stayed relatively constant until a bandwidth of about 50, after which we observed an increase. Therefore, we do not believe that we are biased towards detecting large outlier regions by using a bandwidth of 30. We also did not observe any systematic effect of bandwidth size on the level of overlap detected between genomic landscapes.

SUPPLEMENTARY FIGURES

Figure S1. Boxplots comparing F_{ST} and π between the Z chromosome and autosomes. Outlier values are not shown. Under neutral expectation, the equilibrium level of neutral variability is proportional to the effective population size, and the effective population size of the Z chromosome is expected to be three-fourths that of the autosomes because females only have one copy (Charlesworth 2001). We used a *t*-test for ratios (*t.test.ratio* function in the *mratio*s package) to test whether the ratio of π on the Z chromosome to π on the autosomes was significantly different from 0.75 (Djira et al. 2012). Stonechats and flycatchers showed π ratios between 0.74-0.83; Kenyan, Siberian and Canary Islands stonechats and Pied Flycatchers (*Ficedula hypoleuca*) had π ratios that did not significantly differ from 0.75, while Austrian and Irish stonechats and Collared Flycatchers (*Ficedula albicollis*) showed slightly more diverse Z chromosomes than expected by theory. In all cases, F_{ST} on the Z chromosome was elevated over that of the autosomes, with ratios in stonechats between 1.04-1.29, and a much higher ratio in flycatchers of 1.80.

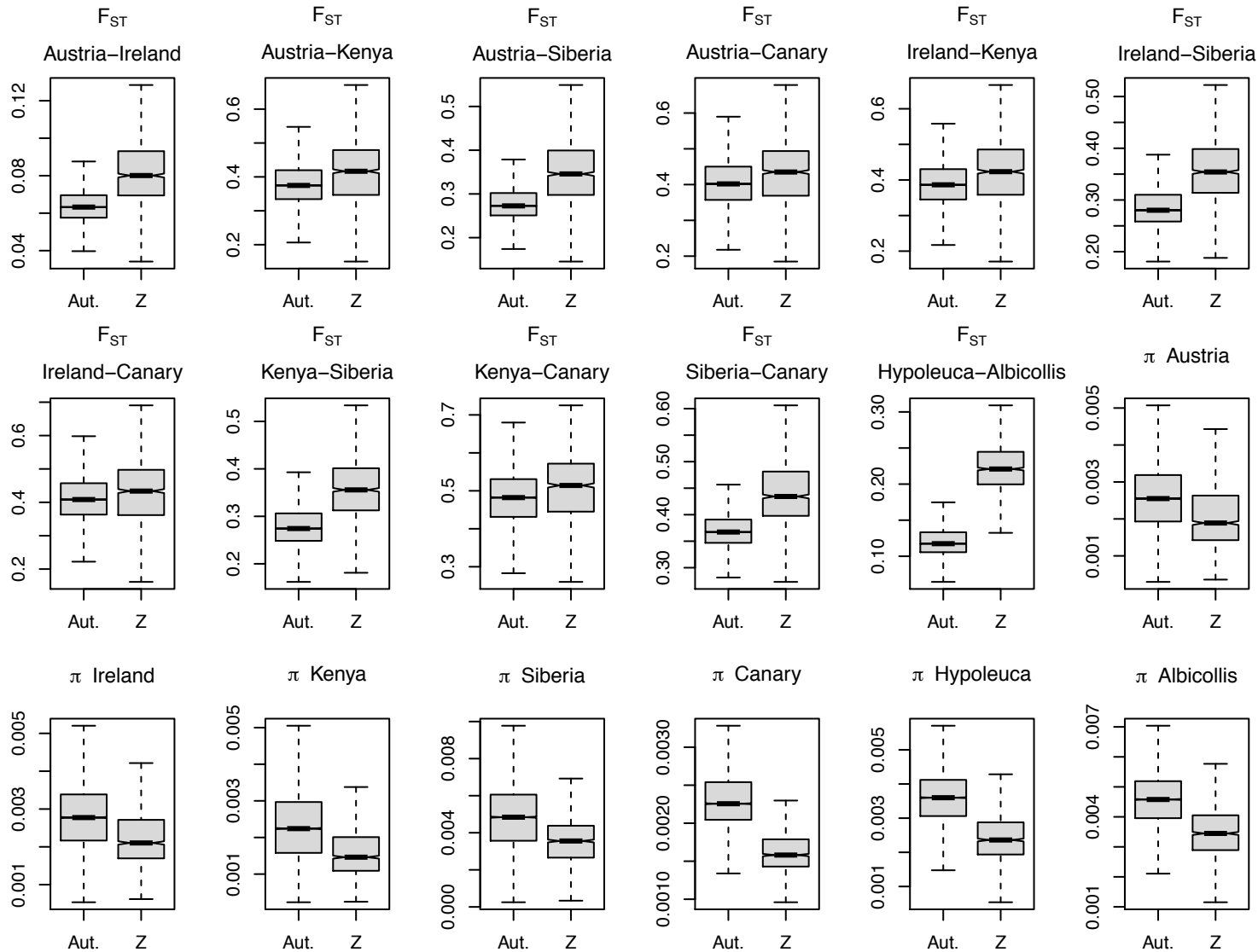
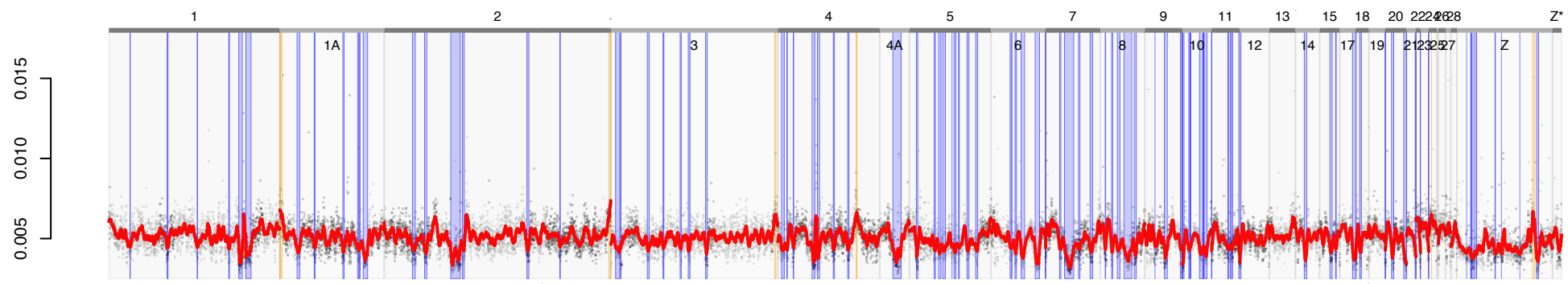
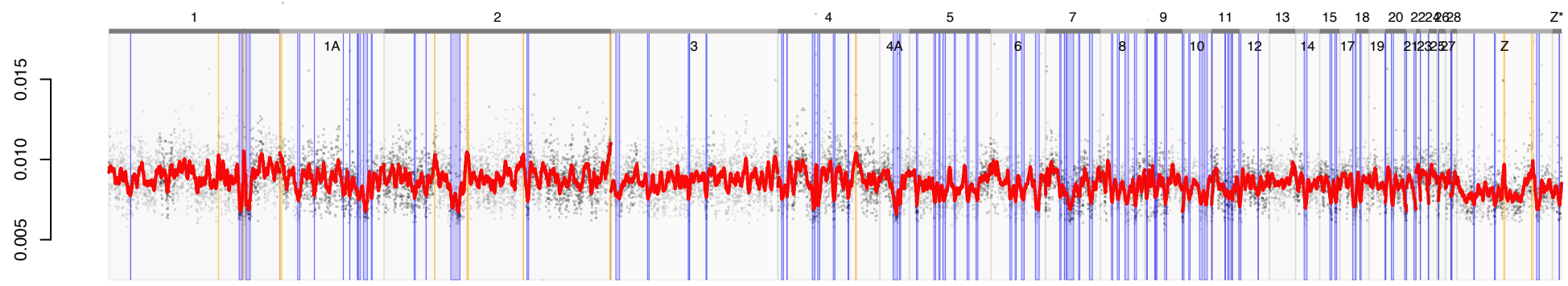


Figure S2. Genome-wide landscape of d_{XY} for pairwise comparisons of stonechats and Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). All stonechat comparisons showed very similar genomic landscapes of d_{XY} . Many outlier regions were also shared with *Ficedula*, especially on the larger chromosomes. For clarity, comparisons including Irish stonechats are not included (with the exception of Austria-Ireland), because of the high degree of similarity between Austrian and Irish taxa. The colored lines are kernel-based density smoothers. Individual points represent 50-Kb windows; scaffolds alternate dark gray and light gray coloring. Chromosomes (based on alignment to *Ficedula albicollis*) are delineated by thick dark gray or light gray lines on the upper border of each plot and are labeled above this line. Z* indicates a flycatcher Z chromosome linkage group that could not be exactly placed in the flycatcher genome assembly. Shaded orange rectangles show d_{XY} peaks and blue rectangles show d_{XY} valleys.

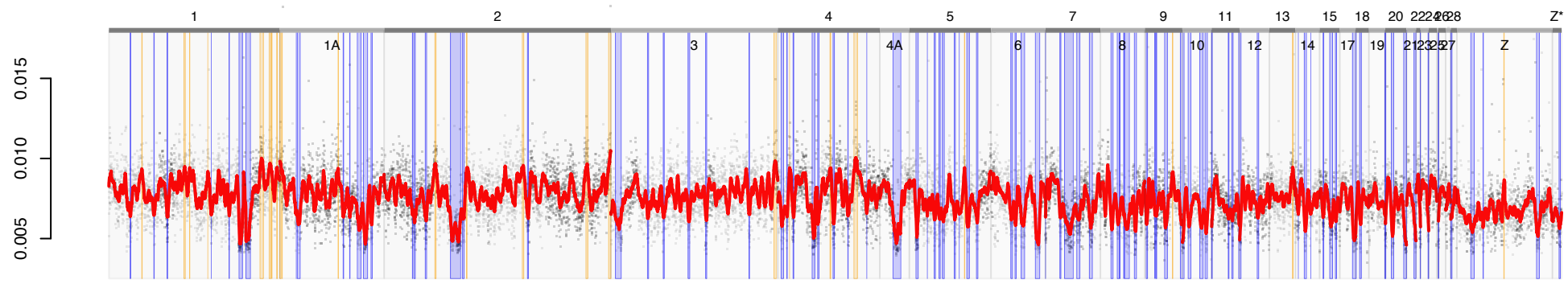
Austria
&
Ireland



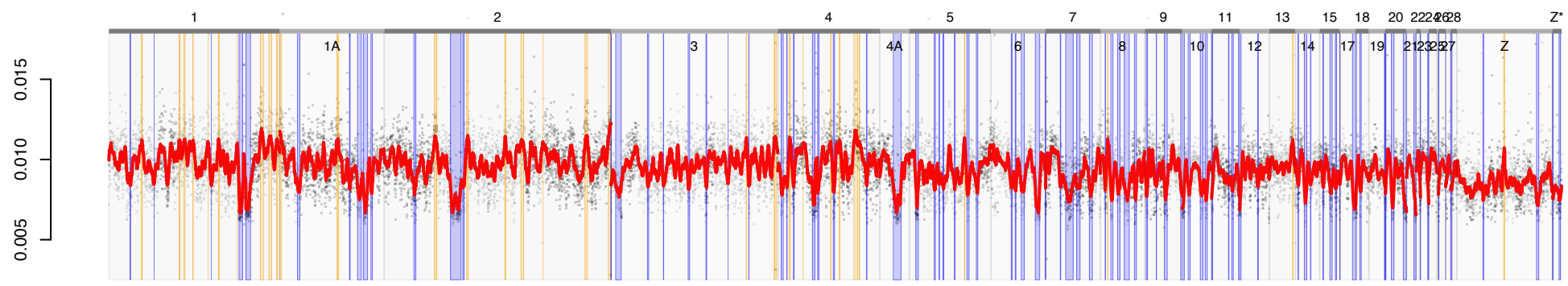
Austria
&
Canary



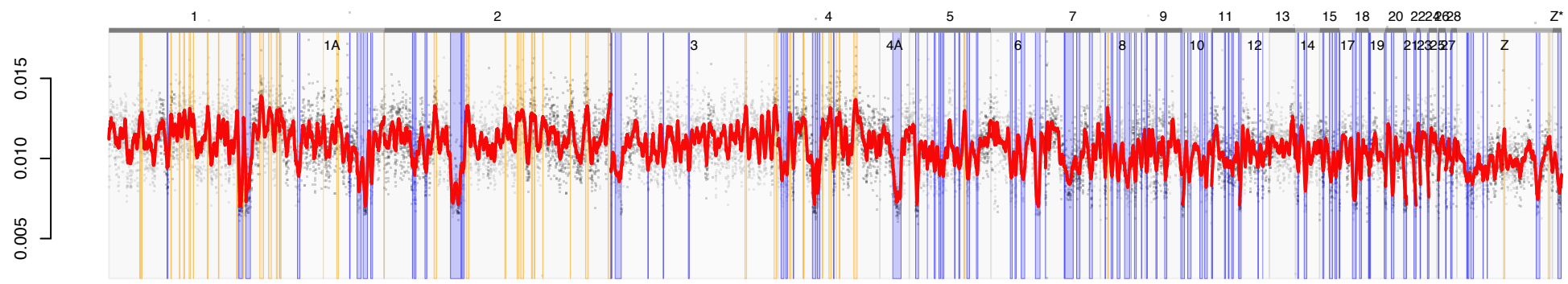
Austria
&
Kenya



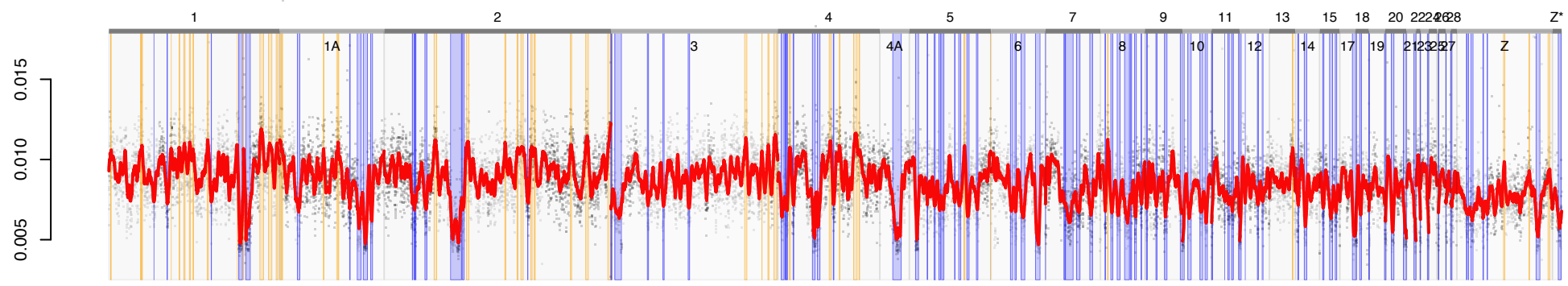
Kenya
&
Canary



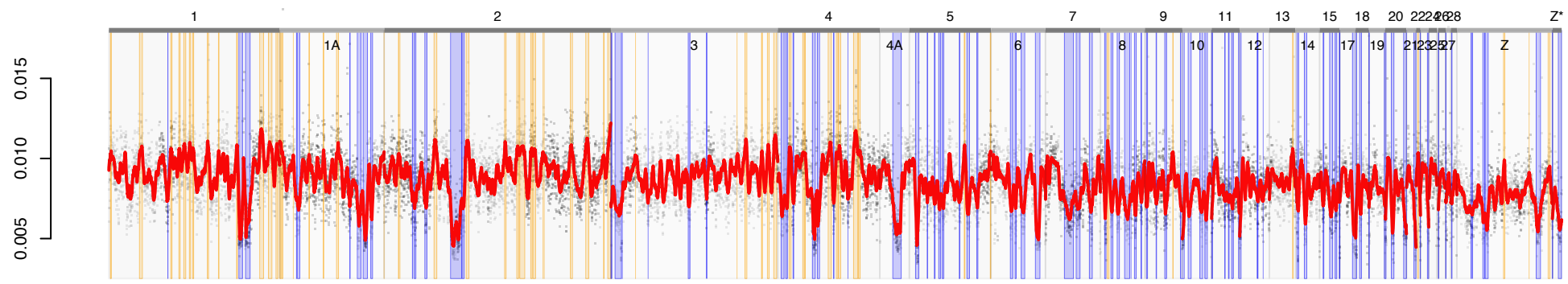
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

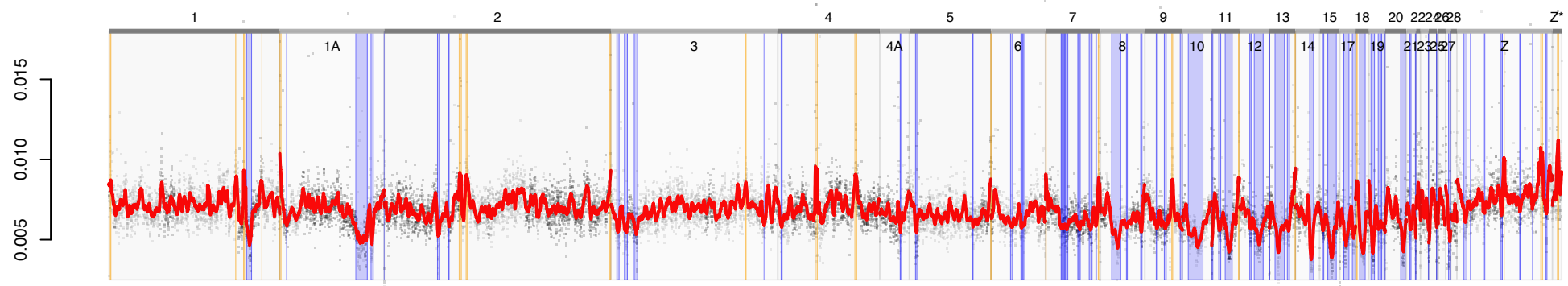
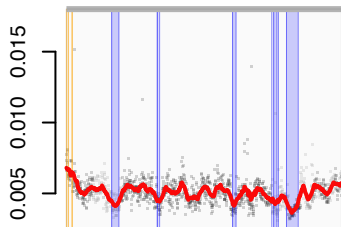
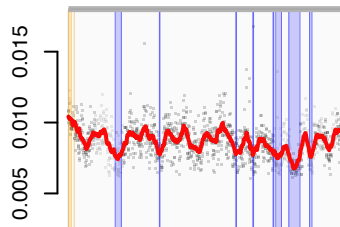


Figure S3. D_{XY} across stonechat chromosome 1A. All stonechat comparisons show very similar fluctuations, including two pronounced valleys. The largest valley is also apparent in the comparisons of Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). Blue rectangles indicate significant d_{XY} valleys. See Figure S2 for other details. For clarity, comparisons including Irish stonechats are not included (with the exception of Austria-Ireland) because of the high degree of similarity between Austrian and Irish populations.

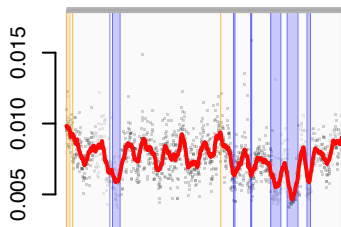
Austria
&
Ireland



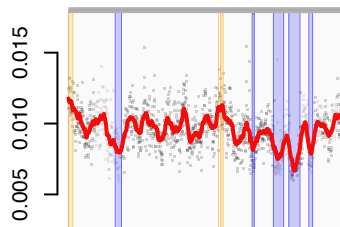
Austria
&
Canary



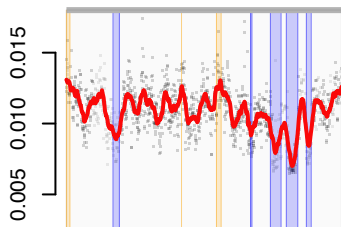
Austria
&
Kenya



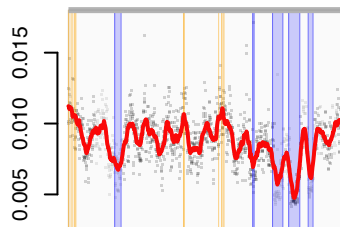
Kenya
&
Canary



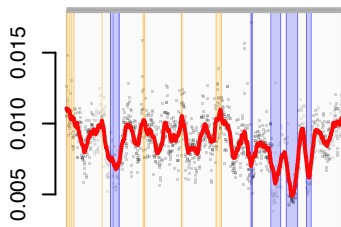
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

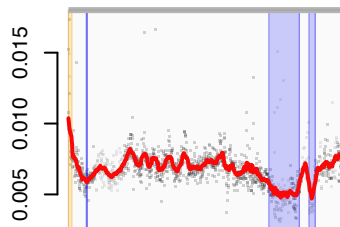


Figure S4. Correlation of high d_{XY} regions among stonechats and flycatchers (*Ficedula hypoleuca* and *albicollis*, or “Hyp.” and “Alb.”). Matrix shows outlier similarity scores, which quantify the number of high- d_{XY} “peaks” shared among different comparisons. Some comparisons including Irish stonechats are not shown because of their similarity to Austrian stonechats. All tests were significant after applying a false discovery rate correction. Cells with yellow backgrounds indicate that four independent taxa are being compared.

Hyp. Alb.	0.46	0.43	0.3	0.43	0.3	0.39	0.39
Aus. Can.	0.71	0.54	0.85	0.77	0.85	0.85	
Aus. Ire.	0.57	0.71	0.57	0.71	0.57	0.57	
Aus. Ken.	0.96	0.79	0.96	0.83			
Aus. Sib.	0.97	0.98	0.91				
Ken. Can.	0.8	0.86					
Ken. Sib.	0.8						
Sib. Can.							


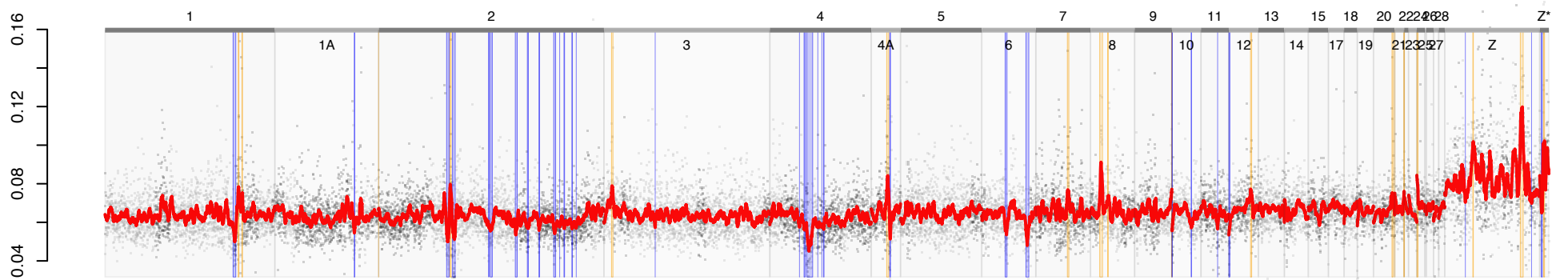
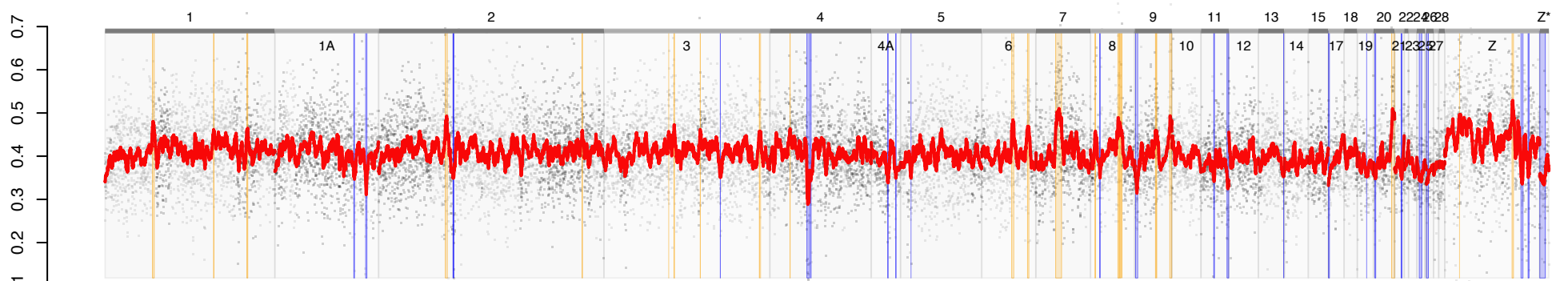
 Four different taxa compared

Figure S5. Genome-wide landscape of F_{ST} for pairwise comparisons of stonechats and Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). Pairs including Siberian stonechats showed the most conspicuous peaks; other comparisons showed less distinct outlier regions. Some comparisons (e.g., Kenya-Canary Is.) showed F_{ST} valleys in the same regions as the F_{ST} peaks of other comparisons. Shaded orange rectangles show F_{ST} peaks and blue rectangles show F_{ST} valleys. See Figure S2 for other details.

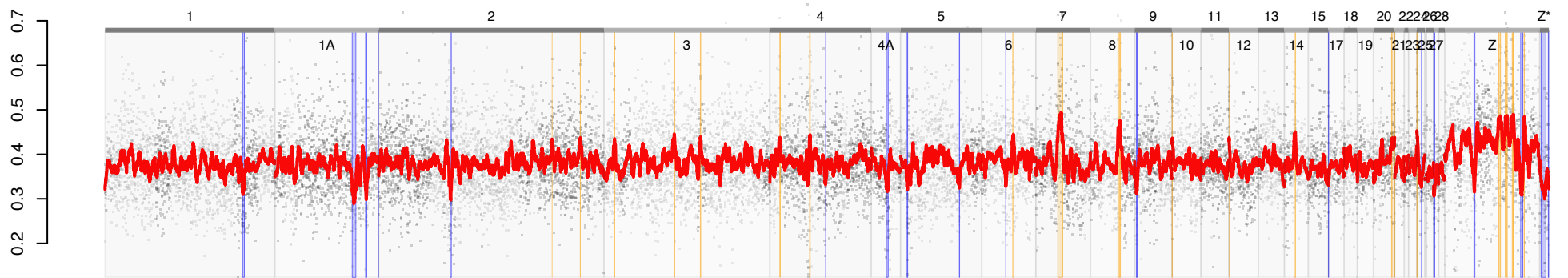
Austria
&
Ireland



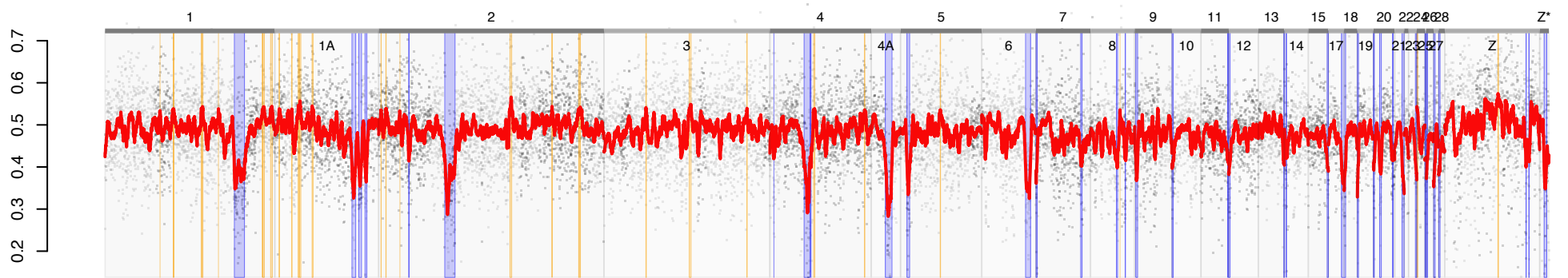
Austria
&
Canary



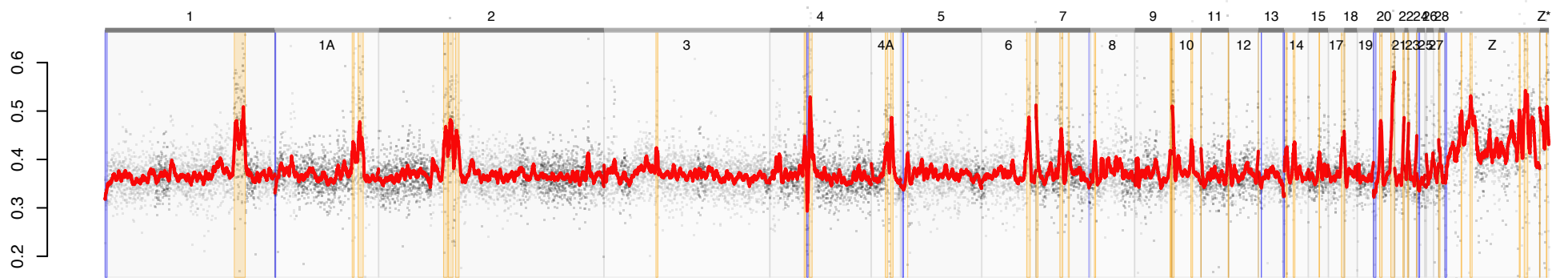
Austria
&
Kenya



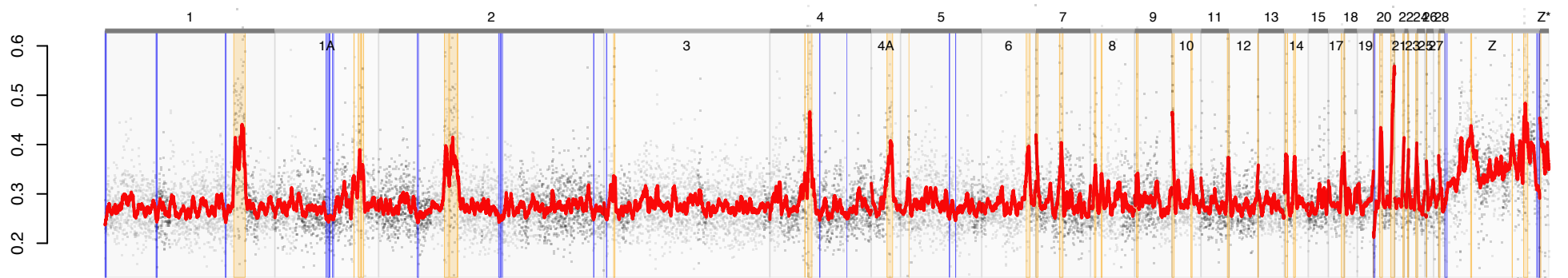
Kenya
&
Canary



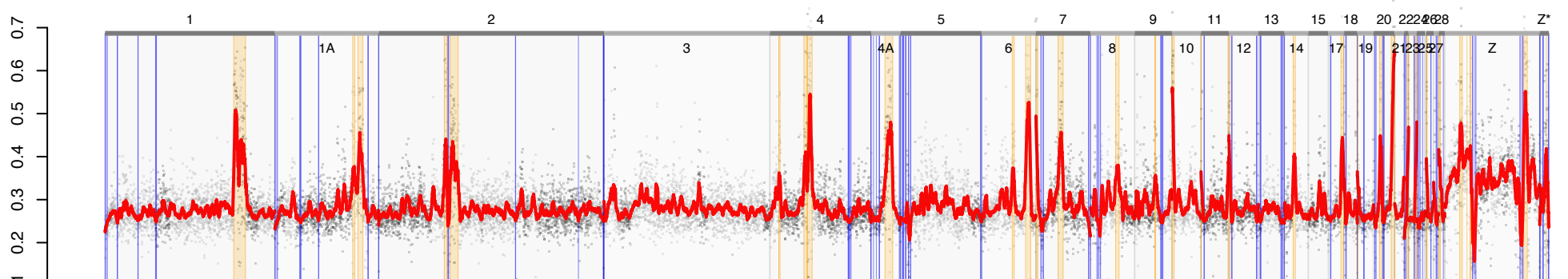
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

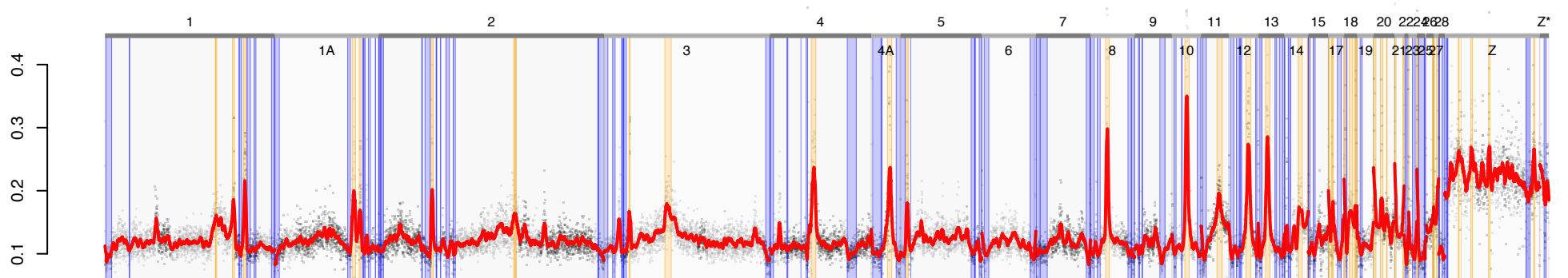


Figure S6. Correlation of nucleotide diversity (π) and d_{XY} among stonechats and flycatchers (*Ficedula hypoleuca* and *albicollis*, or “Hyp.” and “Alb.”). (A) shows outlier similarity scores, which quantify the number of low- d_{XY} “valleys” that coincide with low π “valleys” in the two taxa being compared (top row and bottom row). The top (No. 1) and bottom (No. 2) rows show the results for each of the two taxa being compared. This is necessary because π is a single-population statistic, while F_{ST} and d_{XY} compare two populations. All comparisons were significant after applying a false discovery rate correction. (B,C,D) show scatterplots where each point represents one 50-Kb genomic window. Refer to Figs. 2-3 for details.

A

1:	Hyp.	Aus.	Aus.	Aus.	Aus.	Ken.	Ken.	Sib.
2:	Alb.	Can.	Ire.	Ken.	Sib.	Can.	Sib.	Can.

Low d_{XY}			B		C	D		
Low π (1)	0.75	0.88	0.94	0.9	0.81	0.77	0.76	0.92
Low d_{XY}			B		C	D		
Low π (2)	0.73	0.55	0.96	0.82	0.9	0.56	0.91	0.63

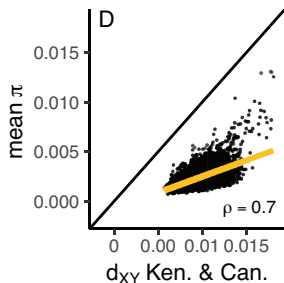
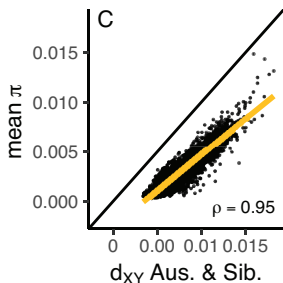
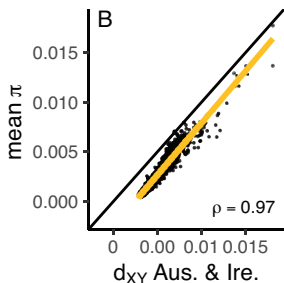


Figure S7. Correlation of nucleotide diversity (π) and standardized nucleotide diversity (π/d_{XY}) among stonechats and flycatchers (*Ficedula hypoleuca* and *albicollis*, or “Hyp.” and “Alb.”). (A,B,C,D,F,G) show scatterplots where each point represents one 50-Kb genomic window. (E) shows outlier similarity scores, which quantify the number of π/d_{XY} valleys shared among different comparisons (upper triangle of matrix) and the number of π valleys shared among different comparisons (lower triangle of matrix). Refer to Figs. 2-3 for details.

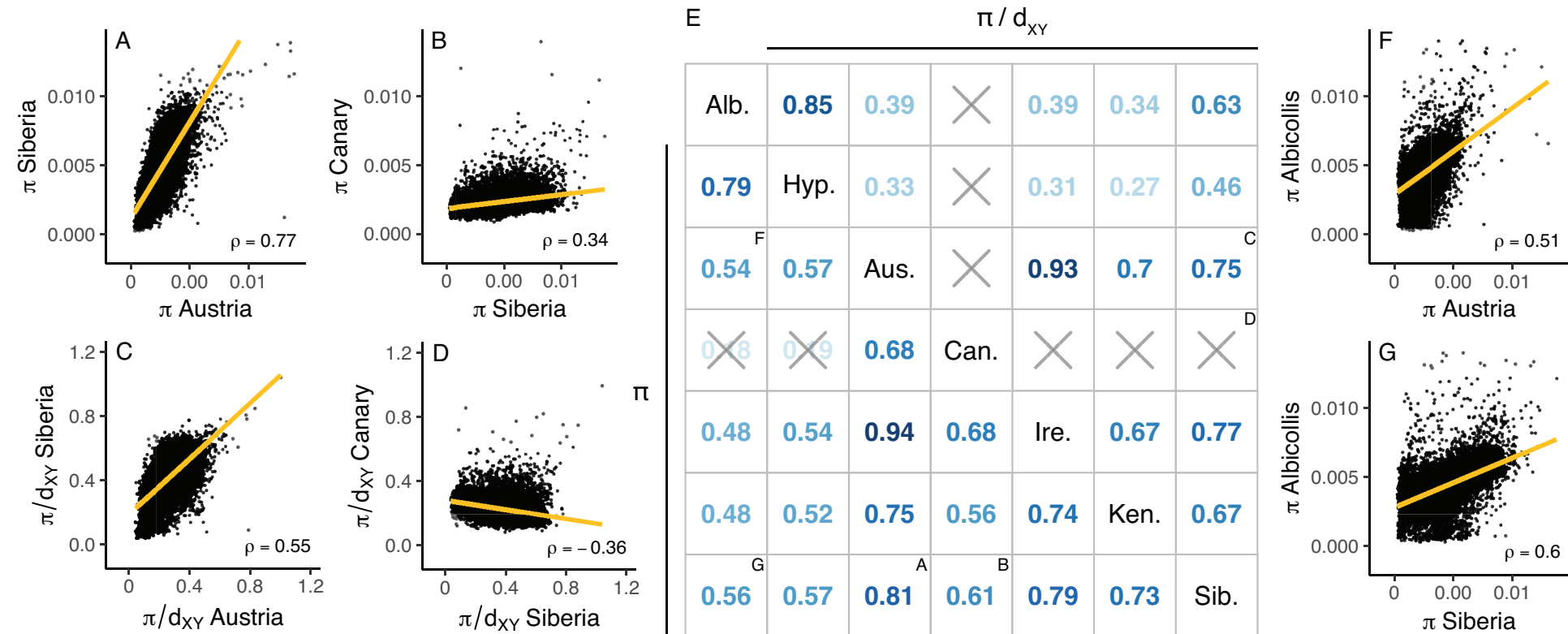
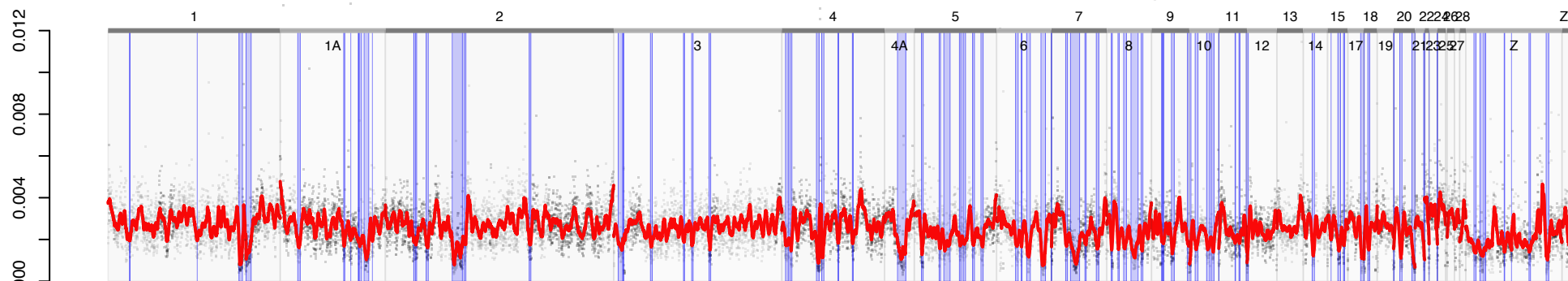
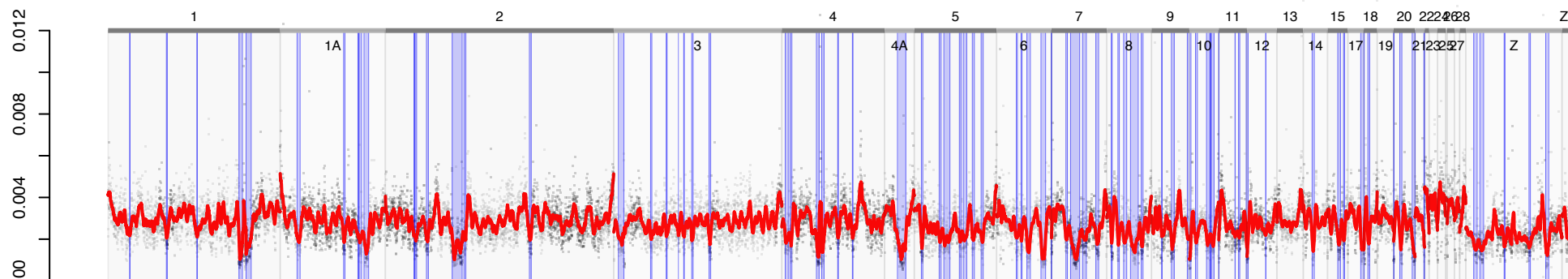


Figure S8. Genome-wide landscape of π for five stonechat taxa. Canary Islands stonechats generally did not share the valleys present in the genomes of the other taxa. See Fig. S2 for other details.

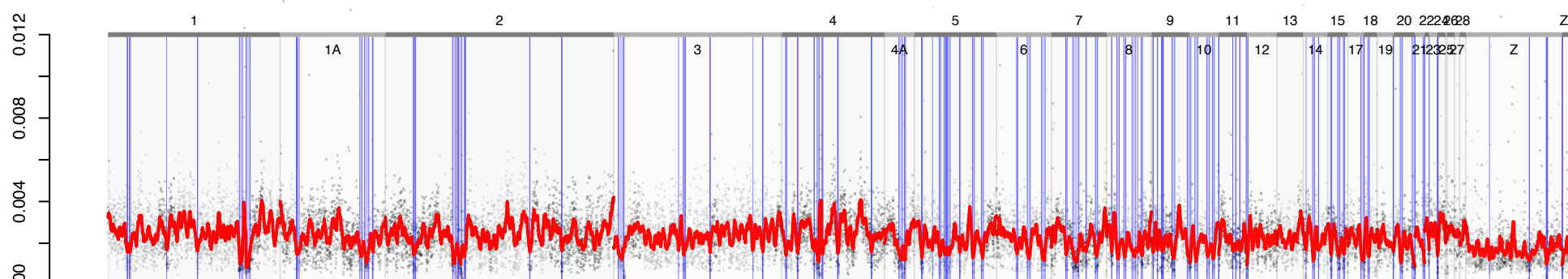
Austria



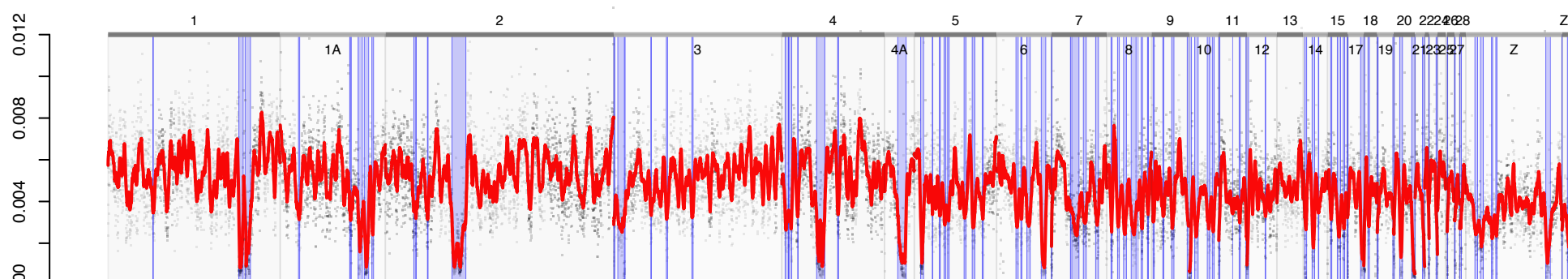
Ireland



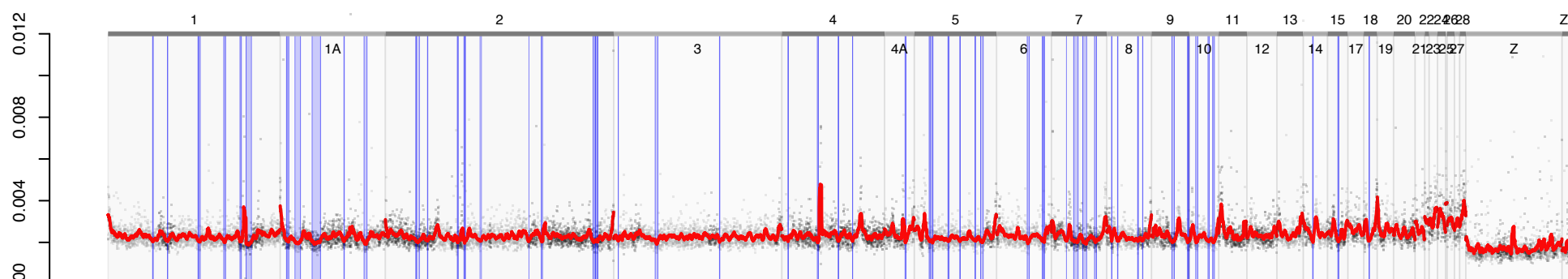
Kenya



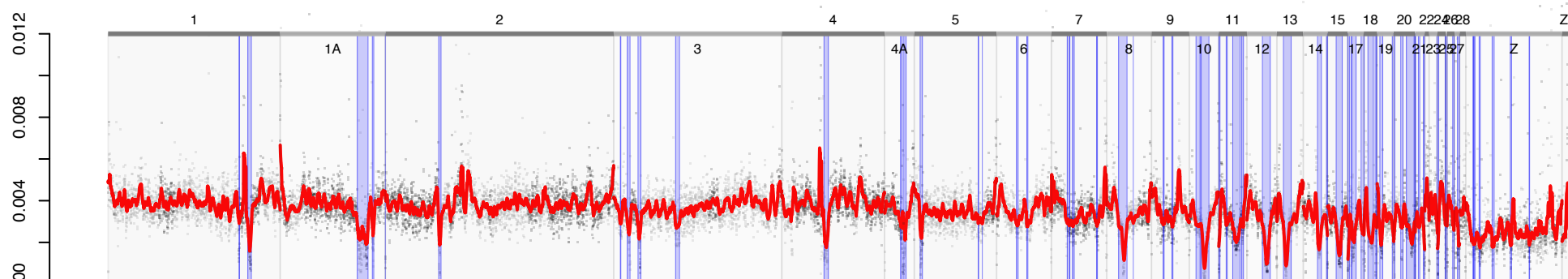
Siberia



Canary



F. hyp.



F. alb.

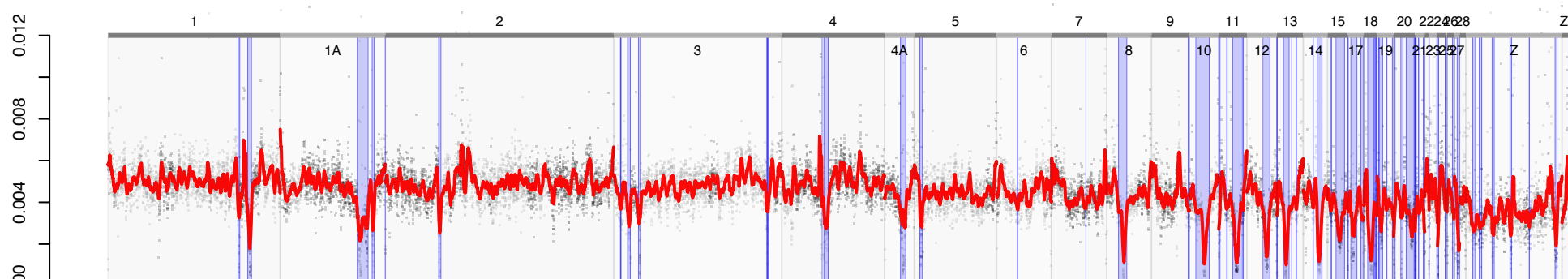
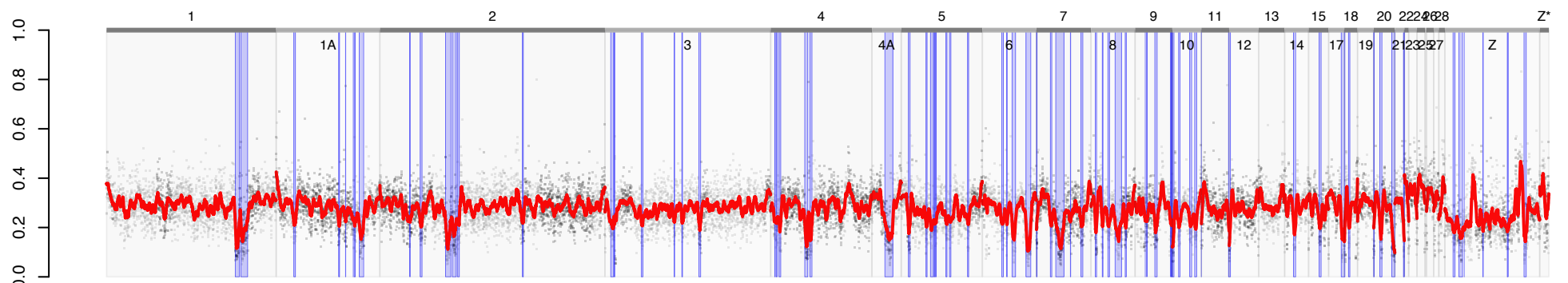
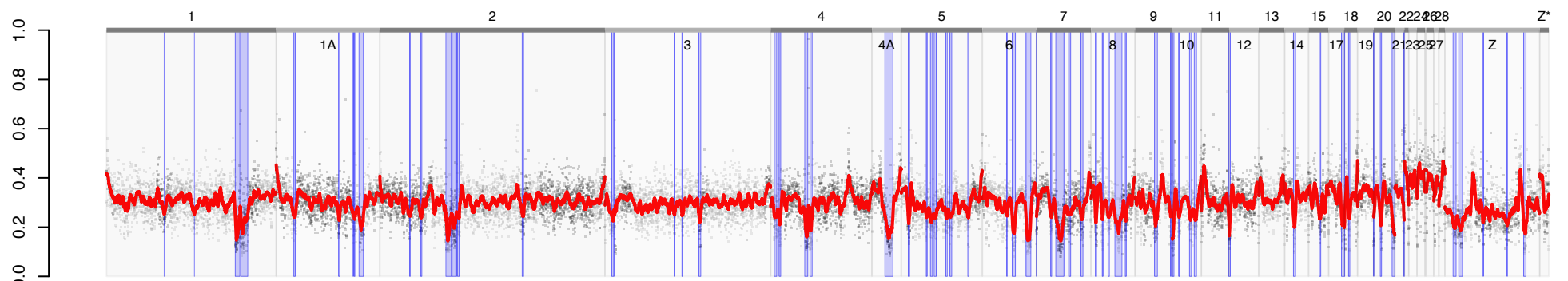


Figure S9. Genome-wide landscape of standardized nucleotide diversity (π/d_{XY}) for five stonechat and two flycatcher taxa. Canary Islands stonechats did not share the valleys present in the genomes of the other taxa. See Fig. S2 for other details.

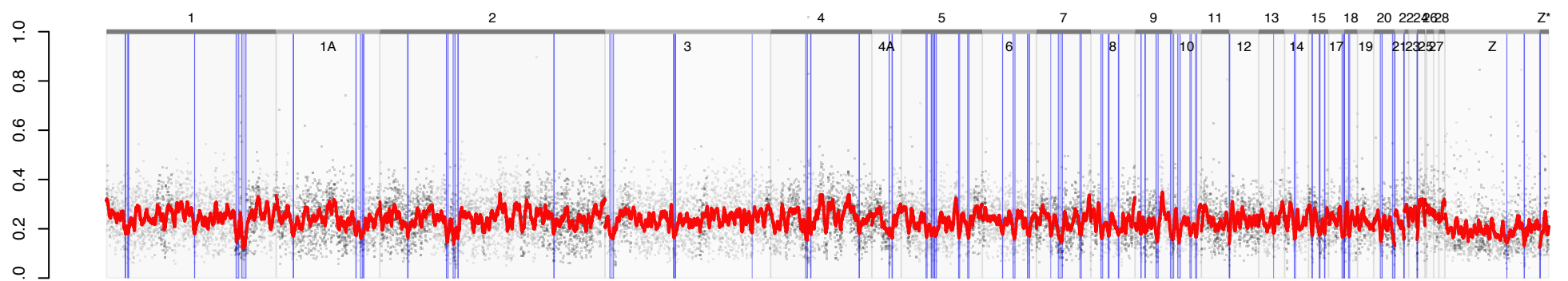
Austria



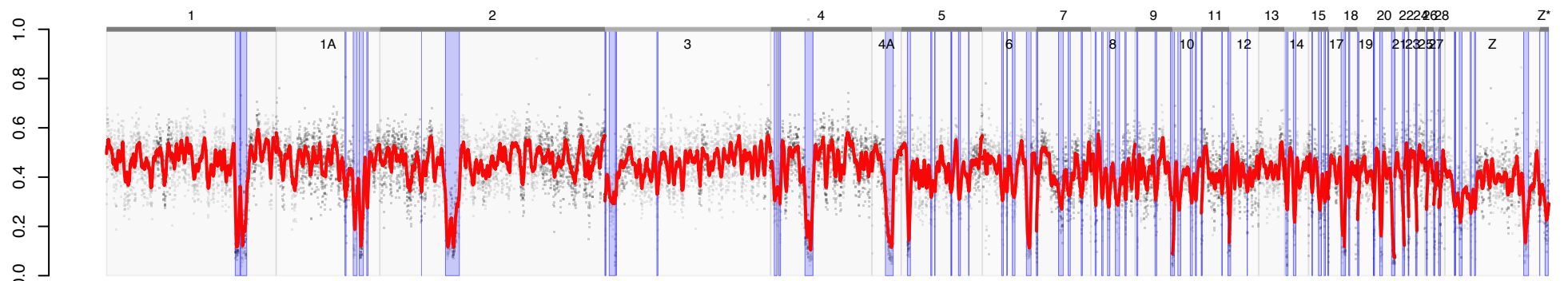
Ireland



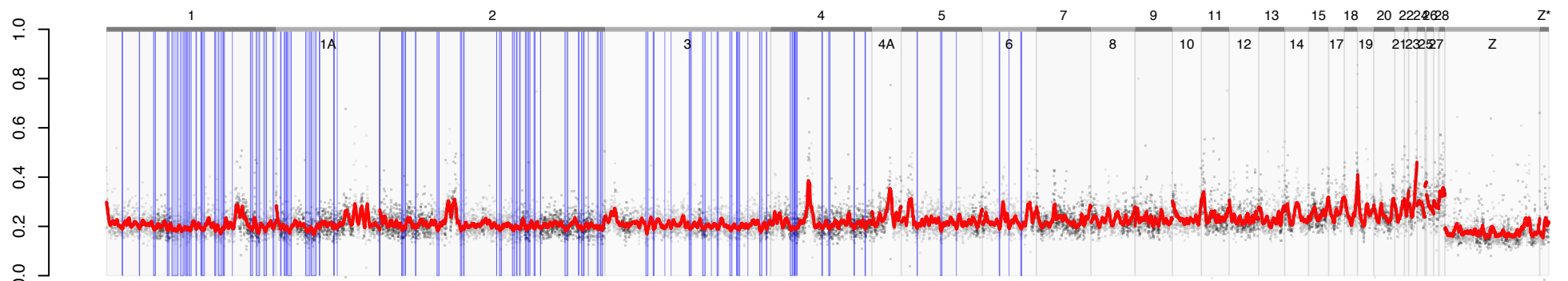
Kenya



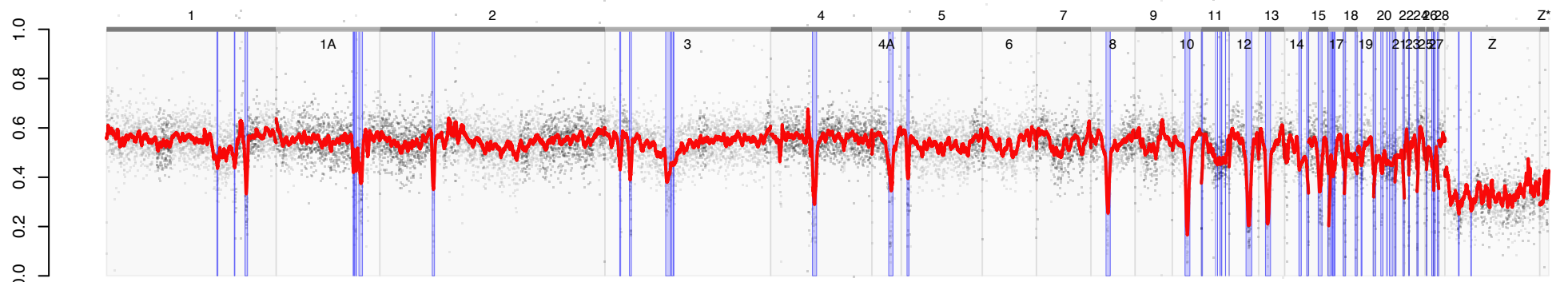
Siberia



Canary



F. hyp.



F. alb.

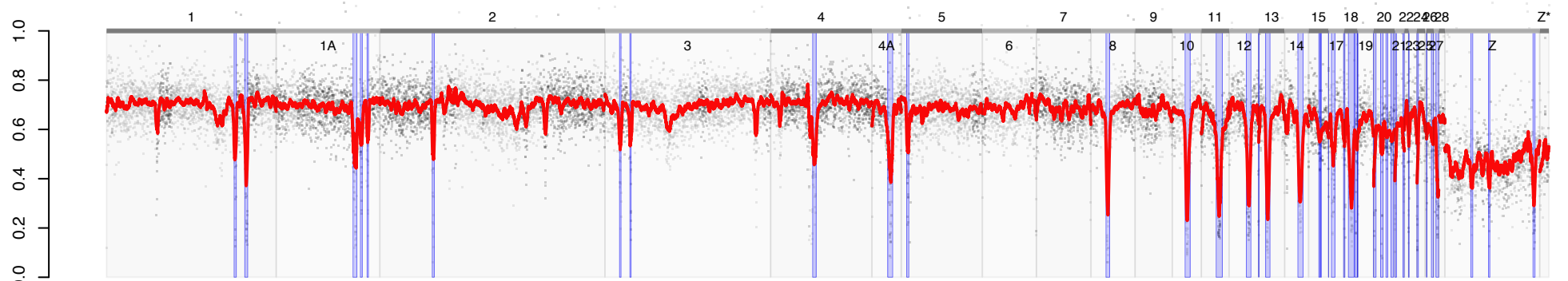


Figure S10. Correlation of nucleotide diversity (π) with Tajima's D and Fay & Wu's H statistics. Shown are outlier similarity scores, which quantify the number of low Fay & Wu's H "valleys" that coincide with low π (top section) and low Tajima's D (bottom section). Refer to Figs. 2-3 for details.

Aus. Can. Ire. Ken. Sib.

LOW π
Low F & W 's H

0.79

~~0.54~~

0.71

0.6

0.74

Low Tajima's D
Low F & W 's H

0.6

~~0.17~~

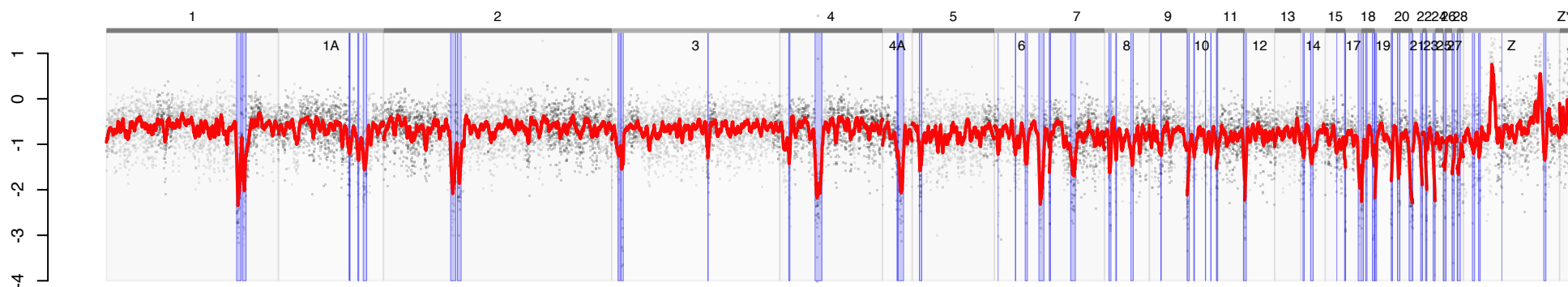
0.44

0.4

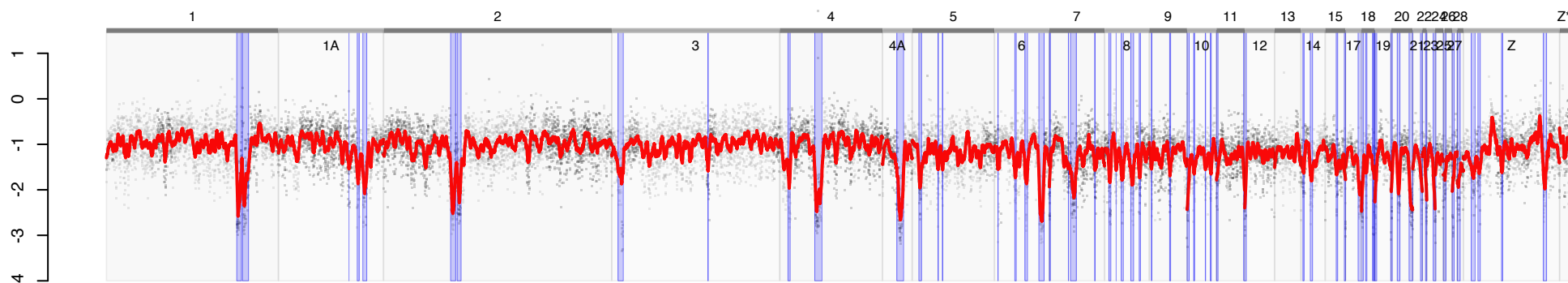
0.52

Figure S11. Genome-wide landscape of Tajima's D for five stonechat and two flycatcher taxa. Austrian, Irish, Siberian, and Kenyan stonechats shared similar genomic landscapes of Tajima's D . Canary Islands stonechats showed a different pattern, with very low Tajima's D across the entire genome. See Fig. S2 for other details.

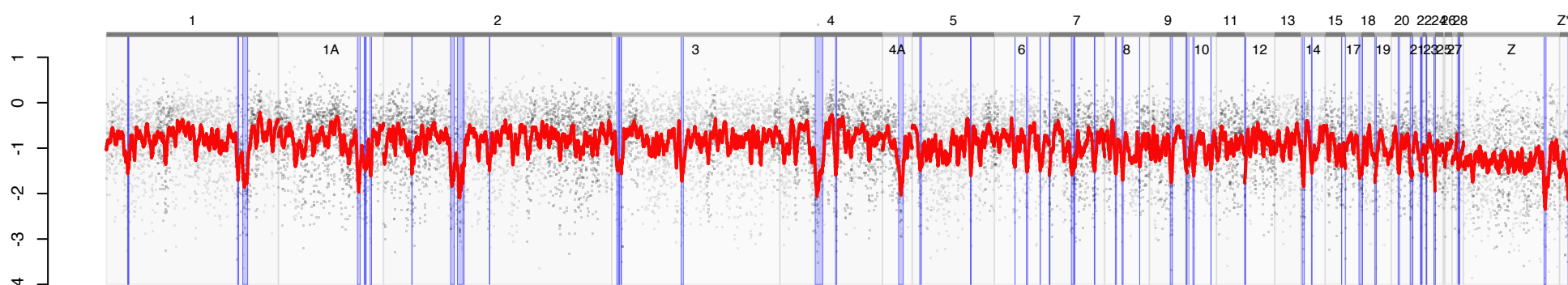
Austria



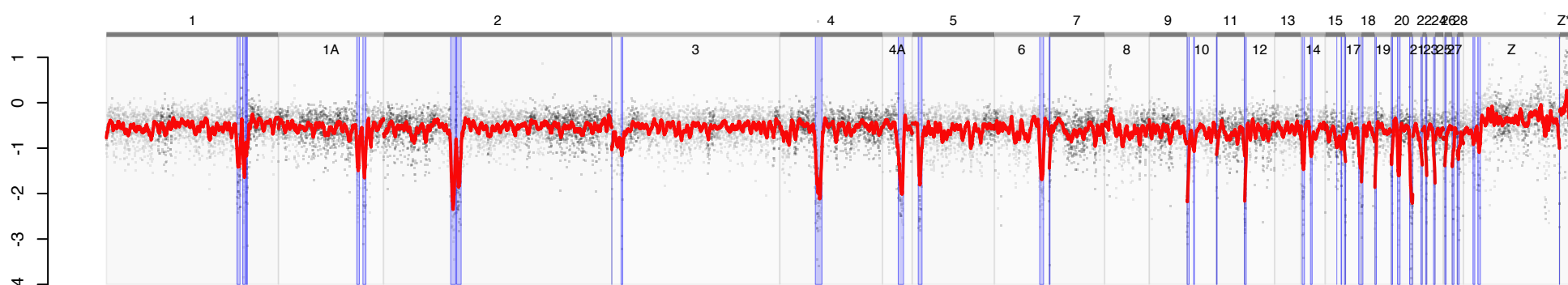
Ireland



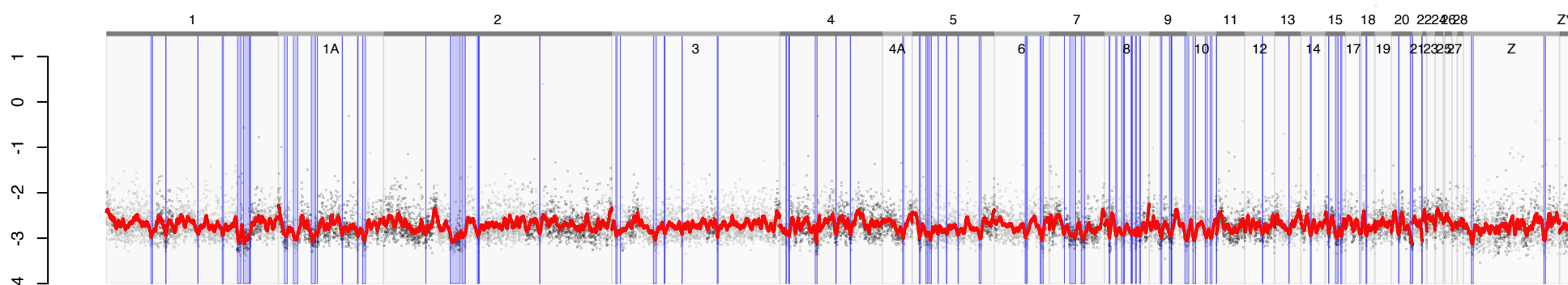
Kenya



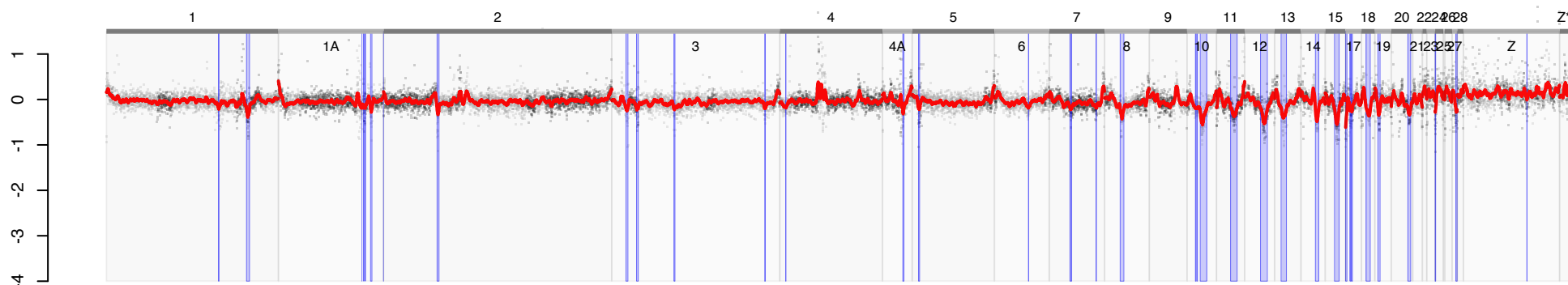
Siberia



Canary



F. hyp.



F. alb.

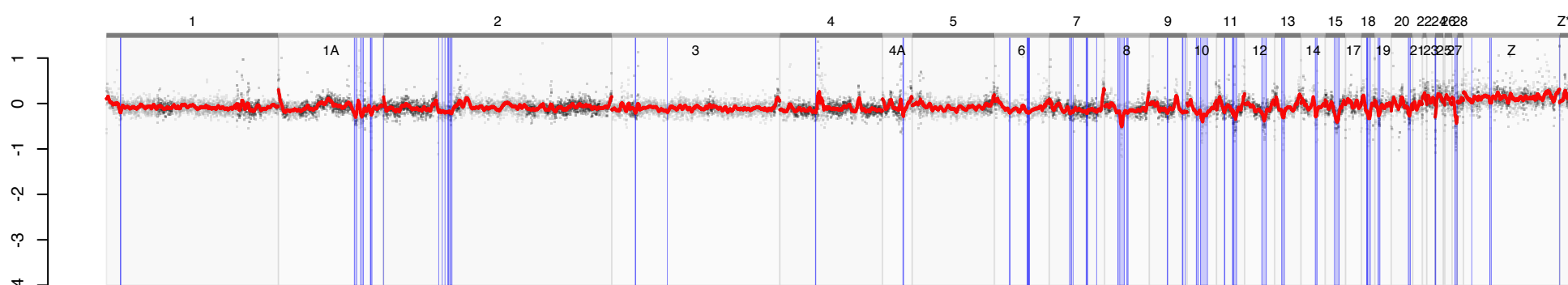
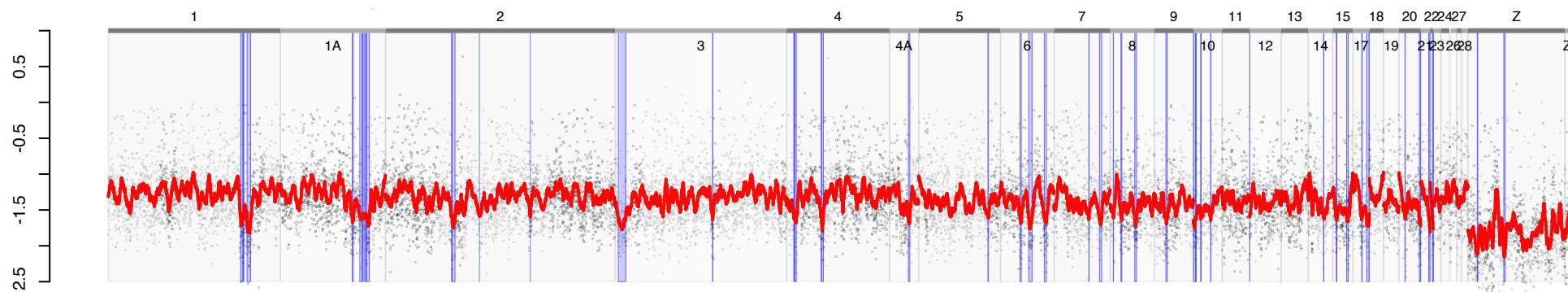
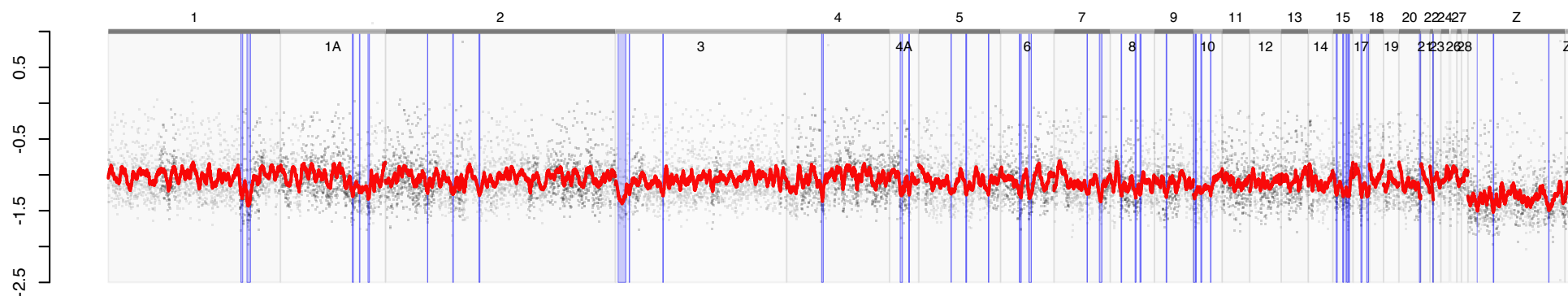


Figure S12. Genomic landscape of Fay and Wu's H for each of five stonechat taxa. See Fig. S2 for other details.

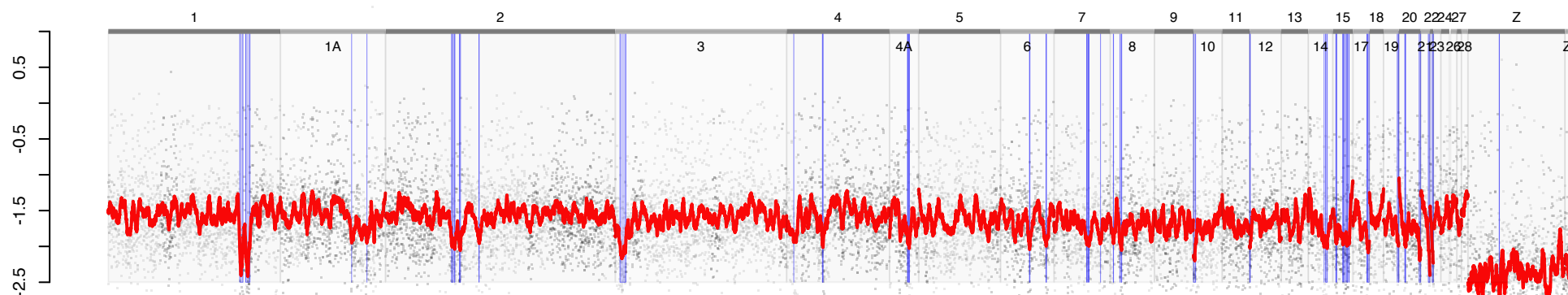
Austria



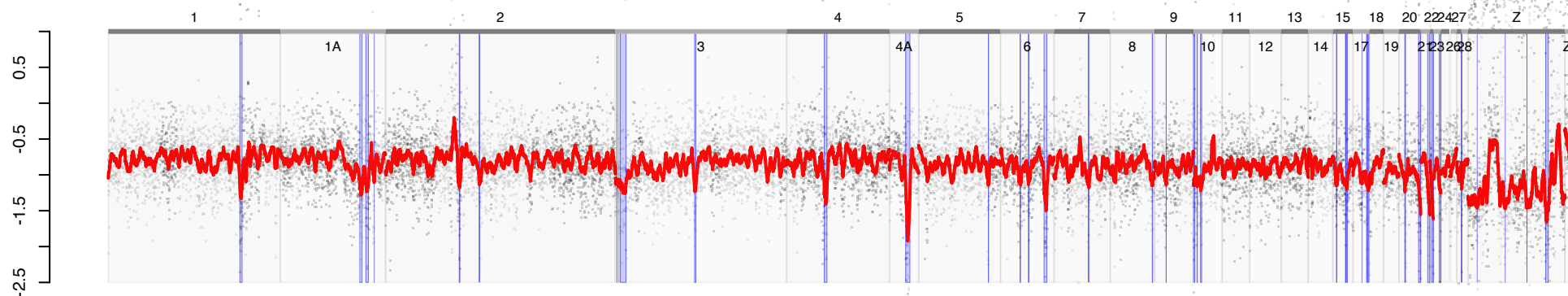
Ireland



Kenya



Siberia



Canary

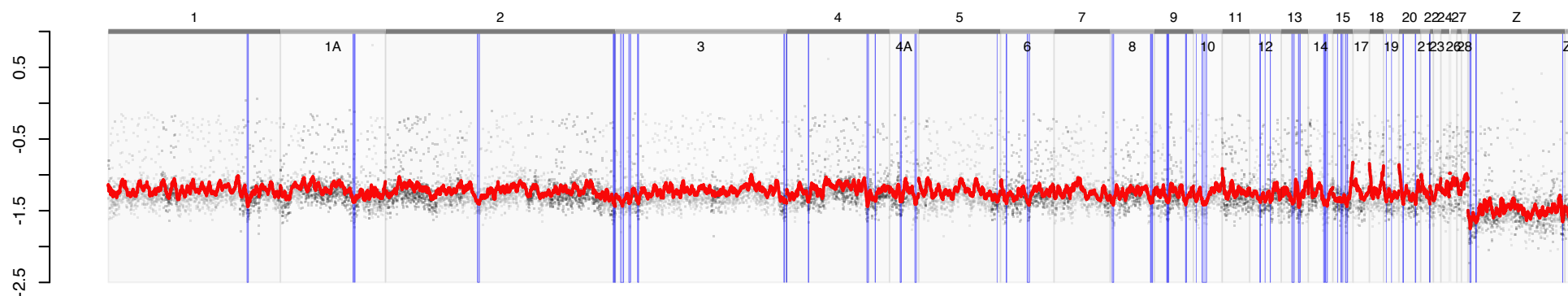


Figure S13. Boxplot of Tajima's D for five stonechat taxa. Each data point represents one 50-Kb window. Canary Islands stonechats showed the lowest median Tajima's D .

Tajima's D

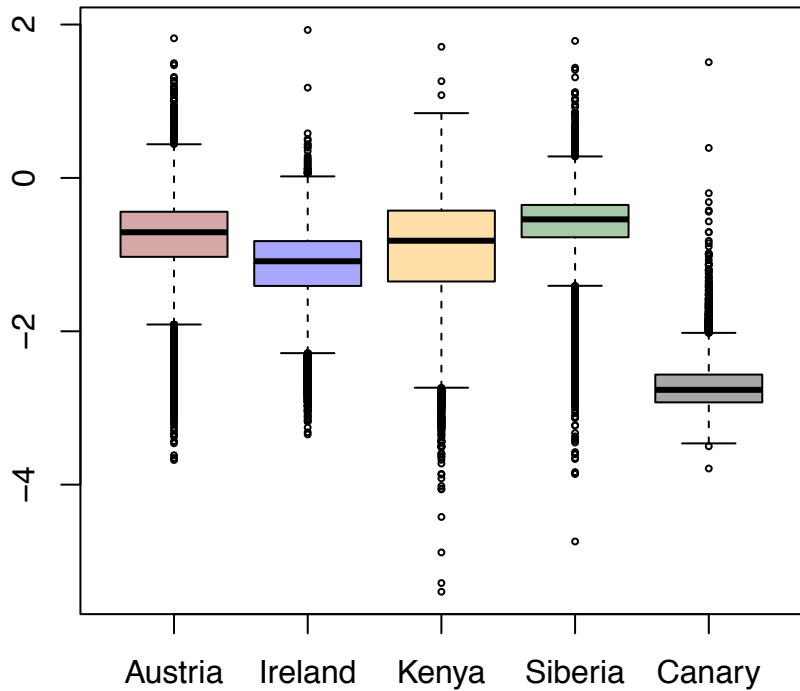
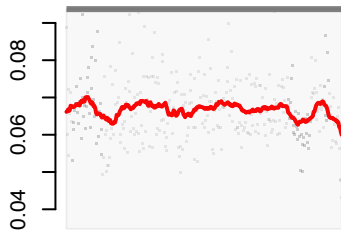
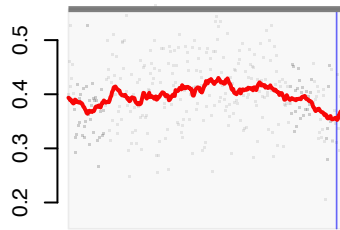


Figure S14. F_{ST} across stonechat chromosome 13, including Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). The *Ficedula* comparison shows a distinct peak, which is not present in any stonechat comparison. This suggests that the evolutionary processes driving differentiation in this chromosome are potentially unique to *Ficedula*. See Fig. S2 for other details.

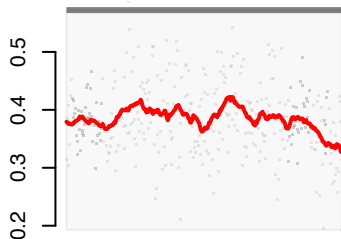
Austria
&
Ireland



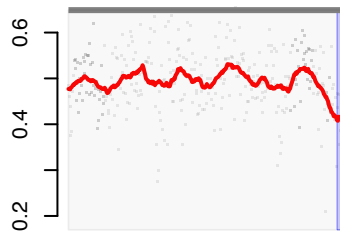
Austria
&
Canary



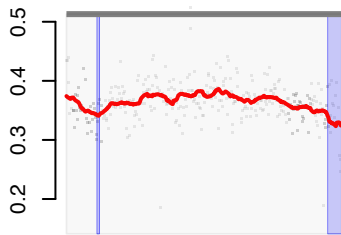
Austria
&
Kenya



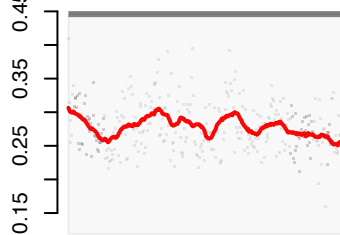
Kenya
&
Canary



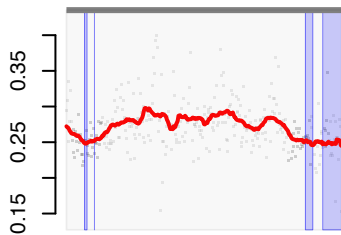
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

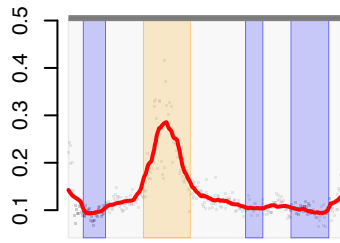
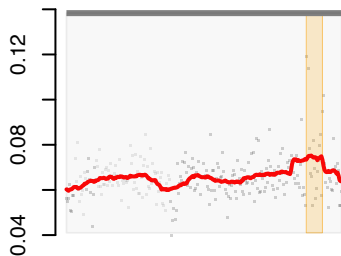
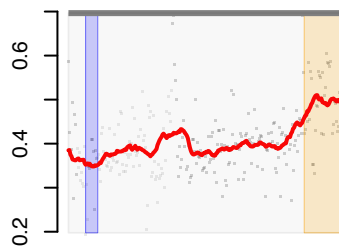


Figure S15. F_{ST} across stonechat chromosome 20, including Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). Comparisons including the Siberian population show a distinct peak at the right end of the chromosome. Notably, this peak is absent in *Ficedula*, suggesting that the evolutionary processes driving divergence in this chromosome are potentially unique to the stonechat radiation. Also note that the Kenya-Canary comparison shows a valley in the center of the chromosome, where there is a peak in other comparisons. See Fig. S2 for other details.

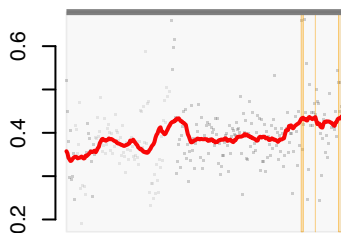
Austria
&
Ireland



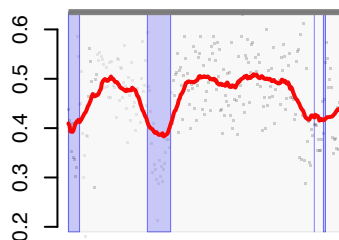
Austria
&
Canary



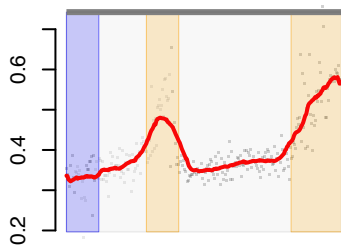
Austria
&
Kenya



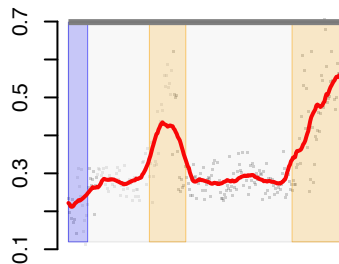
Kenya
&
Canary



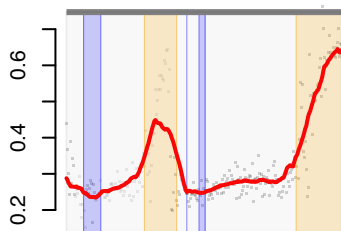
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

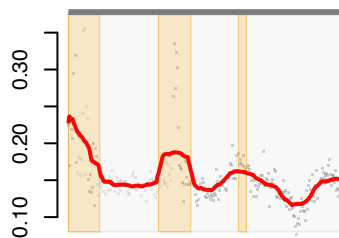
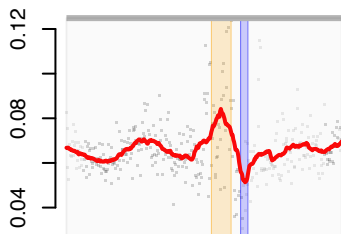
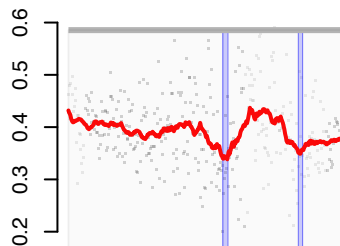


Figure S16. F_{ST} across stonechat chromosome 4A, including Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). Comparisons including Siberian stonechats, *Ficedula*, and Austria-Ireland show a distinct peak. The other comparisons show a valley in the same region. See Fig. S2 for other details.

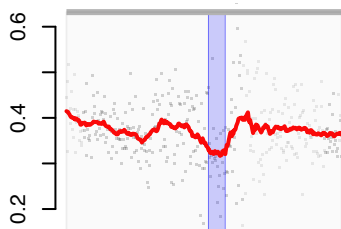
Austria
&
Ireland



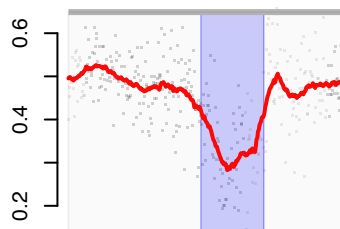
Austria
&
Canary



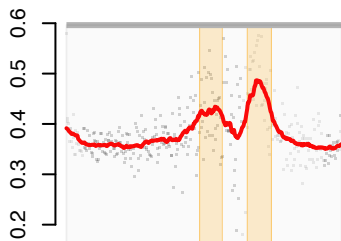
Austria
&
Kenya



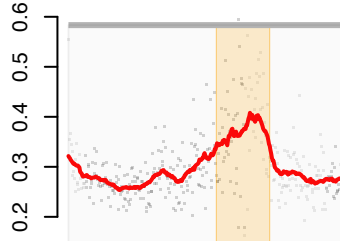
Kenya
&
Canary



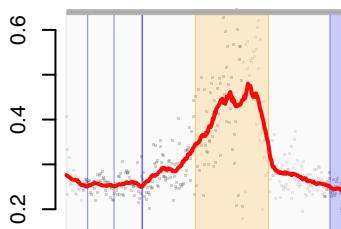
Siberia
&
Canary



Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis

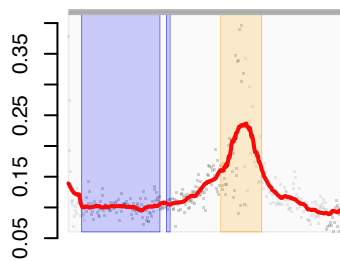
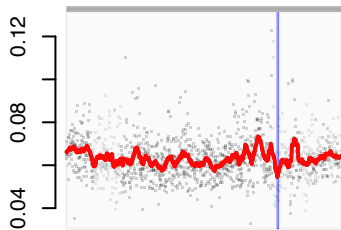
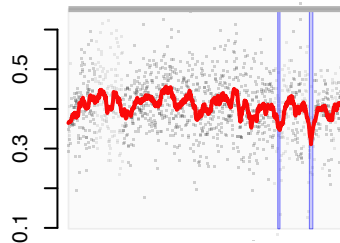


Figure S17. F_{ST} across stonechat chromosome 1A, including Pied and Collared Flycatchers (*Ficedula albicollis* and *F. hypoleuca*). Comparisons including Siberian stonechats and *Ficedula* show a distinct peak. Other comparisons show a valley in the same region. See Fig. S2 for other details.

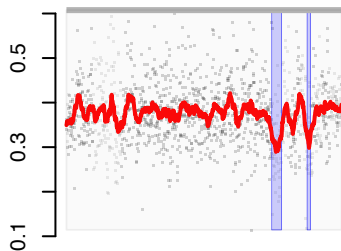
Austria
&
Ireland



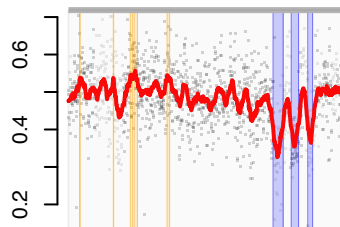
Austria
&
Canary



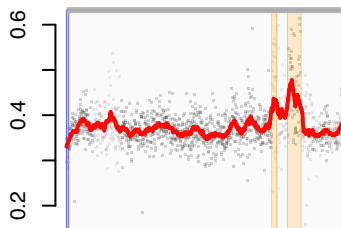
Austria
&
Kenya



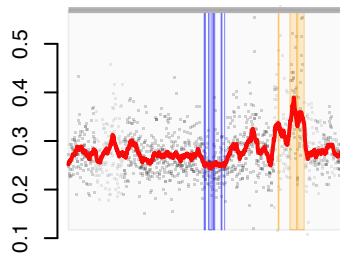
Kenya
&
Canary



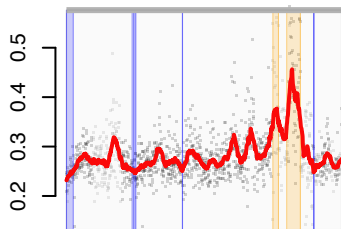
Siberia
&
Canary



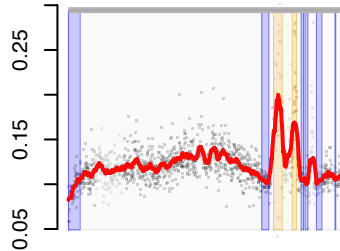
Kenya
&
Siberia



Austria
&
Siberia



F. hypoleuca
&
F. albicollis



SUPPLEMENTARY TABLES

Table S1. Origin, sex, and relatedness information of stonechats included in this study. Kinship matrices were calculated with the *kinship* function in the R package *kinship2* (Therneau and Sinnwell 2015) using a pedigree of captive stonechats, and values presented are the mean (+SD) values from each kinship matrix. Inbreeding coefficients were calculated with the *calcInbreeding* function in the R package *pedigree* (Coster 2013). IQR stands for interquartile range.

	Origins		Sex		Relatedness				Mean inbreeding	SD inbreeding
	Direct from wild	Hatched in captivity	Male	Female	Mean kinship	SD kinship	Median kinship	IQR kinship		
Austria	1	48	27	22	0.014	0.045	0	0	0.006	0.018
Ireland	27	27	26	28	0.009	0.043	0	0	0.002	0.012
Kenya	1	50	18	33	0.009	0.039	0	0	0.012	0.05
Siberia	0	52	30	22	0.033	0.064	0	0.0625	0.005	0.017
Canary	56	0	38	18	-	-	-	-	-	-

Table S2. Summary of alignment of Illumina 150-bp reads from five stonechat taxa to the draft reference genome. Mapping quality is given after filtering out alignments with a mapping quality of 20 or lower.

Taxon	Reads Mapped	Mean (Median) Coverage	Mean Mapping Quality
Kenya	98,758,285	13.8 (12.7)	45.61
Ireland	185,976,416	26.1 (24.8)	45.26
Austria	135,110,173	18.8 (17.7)	45.14
Siberia	107,623,583	14.9 (13.9)	45.80
Canary Islands	176,167,216	24.7 (23.9)	45.64

Table S3. *Ficedula* individuals included in the study (data from the Sequence Read Archive, or SRA, project ERP007074, published in Smeds et al. (2015)).

Species	SRA Run	Mean coverage of <i>Ficedula albicollis</i> genome
<i>F. hypoleuca</i>	ERR637490	16x
	ERR637491	12x
	ERR637492	11x
	ERR637493	11x
	ERR637494	13x
	ERR637495	16x
	ERR637496	18x
	ERR637501	14x
	ERR637503	14x
	ERR637504	12x
	<i>F. albicollis</i>	ERR637505
ERR637506		14x
ERR637508		10x
ERR637511		16x
ERR637512		15x
ERR637513		14x
ERR637515		11x
ERR637519		12x
ERR637522		11x
ERR637523		13x

SUPPLEMENTARY REFERENCES

- Charlesworth, B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. *Genetics Research* 77:153-166.
- Coster, A. 2013. pedigree: Pedigree functions. R package version 1.4.
- Davey, J. W., M. Chouteau, S. L. Barker, L. Maroja, S. W. Baxter, F. Simpson, M. Joron, J. Mallet, K. K. Dasmahapatra, and C. D. Jiggins. 2016. Major Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3: Genes|Genomes|Genetics*.
- DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-498.
- Derks, M. F. L., S. Smit, L. Salis, E. Schijlen, A. Bossers, C. Mateman, A. S. Pijl, D. de Ridder, M. A. M. Groenen, M. E. Visser, and H.-J. Megens. 2015. The Genome of Winter Moth (*Operophtera brumata*) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. *Genome Biology and Evolution* 7:2321-2332.
- Djira, G. D., M. Hasler, D. Gerhard, and F. Schaarschmidt. 2012. mratios: Inferences for ratios of coefficients in the general linear model. R package version 1.3. 17.
- Ellegren, H. 2013. The Evolutionary Genomics of Birds. *Annual Review of Ecology, Evolution, and Systematics* 44:239-259.
- Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. *Systematic Biology*.
- Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, and D. B. Jaffe. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* 108:1513-1518.
- Grabherr, M. G., P. Russell, M. Meyer, E. Mauceli, J. Alföldi, F. Di Palma, and K. Lindblad-Toh. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26:1145-1151.
- Huang, S., Z. Chen, G. Huang, T. Yu, P. Yang, J. Li, Y. Fu, S. Yuan, S. Chen, and A. Xu. 2012. HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome research* 22:1581-1588.
- Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, P. Olason, and H. Ellegren. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology* 23:4035-4058.

- Knief, U. and W. Forstmeier. 2015. Mapping centromeres of microchromosomes in the zebra finch (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma*:1-12.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20:1297-1303.
- Smeds, L., V. Warmuth, P. Bolivar, S. Uebbing, R. Burri, A. Suh, A. Nater, S. Bures, L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnstrom, M. Ruzic, S.-A. Saether, G.-P. Saetre, J. Torok, and H. Ellegren. 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun* 6.
- Smit, A., R. Hubley, and P. Green. 2013-2015. RepeatMasker Open-4.0.
- Stamatakis, A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*.
- Therneau, T. M. and J. Sinnwell. 2015. kinship2: Pedigree Functions. R package version 1.6.4.
- Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Volker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, and R. K. Wilson. 2010. The genome of a songbird. *Nature* 464:757-762.