

1 Signatures of human-commensalism in the house sparrow genome

2

3 Mark Ravinet¹, Tore Oldeide Elgvin^{1,2}, Cassandra Trier¹, Mansour Aliabadian³ Andrey

4 Gavrilov⁴ & Glenn-Peter Sætre¹

5

6 ¹Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway

7 ²Natural History Museum, University of Oslo, Norway

8 ³Ferdowsi University of Mashad, Iran

9 ⁴Institute of Zoology, Ministry of Education and Science of the Republic of Kazakhstan,

10 Kazakhstan

11

12

13 **Corresponding author:** mark.ravinet@ibv.uio.no

14

15 **Running head:** Genomic signature of commensalism in house sparrows

16

17 **Keywords:** Passer, anthrodependency, selective sweeps, amylase genes, Neolithic
18 revolution

19

20

21 **Word count:** (limit = approx 6000 words)

22

23 **Number of figures:** 4

24

25 **Number of tables:** 1

26

27

28

29

30

31

32

33

34 **Abstract**

35

36 House sparrows (*Passer domesticus*) are a hugely successful anthrodependent species;
37 occurring on nearly every continent. Yet, despite their ubiquity and familiarity to
38 humans, surprisingly little is known about their origins. We sought to investigate the
39 evolutionary history of the house sparrow and to identify the processes involved in its
40 transition to a human-commensal niche.

41

42 We used a whole genome resequencing dataset of 120 individuals from three Eurasian
43 species, including three populations of Bactrianus sparrows, a non-commensal,
44 divergent house sparrow lineage occurring in the Near East. Coalescent modelling
45 supports a split between house and Bactrianus sparrow 11 Kya and an expansion in the
46 house sparrow

47 6 Kya, consistent with the spread of agriculture following the Neolithic revolution.
48 Commensal house sparrows therefore likely moved into Europe with the spread of
49 agriculture following this period.

50

51 Using the Bactrianus sparrow as a proxy for a pre-commensal, ancestral house
52 population, we performed a comparative genome scan to identify genes potentially
53 involved with adaptation to an anthropogenic niche. We identified clear signatures of
54 recent, positive selection in the genome of the commensal house sparrow that are
55 absent in Bactrianus populations. The strongest selected region encompasses two major
56 candidate genes; *COL11A* - which regulates craniofacial and skull development and
57 *AMY2A*, part of the amylase gene family which has previously been linked to adaptation
58 to high-starch diets in humans and dogs.

59

60 Our work examines human-commensalism in an evolutionary framework, identifies
61 genomic regions likely involved in rapid adaptation to this new niche and ties the
62 evolution of this species to the development of modern human civilization.

63

64

65

66

67

68 **Introduction**

69

70 Human activity has had a dramatic impact on life on earth, both negatively and
71 positively with respect to biodiversity. With the advent of agriculture and establishment
72 of more permanent settlements following the Neolithic revolution, came the creation of
73 novel niches that a number of species have been able to utilise. Species that have
74 adapted to a life in anthropogenic surroundings, ranging from pests such as bedbugs or
75 head lice to the precursors to domesticated animals, have in turn had a profound impact
76 on our own societies. However, in many cases the ecological context and evolutionary
77 dynamics of adaptation to a human niche are poorly understood [1].

78

79 Anthrodependent or human-commensal taxa differ from domesticated species in that
80 humans do not play a direct role their reproduction, i.e. they do not experience artificial
81 selection [2,3]. Nonetheless, these taxa likely benefit from a close relationship with
82 humans. Archetypal examples include our most common rodents such as the house
83 mouse, black and brown rats that have spread with agriculture, colonialism and
84 urbanisation [4-6]. Such species were likely predisposed to utilise human resources as
85 opportunistic scavengers and subsequently adapted to a dependant relationship with
86 humans [7].The evolutionary origins of several domesticated species – i.e. dogs, cats and
87 cattle – are now reasonably well understood [8]; however, relatively little is known
88 about several charismatic and familiar species which may have had a long association
89 with human civilisation.

90

91 Understanding the origins of anthrodependent species is of particular interest since they
92 may act as bioproxies for our own history [9]. The distributions of human-commensals
93 are largely linked to human activity and so their evolutionary history should reflect
94 large-scale human movements. House mice populations of the Northern and Western
95 British Isles harbour an mtDNA lineage that is also present in Norway, suggesting mice
96 were likely transported to the regions as stowaways on Norwegian Viking ships [5].
97 Analysis of some pest species has also shed light on human evolution; divergence in lice
98 lineages reflect splits in the *Homo* genus [10] and genetic diversity in *Helicobacter pylori*
99 reflects human migrations from prehistory to the modern era [11].

100

101 The house sparrow (*Passer domesticus*) is a ubiquitous human-commensal bird species
102 occupying cities and farmland where it feeds on food waste and crops. Its native range
103 covers Western and Central Eurasia; however, due to deliberate and accidental
104 introductions by humans, its current distribution also encompasses Southern Africa,
105 Australia, New Zealand and the Americas. It is strongly associated with human
106 settlements with a clearly human associated ecology; the species is known to go locally
107 extinct in abandoned settlements [12,13]. The house sparrow is a well-studied model for
108 quantitative genetics [14,15] and also for adaptation to urban environments [16,17].
109 However, we know surprisingly little about the evolutionary history of this charismatic
110 companion species.

111

112 House sparrow human-commensalism is thought to have arisen once in the Middle and
113 Near East with the Neolithic revolution [18]. The species likely spread with the
114 subsequent introduction of agriculture and establishment of fixed settlements in Europe
115 [12]. This is also thought to have played an instrumental role in the hybrid origin of the
116 Italian sparrow (*P. italiae*) in the Mediterranean region as a result of admixture with the
117 Spanish sparrow (*P. hispaniolensis*), which was probably already present in Southern
118 Europe ([19-23], Fig 1A). Furthermore, genomic evidence suggesting multiple,
119 independent hybrid speciation events is consistent with a stepwise introduction of
120 agriculture via Mediterranean islands [24,25].

121

122 Intriguingly, a house sparrow subspecies - *P. d. bactrianus* (hereafter Bactrianus
123 sparrow) - occurs in the Middle East and Central Asian steppes (Fig 1A). Bactrianus
124 sparrows are quite unlike European house sparrows - they migrate, are less bold, and
125 are not associated with human settlements [12,18]. Furthermore, Bactrianus skull and
126 beak morphology are less robust compared to the more human-associated house
127 sparrow [26]. This difference is consistent with divergent foraging- i.e. Bactrianus
128 sparrows mainly feed on natural grass seeds, whereas the house sparrow is expected to
129 have adapted to feeding on tougher seeds from domesticated crops [26]. Taken together,
130 this suggests the Bactrianus sparrow represent a branch of the house sparrow lineage
131 that diverged prior to the evolution of human-commensalism. If this is the case, a

132 comparative approach using both European house and Bactrianus sparrows may shed
133 light on how the ubiquitous house sparrow became associated with humans.

134

135 Here we examine population genomic variation in three Eurasian *Passer* sparrow
136 species, the Spanish, Italian and house sparrow; we also include non-commensal house
137 sparrow lineage, the Bactrianus sparrow. We first investigate population genomic
138 structuring and evolutionary relationships among species, testing for evidence of
139 admixture across the European distribution. We then use coalescent modelling to infer
140 demographic history of the house and Spanish lineages, testing the split date between
141 the house and Bactrianus sparrow, when the house sparrow underwent a population
142 size expansion and whether this occurred in the face of gene flow from the Spanish
143 sparrow. Finally, we perform a comparative genome scan between the house and
144 Bactrianus sparrow, revealing strong signatures of divergent selection between these
145 two *P. domesticus* lineages that points to several intriguing candidate genes that may
146 underlie adaptation to a human niche.

147

148 **Methods**

149 *Sample collection*

150 We collected house ($n = 46$), Spanish ($n = 43$), Italian ($n = 31$) and Bactrianus ($n = 19$)
151 sparrows from across Eurasia (Fig 1A) using mist nets at each sampling location.
152 Captured individuals were recorded and measured (data not presented here) before a
153 blood sample was taken. See Table S1 for a breakdown of all samples and sampling
154 locations. Blood was immediately stored in Queen's Lysis Buffer for preservation and
155 individuals were released quickly, without harm, to minimise stress. In all cases,
156 sparrows were collected under appropriate permissions and guidelines.

157

158 *DNA extraction and sequencing*

159 High quality genomic DNA was extracted from blood samples using a Qiagen Blood &
160 Tissue DNEasy kit (Qiagen, California, USA). Extracted DNA was prepared for sequencing
161 using an Illumina TruSeq gDNA 180 bp kit (Illumina, California, USA). Sequencing was
162 conducted on Illumina Hi-Seq 2000 and Illumina Hi-Seq X machines at Genome Quebec,
163 McGill University, Canada.

164

165 *Read alignment, variant calling and filtering*

166 Raw reads were trimmed and filtered for Illumina adapters using Trimmomatic 0.36.
167 [27]. Base calls at the start and end of reads with a phred quality score of less than 5
168 were removed and any reads with an average quality of less than 10 across 5 bp sliding
169 windows were thrown out. Trimmed and filtered reads were then aligned to the house
170 sparrow reference genome [19] using bwa 0.7.10 [28]; both paired and unpaired reads
171 were mapped separately and then merged to produce a final bam for each individual.
172 Bams were then sorted, marked for duplicates and indexed using Picard 2.7.1
173 (<https://github.com/broadinstitute/picard>).

174

175 Bams were realigned around indels to prevent false positive variant detection; calls
176 were then made for all sites (variant and invariant) using the GATK (2.7.1)
177 HaplotypeCaller [29]. The raw vcf created by this pipeline was then filtered to remove
178 all indels and annotated with filter thresholds (see supplementary methods). Additional
179 filters were applied to create two, high-quality datasets for different downstream
180 analyses. The first dataset (hereafter 'variant only') included only polymorphic, biallelic
181 SNPs occurring in at least 80% of individuals, with minimum site and genotype quality
182 scores of 20, a mean site depth of between 10-40X. We additionally masked all
183 genotypes with a depth below 5x and above 60x. We applied three different minor allele
184 thresholds – no threshold (for demographic analyses), 0.02 and 0.05. The second dataset
185 was filtered for the same criteria but included calls at all sites – i.e. variant and invariant
186 – and with no MAF thresholds (hereafter 'all sites'). All filtering was conducted using
187 vcfutils 0.1.13 [30] and bcftools 1.1 [31] and scripts are available at
188 www.github.com/markravinet.

189

190 *Population structure analysis*

191 To investigate population structure, we performed linkage pruning on the MAF 0.05
192 filtered variant only dataset using PLINK 1.9 [32], filtering for all loci within 100 Kb
193 windows with an r^2 exceeding 0.1, the baseline for genome wide LD [19]. PCA was then
194 performed for autosomal SNPs only using PLINK 1.9. For a model-based analysis of
195 population structure on autosomal SNPs, we used admixture 1.3 [33], setting the
196 number of assumed populations (K) between 1-6 and using leave-one-out cross-

197 validation in order to determine the best supported value. Scripts for population
198 structure analyses are available at www.github.com/markravinet.

199

200 *Genome scans for signatures of selection and introgression*

201 In order to identify genomic regions under selection, we first estimated F_{ST} , a relative
202 measure of differentiation in 100 Kb sliding windows with a 25 Kb step using `vcftools`
203 from our MAF 0.02 filtered variant only dataset. We additionally estimated d_{XY} , an
204 absolute measure of divergence for the same windows using `popgenWindows.py`
205 (https://github.com/simonhmartin/genomics_general), [34] on our all sites dataset.

206

207 Since both F_{ST} and d_{XY} are prone to biases, particularly with regard to genome-wide
208 recombination rate variation, we also calculated long-range haplotype (LRH) estimates
209 that incorporate such information [35]. This approach requires phased haplotypes, so
210 we statistically phased the MAF = 0.02 threshold variant-only dataset. Bactrianus, house
211 and Spanish sparrow individuals were all phased individually with `Shapelt2` [36]. For all
212 autosomes, a previously published linkage map [19] was used to inform phasing – we
213 did not include the Z chromosome. LRH and extended haplotype homozygosity statistics
214 (specifically `iHS`, `xpEHH` and `EHHS`) were calculated using the R package `rehh` [37];
215 scripts for genome scans, phasing and file conversion are available at
216 www.github.com/markravinet.

217

218 Our initial population structure analyses revealed evidence of admixture between house
219 and Spanish lineages in Europe. We therefore investigated fine-scale patterns of
220 introgression within the house genome using a four population test to calculate f_d –i.e.
221 the proportion of introgressed sites within a genome window [34]. Using the test tree
222 topology (((Bactrianus, house), Spanish), Tree)), high levels of f_d therefore represent an
223 enrichment of shared sites between house and Spanish sparrows (see Fig 1D). We
224 calculated f_d three separate times using all Spanish sparrows, using ‘pure’ Spanish
225 sparrows and using Spanish sparrows showing evidence of admixture with European
226 house sparrows. Analyses were performed on 100 Kb sliding windows with a 25 Kb step
227 using `ABBABABAwindows.py` (https://github.com/simonhmartin/genomics_general)
228 [34] on our all sites dataset.

229

230 *Candidate gene identification and gene ontology analysis*

231 To identify candidate genes for adaptation to a human-commensal lifestyle, we first
232 identified outlier SNPs with a $\log_{10} P$ -value of 6 for xpEHH – i.e. SNPs exhibiting clear
233 divergent selection between house and Bactrianus sparrows. Using a custom R script
234 (available at www.github.com/markravinet) we then clustered outlier SNPs occurring
235 within 100 Kb of one another to produce a dataset of outlier regions occurring across
236 the genome. We then identified all known genes within 250 Kb of the peaks of these
237 outlier regions. Retaining only unique gene ids we then performed gene ontology
238 analysis using clueGO in Cytoscape 3.6.0 [38] using a human gene set, medium network
239 specificity, a right-sided hypergeometric test, Benjamini & Hochberg FDR correction
240 and a $P < 0.05$. In addition to GO analysis, we examined the identity of genes in outlier
241 regions manually.

242

243 *Demographic inference*

244 To shed light on the evolutionary origins of Eurasian *Passer* sparrow lineages, we
245 performed maximum likelihood based demographic inference using the site-frequency
246 spectrum with fastsimcoal2 [39]. Because of the complexities of modelling hybrid origin
247 and also the possibility that Italian sparrows have arisen from several independent
248 hybridisation events [24], we focused only on the house, Bactrianus and Spanish
249 populations. Spanish sparrows were split into two further subsets – one with evidence
250 of admixture with the house (Spanish admix) and one without (Spanish pure). We
251 derived the folded four-population multidimensional observed SFS using *easySFS.py*
252 (<https://github.com/isaacovercast/easySFS>) from a set of high quality autosomal SNPs
253 present in >95% of individuals with no MAF filtering. Since sample sizes varied between
254 each of the populations, we projected the SFS down to 30 chromosomes – i.e. 15 diploid
255 individuals per population.

256

257 We tested three main models of divergence between sparrow species – isolation,
258 migration and admixture (see Fig S5). For the isolation model, all species/populations
259 diverged in the absence of gene flow (Fig S5). Under this model, contemporary evidence
260 of admixture is due to incomplete lineage sorting. For the migration model,
261 species/populations diverged in the face of interspecific gene flow throughout their
262 divergence history (Fig S5). In the admixture model, interspecific gene flow occurs as a

263 result of a pulse of admixture following divergence. Since more realistic models of gene
264 flow can improve model performance [40], we also ran versions of the models with
265 intraspecific gene flow included (Fig S5).

266

267 For all models, we drew priors from a loguniform distribution (see Table S2 for a full
268 description of priors). Since we were not interested in divergence time between
269 admixed and pure Spanish lineages, we fixed this to 10 000 generations, assuming
270 population structuring as a result of postglacial range expansion; this also allowed us to
271 constrain admixture and population expansion events to having occurred within this
272 time. The split between *P. domesticus* and *P. hispaniolensis* was allowed to vary between
273 100 000 and 2 million generations. We also set priors so that house/Bactrianus
274 divergence occurred 10 000 to 100 000 generations in the past (see Table S2).

275

276 For each model, we performed 100 independent runs of 100,000 coalescent simulations
277 to estimate the maximum likelihood. We then performed model selection using the run
278 with the highest likelihood for each of the models. Following Meier et al., [40] we first
279 used AIC for model selection but we additionally assessed the likelihood distribution for
280 each model by calculating the likelihoods of 100 expected site frequency spectra,
281 derived from 1,000,000 coalescent simulations. We also derived 95% confidence
282 intervals for parameters estimated from our models using non-parametric block
283 bootstrapping [40]. To achieve this, we split the genome into 3,204 1 Mb windows and
284 resampled windows with replacement until the bootstrapped SFS was the same size as
285 the observed. We created 100 bootstrapped site-frequency spectra and then performed
286 10 independent runs of likelihood estimation on them. Estimates from each of the best
287 runs were then used to derive the 95% confidence intervals around all parameter
288 estimates from our focal model.

289

290 **Results**

291 *Population structure and differentiation*

292 After mapping to the house sparrow reference genome, calling and filtering variants, we
293 retained 21,930,880 SNPs. We used a subset of LD pruned 178,268 biallelic SNPs with
294 an MAF > 0.05 for parametric (ADMIXTURE) and non-parametric (PCA) inference of
295 population structure.

296

297 Species are clearly separated along the first principal component (38.3% PVE) with the
298 Italian sparrow occurring intermediate between Spanish sparrows and house sparrow,
299 consistent with hybrid origin (Fig 1B). Intriguingly, the Bactrianus sparrow is displaced
300 on both PC1 and PC2 (9.1% PVE), forming a separate cluster from European house
301 populations (Fig 1B). For Spanish, house and Italian sparrows, within species population
302 structure is also apparent from the PCA (Fig S1).

303

304 ADMIXTURE analysis strongly supported K values of 2 and 3 (Fig 1C, Fig S2). For $K = 2$,
305 Bactrianus and Spanish sparrows form clear, separate 'pure' clusters and the Italian
306 sparrow is clearly admixed, as expected for a hybrid species. However, European
307 populations of both house and Spanish sparrows show evidence of introgression, which
308 is also apparent from PC1 (Fig S1). For $K = 3$, house sparrows form a separate third
309 cluster but retain a signature of Spanish and Bactrianus ancestry (Fig 1C).

310

311 Mean genome-wide F_{ST} (\pm sd) estimates also support a Bactrianus lineage divergent from
312 both the house (0.103 ± 0.075) and Spanish sparrow (0.326 ± 0.213 ; see also Fig S3).
313 Furthermore, differentiation was lower between European house and putatively
314 admixed Spanish sparrows (0.166 ± 0.171) than pure Spanish populations from Italy and
315 Kazakhstan (0.234 ± 0.208 ; $P < 2.2 \times 10^{-16}$, permutation test: Fig S3). Absolute genomic
316 divergence between species, measured using d_{XY} , was similarly higher between
317 house/Bactrianus and the pure Spanish populations compared to the admixed ($P < 2.2 \times$
318 10^{-16} , permutation tests; see Fig S4A). Finally, f_d (\pm sd) was higher when P3 was set to the
319 Spanish admixed populations (0.38 ± 0.17 ; see Fig 1D) compared to the Spanish pure
320 (0.35 ± 0.16 ; see Fig S4B) – supporting admixture between Spanish and house sparrows
321 in Europe.

322

323 *Demographic inference*

324 Our demographic analysis clearly rejected scenarios where no gene flow has occurred
325 between the house and Spanish sparrow (Figs S5 & S6). The best-supported
326 demographic model using log-likelihood and AIC was one of divergence with migration
327 via secondary contact between the house and Spanish sparrow (Fig 2A, Table 1, Fig S6).
328 Under this model the two Passer species diverged 0.83 million years BP (0.69-0.93 mya

329 95% confidence intervals), assuming a single year for sparrow generation time.
330 Divergence between the Bactrianus and house sparrow was much more recent,
331 occurring 11.1 K years ago, whereas an expansion in the house lineage occurred 5657
332 generations (4292-6308 85% CI; Fig 2A & 2B) in the past. Migration rate estimates also
333 indicated introgression from the Spanish into the house was greater than gene flow in
334 the opposite direction or within species (Fig 2B).

335

336 *Divergent selection between house and Bactrianus sparrows*

337 Divergence between house and Bactrianus sparrows is consistent with the onset of the
338 Holocene; suggesting the latter is potentially a relict of the pre-human commensal house
339 ancestor. A comparative genome scan between these species will help identify selected
340 regions that may have played a role in adaptation to a human niche by the house
341 sparrow. The house sparrow genome shows clear signatures of strong selection on the
342 majority of autosomes (Fig S7). However, cross-population long-range haplotype
343 statistics also point to several regions throughout the genome that exhibit signatures of
344 divergent selection between the house and Bactrianus sparrow, including strikingly high
345 peaks on chromosomes 1 and 8 (Fig 3A, 3B & Fig 4).

346

347 To further investigate these signatures of selection, we identified 705 outlier SNPs
348 (0.06% of 1,033,861 total SNPs analysed) where the \log_{10} P -value of xpEHH between
349 house and Bactrianus was greater than 7 (i.e. points coloured red in Fig 4 and Fig S8).
350 We grouped outlier SNPs occurring within 100 Kb of one another together to identify
351 the peaks outlier regions across the genome (i.e. Fig 4). The majority of these peaks fell
352 on chromosomes 1, 2, 3 & 8; the last of which also harboured the SNPs with the highest
353 signatures of divergent selection (see Tables S3 & S4). Using the highest outlier SNPs on
354 each of these chromosomes, EHHS values show a clear pattern of increased extended
355 haplotype homozygosity in the house but not the Bactrianus sparrow (Fig 3B). This
356 suggests positive selective sweeps have occurred at these regions in the house sparrow
357 genome only.

358

359 Closer inspection of co-variation of other population genomic estimates alongside
360 xpEHH peaks showed little evidence of selection or introgression (Fig 4 and Fig S8).

361 However, in some cases, regions of increased F_{ST} corresponded to extended haplotype
362 homozygosity and slight peaks of f_d (Fig 4).

363

364 *Candidate gene and gene ontology analysis*

365 We identified 153 unique genes falling within 250 Kb of xpEHH outlier peaks from
366 across the genome. GO analysis identified 20 gene pathways with evidence of
367 enrichment among the outlier gene set (see Table S5). These included cartilage
368 condensation ($P = 0.01$) and regulation of circadian rhythm ($P = 0.085$ after FDR
369 correction). A gene of interest from the cartilage condensation pathway is *wnt7a* on
370 chromosome 12; this has previously been linked to feather development and
371 melanogenesis in birds [41,42]. Additionally, a *PARL* transcript (presenilin associated
372 rhomboid-like) on chromosome 9 is upregulated during migration in white-crowned
373 sparrows (*Zonotrichia leucophrys*)[43].

374

375 One of the highest xpEHH peaks in the genome occurred on chromosome 8 between
376 19.01-19.27 Mb (Fig 3B, Fig 4). This peak contains two known genes, *COL11A* and
377 *AMY2A*, Fig 4). *COL11A* – collagen type XII alpha – is associated with craniofacial
378 development; mutations in this gene for humans result in Marshall's Syndrome which is
379 characterised by increased skull thickness and altered facial structure [44]. *AMY2A* –
380 amylase alpha2 is part of the amylase gene family associated with adaptation to a higher
381 starch diet in both humans and dogs [45,46].

382

383 **Discussion**

384

385 Our findings show that a high level of genomic divergence matches the phenotypic,
386 behavioural and ecological differences between the European house and Bactrianus
387 sparrows. Pairwise mean F_{ST} between the house and Bactrianus subspecies is half of
388 that between the house and Spanish sparrow (Fig S3). High differentiation is unlikely to
389 be a factor of distance between the house and Bactrianus; Spanish sparrow populations
390 are similarly spatially isolated but show no evidence of population structuring. Instead,
391 high divergence between the house lineages is likely due to the fact that they split some
392 time ago. Our coalescent analyses show that while this was not ancient, occurring 11

393 Kya, divergence may have occurred prior to the widespread dissemination of agriculture
394 and the evolution of commensalism in the house sparrow.

395

396 Reanalysis of Eurasian sparrow genomic data alongside the Bactrianus has developed
397 our understanding of population genomic structuring among these species. Admixture
398 between the house and Spanish sparrow has played an important role in the hybrid
399 origin of the Italian sparrow. We show that European house populations have
400 experienced some level of Spanish admixture and that conversely, certain Spanish
401 populations harbour house introgression, albeit at a lower level. What has driven gene
402 exchange between these two species remains unclear. Our demographic inference
403 suggests that gene flow has been ongoing since secondary contact, and since Spanish
404 sparrows were already likely present in Southern Europe at the time the house sparrow
405 was introduced with agriculture, this is a likely explanation. However, gene flow is still
406 on-going and there are several parts of the present-day distribution where house x
407 Spanish hybrids occur. This is also consistent with the tentative observation that the
408 proportion of Spanish ancestry in the house appears to be lower in more northerly
409 populations sampled in Norway, compared to those in France ($K = 3$, Fig 1C).
410 Nonetheless, it is also possible that introgression is in fact quite old and that it largely
411 occurred in the native range, where the Spanish sparrow is also present, prior to the
412 expansion of the house into Europe. Concordance between signatures of divergent
413 selection between the house and Bactrianus and measures of Spanish introgression into
414 the house is consistent this explanation. Most likely, Spanish ancestry in the house
415 genome is due to both mechanisms, but more work is now necessary to distinguish them
416 and quantify their relative importance.

417

418 The Bactrianus sparrow may have diverged from the main house lineage prior to, or as a
419 result of the evolution of human-commensalism during the Neolithic revolution. The
420 subspecies migrates, does not associate with human settlements and is less bold [12] –
421 all traits which are common to non-human commensal sparrow species such as the
422 Spanish. Bactrianus sparrows can therefore be considered a proxy for the ancestral, pre-
423 commensal house sparrow. With the rapid expansion and spread of house sparrow
424 populations following the invention of agriculture, it is likely the Bactrianus has
425 remained a relict lineage confined to the Central Asian steppes. This is also supported by

426 evidence of limited interbreeding between co-occurring house and Bactrianus sparrows
427 in Kazakhstan [47,48]. This offers us a unique opportunity to look for signatures of
428 strong selection that have occurred in the house lineage and are absent from the
429 Bactrianus – i.e. a comparative genome scan for adaptation to a human niche.

430

431 Our comparative approach does indeed reveal multiple regions throughout the sparrow
432 genome that show signatures of house-specific positive selective sweeps. Importantly,
433 our use of long-range haplotype tests that incorporate information on linkage
434 disequilibrium and haplotype homozygosity are able to identify strong patterns of
435 selection that are not apparent using more standard measures such as F_{ST} or d_{XY}
436 [35,37,49]. The sweep regions we detected harbour a number of candidate genes
437 consistent with many of the phenotypic traits that are known to be divergent between
438 house and non-commensal Bactrianus sparrows, such as plumage, migratory behaviour,
439 dietary differences and skull morphology. An xpEHH peak on chromosome 12 is in close
440 proximity to *wnt7a*, a gene that has previously shown to be involved in feather
441 development and melanogenesis in birds [41,42]. This gene is also under apparent
442 divergent selection between island populations of the Italian sparrow which differ in the
443 darkness of their plumage [24]. Bactrianus sparrows too have a darker plumage than
444 the house. A further peak on chromosome 9 harbours the *PARL* gene, which shows
445 increased expression in the brains of white-crowned sparrows (*Zonotrichia leucophrys*)
446 during the migratory season[43]. This is also bolstered by our finding of regulation of
447 circadian rhythm as an enriched GO pathway; importantly, Bactrianus sparrows migrate
448 whereas house sparrows do not [12,18].

449

450 One of the most striking peaks of divergent selection sits on chromosome 8 and contains
451 the *COL11A* and *AMY2A* genes (Fig 4). *COL11A* is closely associated with craniofacial
452 development in humans and is linked to Marshall's Syndrome, a genetic disorder which
453 leads to increased skull thickness and abnormal facial structure [44]. This is particularly
454 interesting as skull morphology and craniofacial structure has been shown to differ
455 between Bactrianus and house sparrows, with the latter exhibiting a more robust skull
456 morphology and larger beak [26]. Craniofacial differences are further supported by an
457 enrichment of cartilage condensation in our GO analysis. The shift in skull and beak
458 morphology between sparrow lineages is commonly attributed to the dietary shift from

459 natural seeds to agricultural food waste during the development of commensalism [26].
460 Intriguingly, *AMY2A*, is part of a family of amylase genes that have been linked to the
461 transition to starch-based diets both humans and dogs during the Neolithic revolution
462 [45,50]. In dogs, increased copy number of the closely related *AMY2B* gene is consistent
463 with the spread of agriculture during this period [46,51]. Among human populations,
464 *AMY1* copy number is associated with dietary starch content and the frequency of
465 *AMY2A* deletions is higher in groups with a low starch diet [45,52]. Our findings
466 therefore add to the emerging picture that the Neolithic revolution introduced a
467 common selective pressure that has resulted in parallel adaptations in similar genes for
468 three very different taxa – humans, dogs and potentially, house sparrows.

469

470 At present, it is not clear whether *AMY2A* or *COL11A* or both genes are the target of
471 selection at this region of the genome. However, since both genes occur just 154 Kb from
472 one another, there is a strong possibility they remain in linkage disequilibrium as a co-
473 adapted gene complex for a human-commensal diet. It is now necessary to investigate
474 whether this is the case and to clearly link these genes to their putative role as
475 adaptations to a human niche in house sparrows. Furthermore, determining the age of
476 the selective sweep and testing whether selection is also apparent in the Italian sparrow
477 is now necessary to conclusively link this adaptation to the onset of the Neolithic
478 revolution. Nonetheless, our current findings place the origins of commensalism in
479 house sparrows in an evolutionary context and show that understanding how this
480 species came to be is informative for our understanding of our own recent evolutionary
481 history.

482

483 **References**

- 484 1. Hulme-Beaman, A., Dobney, K., Cucchi, T. & Searle, J. B. 2016 An Ecological and
485 Evolutionary Framework for Commensalism in Anthropogenic Environments. **31**,
486 1–13. (doi:10.1016/j.tree.2016.05.001)
- 487 2. Vigne, J.-D. 2015 Early domestication and farming: what should we know or do for
488 a better understanding? *Anthropozoologica* **50**, 123–150.
489 (doi:10.5252/az2015n2a5)
- 490 3. Zeder, M. A. 2015 Core questions in domestication research. *Proc Natl Acad Sci*
491 *USA* **112**, 3191–3198. (doi:10.1073/pnas.1501711112)

- 492 4. Combs, M., Puckett, E. E., Richardson, J., Mims, D. & Munshi-South, J. 2017 Spatial
493 population genomics of the brown rat (*Rattus norvegicus*) in New York City. **224**,
494 335–16. (doi:10.1111/mec.14437)
- 495 5. Searle, J. B. et al. 2009 Of mice and (Viking?) men: phylogeography of British and
496 Irish house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**, 201–
497 207. (doi:10.1098/rspb.2008.0958)
- 498 6. Gray, M. M., Wegmann, D., Haasl, R. J., White, M. A., Gabriel, S. I., Searle, J. B.,
499 Cuthbert, R. J., Ryan, P. G. & Payseur, B. A. 2014 Demographic history of a recent
500 invasion of house mice on the isolated Island of Gough. *Mol Ecol* **23**, 1923–1939.
501 (doi:10.1111/mec.12715)
- 502 7. Johnson, M. T. J. & Munshi-South, J. 2017 Evolution of life in urban environments.
503 *Science* **358**, eaam8327. (doi:10.1126/science.aam8327)
- 504 8. Larson, G. & Fuller, D. Q. 2014 The Evolution of Animal Domestication. *Annu. Rev.*
505 *Ecol. Evol. Syst.* **45**, 115–136. (doi:10.1146/annurev-ecolsys-110512-135813)
- 506 9. Jones, E. P., Eager, H. M., Gabriel, S. I., Jóhannesdóttir, F. & Searle, J. B. 2013 Genetic
507 tracking of mice and other bioproxies to infer human history. *Trends in Genetics*
508 **29**, 298–308. (doi:10.1016/j.tig.2012.11.011)
- 509 10. Reed, D. L., Smith, V. S., Hammond, S. L., Rogers, A. R. & Clayton, D. H. 2004 Genetic
510 Analysis of Lice Supports Direct Contact between Modern and Archaic Humans.
511 *PLoS Biol* **2**, e340–12. (doi:10.1371/journal.pbio.0020340)
- 512 11. Falush, D. et al. 2003 Traces of Human Migrations in *Helicobacter pylori*
513 Populations. *Science* **299**, 1582–1585. (doi:10.1126/science.1080857)
- 514 12. Summers-Smith, D. 1988 *The Sparrows*. Calton, UK: T & AD Poyser.
- 515 13. Anderson, T. R. 2006 *Biology of the Ubiquitous House Sparrow*. Oxford University
516 Press.
- 517 14. Jensen, H., Sæther, B. E., Ringsby, T. H., Tufto, J., Griffith, S. C. & Ellegren, H. 2003
518 Sexual variation in heritability and genetic correlations of morphological traits in
519 house sparrow (*Passer domesticus*). *J. Evol. Biol.* **16**, 1296–1307.
520 (doi:10.1046/j.1420-9101.2003.00614.x)
- 521 15. Ben Cohen, S. & Dor, R. 2018 Phenotypic divergence despite low genetic
522 differentiation in house sparrow populations. *Sci Rep* **8**, 394.
523 (doi:10.1038/s41598-017-18718-8)
- 524 16. Shaw, L. M., Chamberlain, D. & Evans, M. 2008 The House Sparrow *Passer*
525 *domesticus* in urban areas: reviewing a possible link between post-decline
526 distribution and human socioeconomic status. *J Ornithol* **149**, 293–299.
527 (doi:10.1007/s10336-008-0285-y)

- 528 17. Swaileh, K. M. & Sansur, R. 2006 Monitoring urban heavy metal pollution using the
529 House Sparrow (*Passer domesticus*). *Journal of Environmental Monitoring* **8**, 209–
530 213. (doi:10.1039/B510635D)
- 531 18. Sætre, G. P. et al. 2012 Single origin of human commensalism in the house
532 sparrow. *J. Evol. Biol.* **25**, 788–796. (doi:10.1111/j.1420-9101.2012.02470.x)
- 533 19. Elvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J.,
534 Ravinet, M., Jensen, H. & Sætre, G.-P. 2017 The genomic mosaicism of hybrid
535 speciation. *Science Advances* **3**, e1602996. (doi:10.1126/sciadv.1602996)
- 536 20. Elgvin, T. O., Hermansen, J. S., Fijarczyk, A., Bonnet, T., Norge, T., Saether, S. A.,
537 Voje, E, K. L. & Sætre, G. P. 2011 Hybrid speciation in sparrows II: a role for sex
538 chromosomes? **20**, 3823–3837. (doi:10.1111/j.1365-294X.2011.05182.x)
- 539 21. Hermansen, J. S., Saether, S. A., Elgvin, T. O., Borge, T., Hjelle, E. & Sætre, G.-P. 2011
540 Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and
541 barriers to gene flow. *Mol Ecol* **20**, 3812–3822. (doi:10.1111/j.1365-
542 294X.2011.05183.x)
- 543 22. Hermansen, J. S., Haas, F., Trier, C. N., Bailey, R. I., Nederbragt, A. J., Marzal, A. &
544 Sætre, G.-P. 2014 Hybrid speciation through sorting of parental incompatibilities
545 in Italian sparrows. **23**, 5831–5842. (doi:10.1111/mec.12910)
- 546 23. Trier, C. N., Hermansen, J. S., Sætre, G.-P. & Bailey, R. I. 2014 Evidence for Mito-
547 Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian
548 Sparrow and Its Parent Species. *PLoS Genet* **10**, e1004075–10.
549 (doi:10.1371/journal.pgen.1004075)
- 550 24. Runemark, A., Trier, C. N., Eroukhanoff, F., Hermansen, J. S., Matschiner, M.,
551 Ravinet, M., Elgvin, T. O. & Sætre, G.-P. 2018 Variation and constraints in hybrid
552 genome formation. *Nat. ecol. evol.* **20**, 229. (doi:10.1038/s41559-017-0437-7)
- 553 25. Paschou, P. et al. 2014 Maritime route of colonization of Europe. *Proc Natl Acad Sci*
554 *USA* **111**, 9211–9216. (doi:10.1073/pnas.1320811111)
- 555 26. Riyahi, S., Hammer, Ø., Arbabi, T., Sánchez, A., Roselaar, C. S. & Aliabadian, M. 2013
556 Beak and skull shapes of human commensal and non-commensal house sparrows
557 *Passer domesticus*. *BMC Evolutionary Biology* **13**, 200. (doi:10.1186/1471-2148-
558 13-200)
- 559 27. Bolger, A. M., Lohse, M. & Usadel, B. 2014 Trimmomatic: a flexible trimmer for
560 Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
561 (doi:10.1093/bioinformatics/btu170)
- 562 28. Li, H. & Durbin, R. 2009 Fast and accurate short read alignment with Burrows-
563 Wheeler transform. *Bioinformatics* **25**, 1754–1760.
564 (doi:10.1093/bioinformatics/btp324)

- 565 29. DePristo, M. A. et al. 2011 A framework for variation discovery and genotyping
566 using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498.
567 (doi:10.1038/ng.806)
- 568 30. Danecek, P. et al. 2011 The variant call format and VCFtools. *Bioinformatics* **27**,
569 2156–2158. (doi:10.1093/bioinformatics/btr330)
- 570 31. Danecek, P. & McCarthy, S. A. 2017 BCFtools/csq: haplotype-aware variant
571 consequences. *Bioinformatics* **33**, 2037–2039.
572 (doi:10.1093/bioinformatics/btx100)
- 573 32. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. 2015
574 Second-generation PLINK: rising to the challenge of larger and richer datasets.
575 *GigaScience* 2015 4:1 **4**, 7. (doi:10.1186/s13742-015-0047-8)
- 576 33. Alexander, D. H., Novembre, J. & Lange, K. 2009 Fast model-based estimation of
577 ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664.
578 (doi:10.1101/gr.094052.109)
- 579 34. Martin, S. H., Davey, J. W. & Jiggins, C. D. 2014 Evaluating the Use of ABBA–BABA
580 Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution* **32**, 244–
581 257. (doi:10.1093/molbev/msu269)
- 582 35. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. 2013. Detecting natural selection in
583 genomic data. *Annu. Rev. Genet.* **47**, 97–120. (doi:10.1146/annurev-ge-47-120213-
584 200001)
- 585 36. O'Connell, J. et al. 2014 A General Approach for Haplotype Phasing across the Full
586 Spectrum of Relatedness. *PLoS Genet* **10**, e1004234–21.
587 (doi:10.1371/journal.pgen.1004234)
- 588 37. Gautier, M., Klassmann, A. & Vitalis, R. 2016 rehh2.0: a reimplement of the R
589 package rehh to detect positive selection from haplotype structure. *Mol Ecol*
590 *Resour* **17**, 78–90. (doi:10.1111/1755-0998.12634)
- 591 38. Bindea, G. et al. 2009 ClueGO: a Cytoscape plug-in to decipher functionally
592 grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**,
593 1091–1093. (doi:10.1093/bioinformatics/btp101)
- 594 39. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. 2013 Robust
595 Demographic Inference from Genomic and SNP Data. *PLoS Genet* **9**, e1003905–17.
596 (doi:10.1371/journal.pgen.1003905)
- 597 40. Meier, J. I., Sousa, V. C., Marques, D. A., Selz, O. M., Wagner, C. E., Excoffier, L. &
598 Seehausen, O. 2016 Demographic modelling with whole-genome data reveals
599 parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol Ecol*
600 **26**, 123–141. (doi:10.1111/mec.13838)
- 601 41. Guo, H., Xing, Y., Liu, Y., Luo, Y., Deng, F., Yang, T., Yang, K. & Li, Y. 2016 Wnt/ β -
602 catenin signaling pathway activates melanocyte stem cells in vitro and in vivo.

- 603 *Journal of Dermatological Science* **83**, 45–51.
604 (doi:10.1016/j.jdermsci.2016.04.005)
- 605 42. S Noramly, A. F. A. B. A. M. 1999 -catenin signaling can initiate feather bud
606 development. *Development* **126**, 3509–3521.
- 607 43. Jones, S., Pfister-Genskow, M., Cirelli, C. & Benca, R. M. 2008 Changes in brain gene
608 expression during migration in the white-crowned sparrow. *Brain Research*
609 *Bulletin* **76**, 536–544. (doi:10.1016/j.brainresbull.2008.03.008)
- 610 44. Griffith, A. J., Sprunger, L. K., Sirko-Osadsa, D. A., Tiller, G. E., Meisler, M. H. &
611 Warman, M. L. 1998 Marshall Syndrome Associated with a Splicing Defect at the
612 COL11A1 Locus. *The American Journal of Human Genetics* **62**, 816–823.
613 (doi:10.1086/301789)
- 614 45. Perry, G. H. et al. 2007 Diet and the evolution of human amylase gene copy
615 number variation. *Nat Genet* **39**, 1256–1260. (doi:10.1038/ng2123)
- 616 46. Reiter, T., Jagoda, E. & Capellini, T. D. 2016 Dietary Variation and Evolution of
617 Gene Copy Number among Dog Breeds. *PLoS ONE* **11**, e0148899–19.
618 (doi:10.1371/journal.pone.0148899)
- 619 47. Gavrilov, E. I. & Korelov, M. N. 1968 The Indian sparrow as a distinct good species.
620 *Byulleten Moskovskogo Obshchestva Ispytateley Prirody Otdel Biologicheskiiy* **73**,
621 115–122.
- 622 48. Yakobi, V. E. 1978 On the species independence of the Indian sparrow. *Zoological*
623 *Zhurnal* **58**, 136–137.
- 624 49. Vatsiou, A. I., Bazin, E. & Gaggiotti, O. E. 2015 Detection of selective sweeps in
625 structured populations: a comparison of recent methods. *Mol Ecol* **25**, 89–103.
626 (doi:10.1111/mec.13360)
- 627 50. Axelsson, E. et al. 2013 The genomic signature of dog domestication reveals
628 adaptation to a starch-rich diet. *Nature* **495**, 360–364.
629 (doi:10.1038/nature11837)
- 630 51. Arendt, M., Cairns, K. M. & Ballard, J. W. O. 2016 Diet adaptation in dog reflects
631 spread of prehistoric agriculture. *Heredity* **117**, 301–306.
632 (doi:10.1038/hdy.2016.48)
- 633 52. Inchley, C. E. et al. 2016 Selective sweep on human amylase genes postdates the
634 split with Neanderthals. *Sci Rep* **6**, 1–10. (doi:10.1038/srep37198)

635

636 **Tables**

637 Table 1: Model selection for SFS based demographic analyses.

Model	Gene flow	LogLikelihood	N _{Params}	AIC	ΔAIC	ΔlogLik
-------	-----------	---------------	---------------------	-----	------	---------

Migration	intraspecific	-80571,91	15	161173,81	0,00	9604,36
Admixture	intraspecific	-83288,77	17	166611,54	5437,72	12321,22
Isolation	intraspecific	-86388,44	13	172802,88	11629,07	15420,90
Migration	none	-80730,16	13	161486,32	312,51	9762,62
Admixture	none	-82412,51	15	164855,01	3681,20	11444,96
Isolation	none	-85949,12	11	171920,24	10746,42	14981,57

638

639

640 **Figures**

641 Figure 1: A) Sample sites for Eurasian sparrow species, point colours correspond to
 642 species shown in B. B) principle component analysis of high-quality, linkage pruned
 643 SNPs separate species along PC1. C) Population clustering using ADMIXTURE for two
 644 best supported values of K . D) Schematic of the four-population test used in order to
 645 calculate genome-window measures of introgression (f_a) between the house and
 646 Spanish sparrow.

647

648 Figure 2: Site-frequency-spectrum based coalescent analyses best support a model of
 649 divergence in isolation and then migration between the house and Spanish sparrow
 650 upon secondary contact (A). Parameter estimates and their 95% CI, derived from non-
 651 parametric block bootstraps.

652

653 Figure 3: Signatures of divergent selection between house and bactrianus sparrows. (A)
 654 Manhattan plot of xpEHH shows clear peaks of divergent haplotype homozygosity
 655 occurring throughout the genome. N.B: the full dataset has been downsampled to 20% of
 656 the original data to aid visualisation. (B) Examination of the four snps with highest
 657 xpEHH scores shows a clear signature of extended haplotype homozygosity in the house
 658 but not the bactrianus sparrow.

659

660 Figure 4: A closer look at genomic divergence between house and bactrianus sparrows
 661 along chromosome 8; top panel – $\log_{10}(p)$ xpEHH (blue = background, red = outliers
 662 where $P < 0.0001$); second panel – mean absolute nucleotide divergence (d_{XY}); third
 663 panel – relative differentiation (F_{ST}); fourth panel – proportion of putatively
 664 introgressed per 100 Kb window (f_a).

665

666 **Supplementary material**

667 Table S1: Breakdown of samples used.

668 Table S2: Priors for demographic inference.

669 Table S3: Number of xpEHH outlier regions per chromosome.

670 Table S4: Table of highest xpEHH outlier peaks per chromosome.

671 Table S5: Enriched GO terms for genes within 250 Kb of xpEHH peaks between the
672 house and Bactrianus sparrow.

673

674 Figure S1: PCA showing within-species population structuring.

675 Figure S2: Cross-validation of ADMIXTURE runs showing highest support for $K = 2$ & 3.676 Figure S3: Boxplots of pairwise genome-wide F_{ST} estimates from 100 Kb sliding
677 windows with a 25 Kb step.678 Figure S4: Boxplots of pairwise genome-wide d_{XY} (a) and f_d (b) estimates from 100 Kb
679 sliding windows with a 25 Kb step.

680 Figure S5: Demographic models tested using site-frequency spectrum methods.

681 Figure S6: Log-likelihood distributions for demographic models.

682 Figure S7: Manhattan plot of iHS across the house sparrow genome. The full dataset has
683 been downsampled to 20% (176, 701 SNPs) in order to aid visualisation.684 Figure S8: Closer examination of genomic divergence between house and Bactrianus
685 sparrows for chromosomes 1, 2, 3 & 8.

686

687 **Acknowledgments**

688 We are grateful to Daniel Wegmann for his advice on designing demographic models
689 and Joana Meier for her assistance with running SFS analysis. We thank Camilla Sætre,
690 Juan Carlos Illera Cobo, Sonia Ramos, Caroline Øien Guldvog, Jo Skeie Hermansen, Stein
691 Are Sæther, Richard Bailey, Fabrice Eroukhmanhoff, Anna Runemark, Sepand Riyahi,
692 Almat Abayev and Syrymgul Zaripova for assistance with fieldwork. We are grateful to
693 Fabrice Eroukhmanhoff for his comments on a draft version of this manuscript. Data
694 analysis was conducted on the UiO Abel cluster [Norwegian Metacenter for High
695 Performance Computing (NOTUR) and the University of Oslo] operated by the Research
696 Computing Services group at The University Center for Information Technology

697 (www.hpc.uio.no/)and the DDBJ Supercomputer at the National Institute of Genetics,
698 Mishima, Japan.

699

700 **Data availability statement**

701 Scripts for analyses are available at www.github.com/markravinnet and Dryad. Datasets
702 for manuscript will be deposited on Dryad following manuscript decision. Sequence data
703 will be deposited on the European Nucleotide archive.

704

705 **Author contributions**

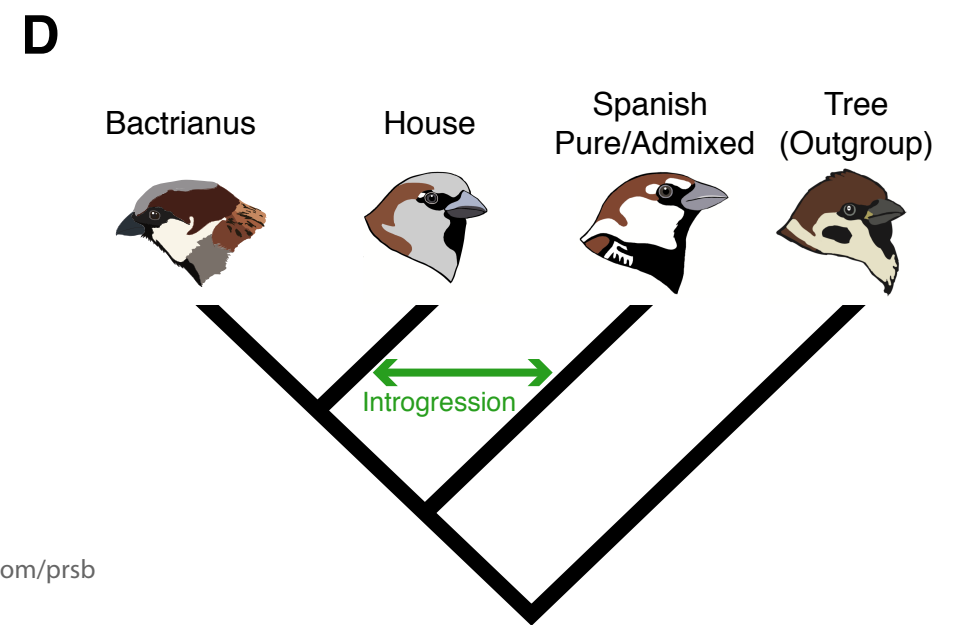
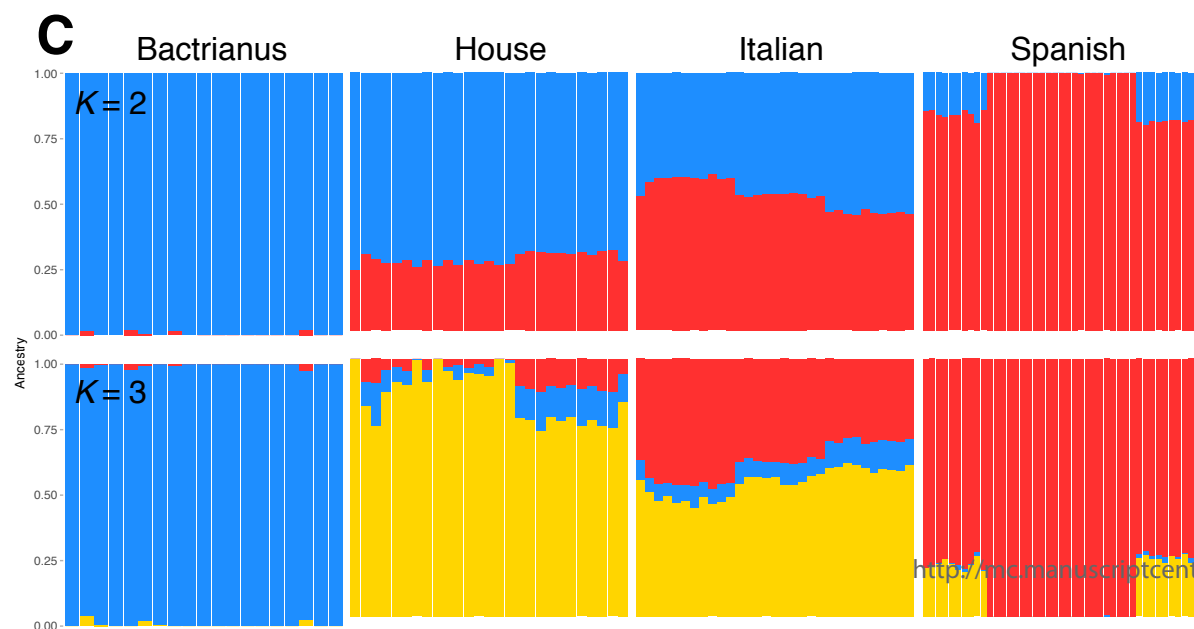
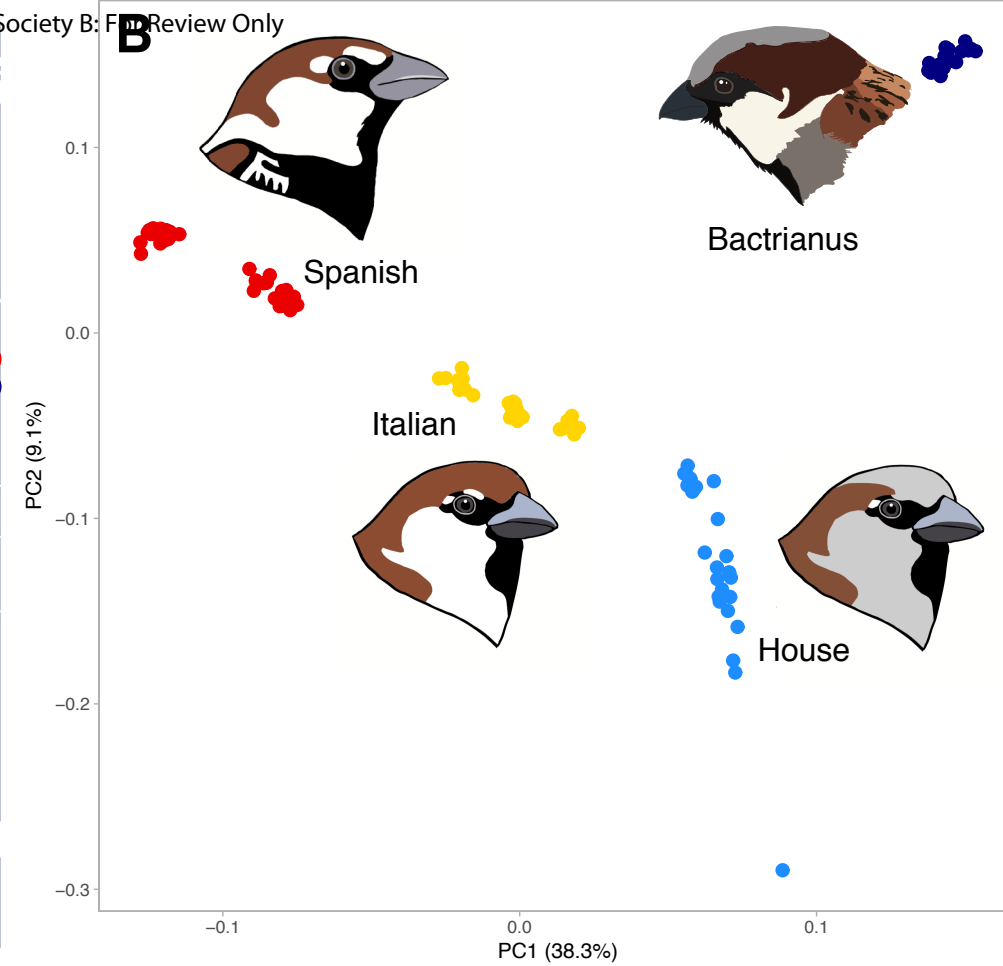
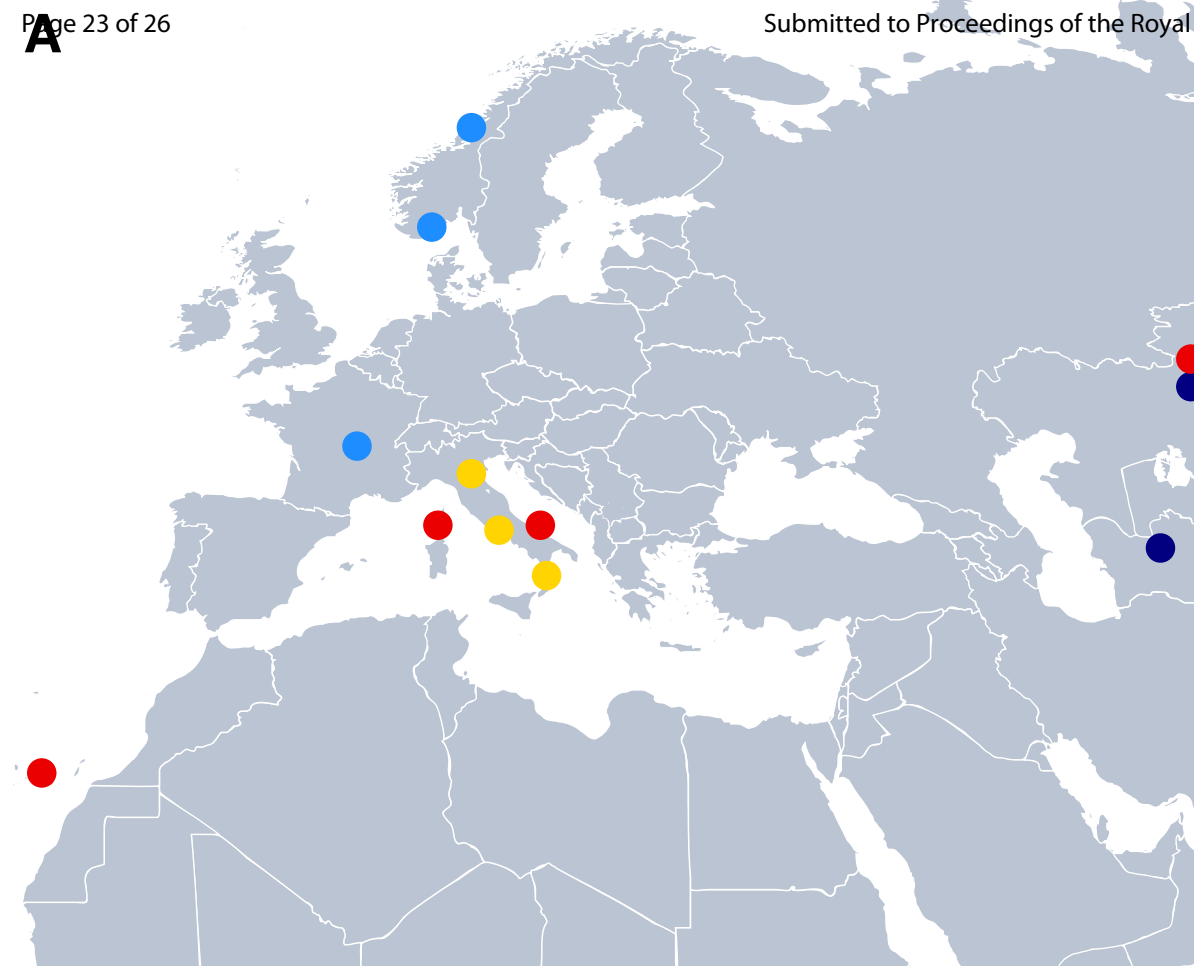
706 MR, TOE and GPS designed the study. TOE, CNT and MR conducted fieldwork and
707 prepared samples for sequencing. GPS, MA & AG organised sampling and conducted
708 fieldwork. MR analysed the data. MR & TOE wrote the manuscript. All authors
709 commented on the manuscript and gave final approval for publication.

710

711 **Funding information**

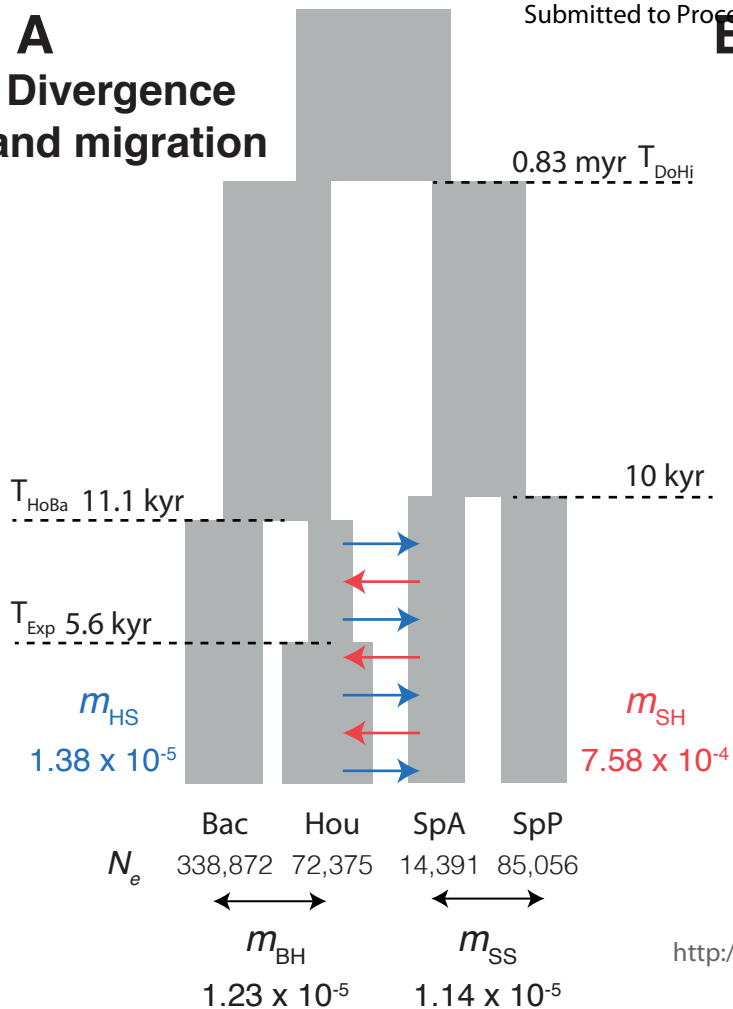
712 This study was supported grant to GPS from the Research Council of Norway.

713

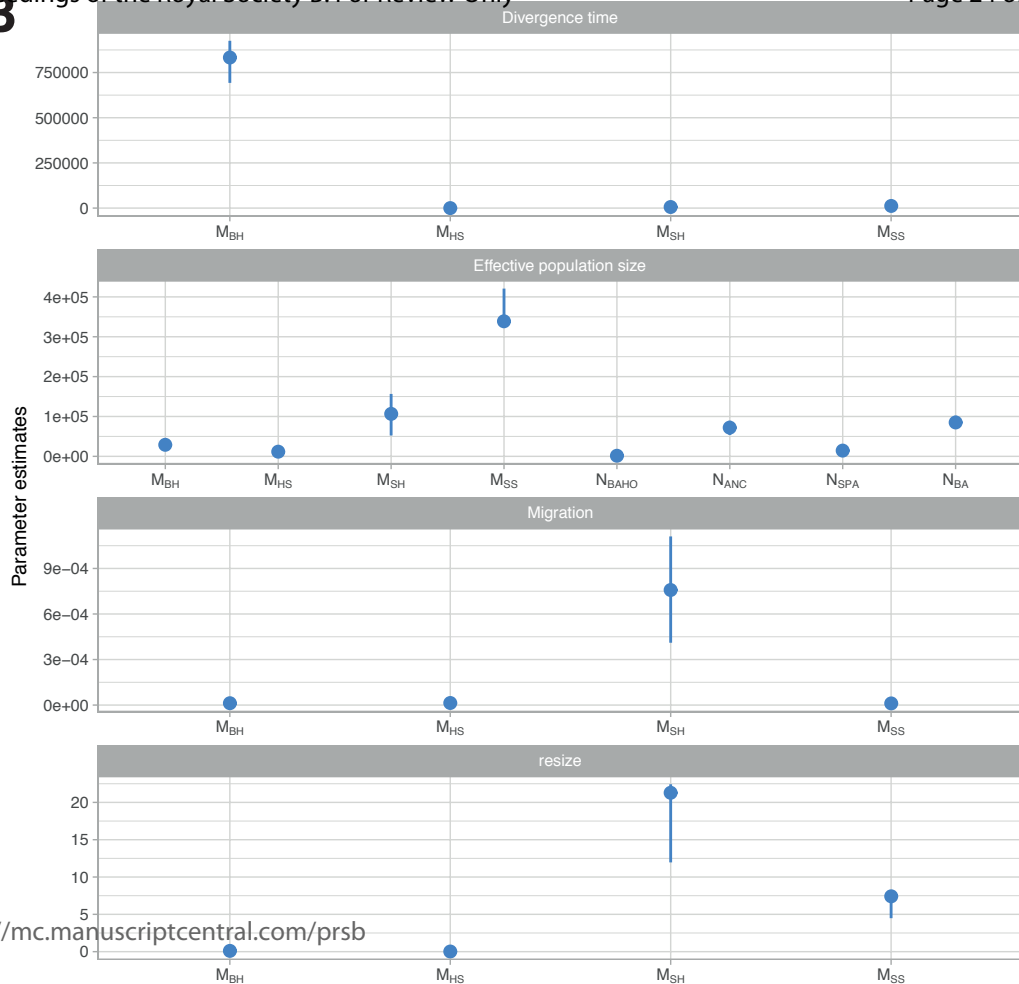


A

Divergence and migration



B



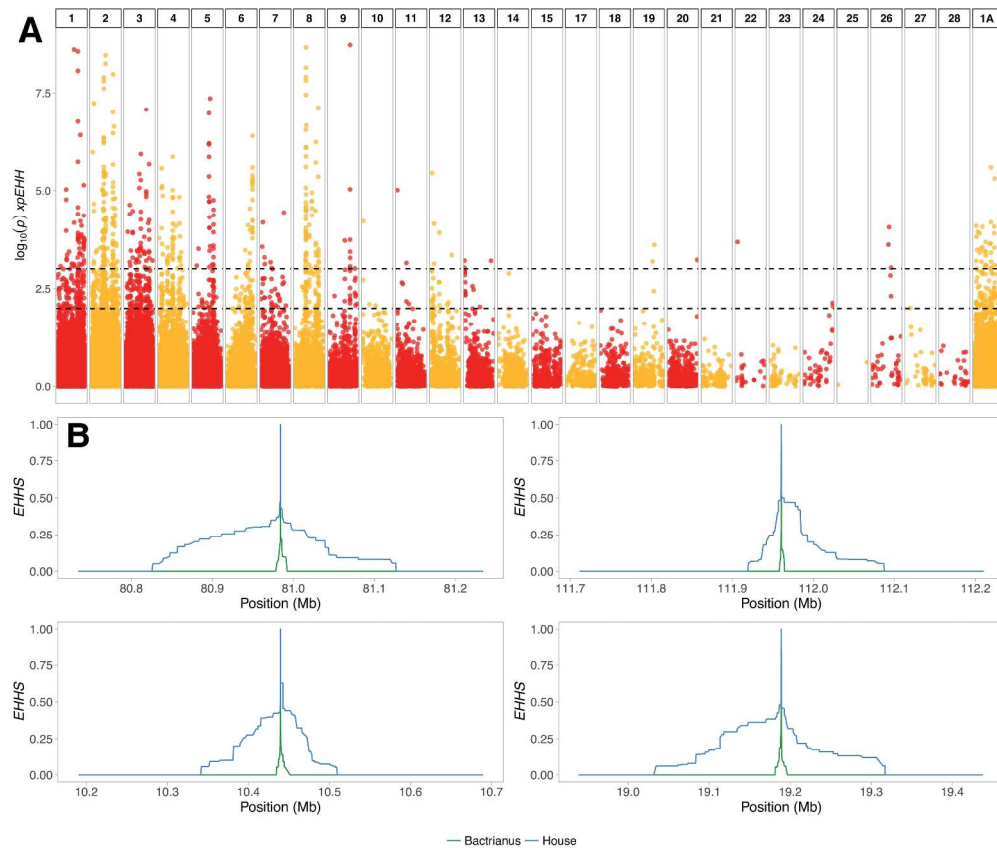


Figure 3: Signatures of divergent selection between house and bactrianus sparrows. (A) Manhattan plot of xpEHH shows clear peaks of divergent haplotype homozygosity occurring throughout the genome. N.B: the full dataset has been downsampled to 20% of the original data to aid visualisation. (B) Examination of the four snps with highest xpEHH scores shows a clear signature of extended haplotype homozygosity in the house but not the bactrianus sparrow.

302x257mm (300 x 300 DPI)

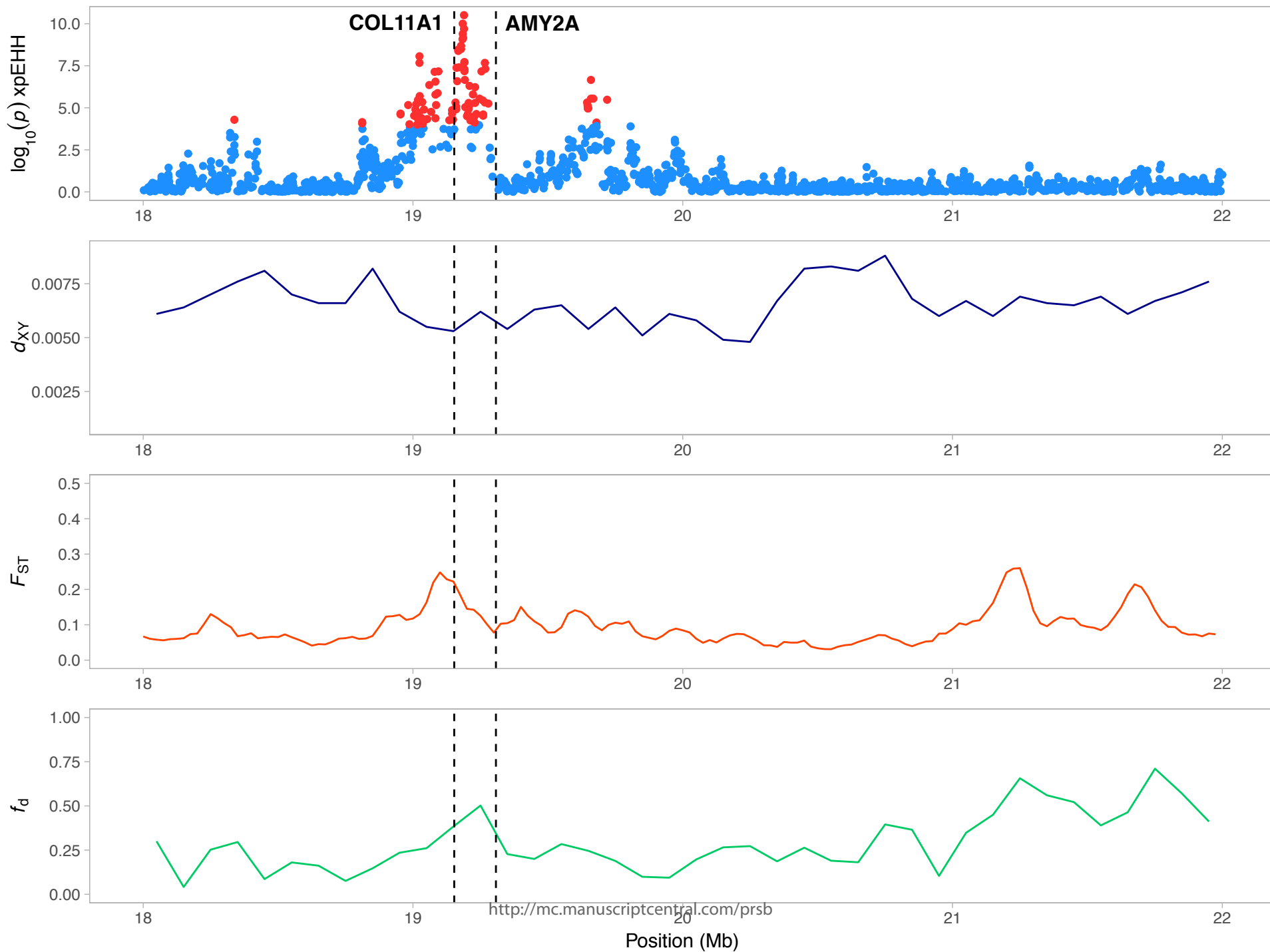


Figure S1: PCA showing within-species population structuring.

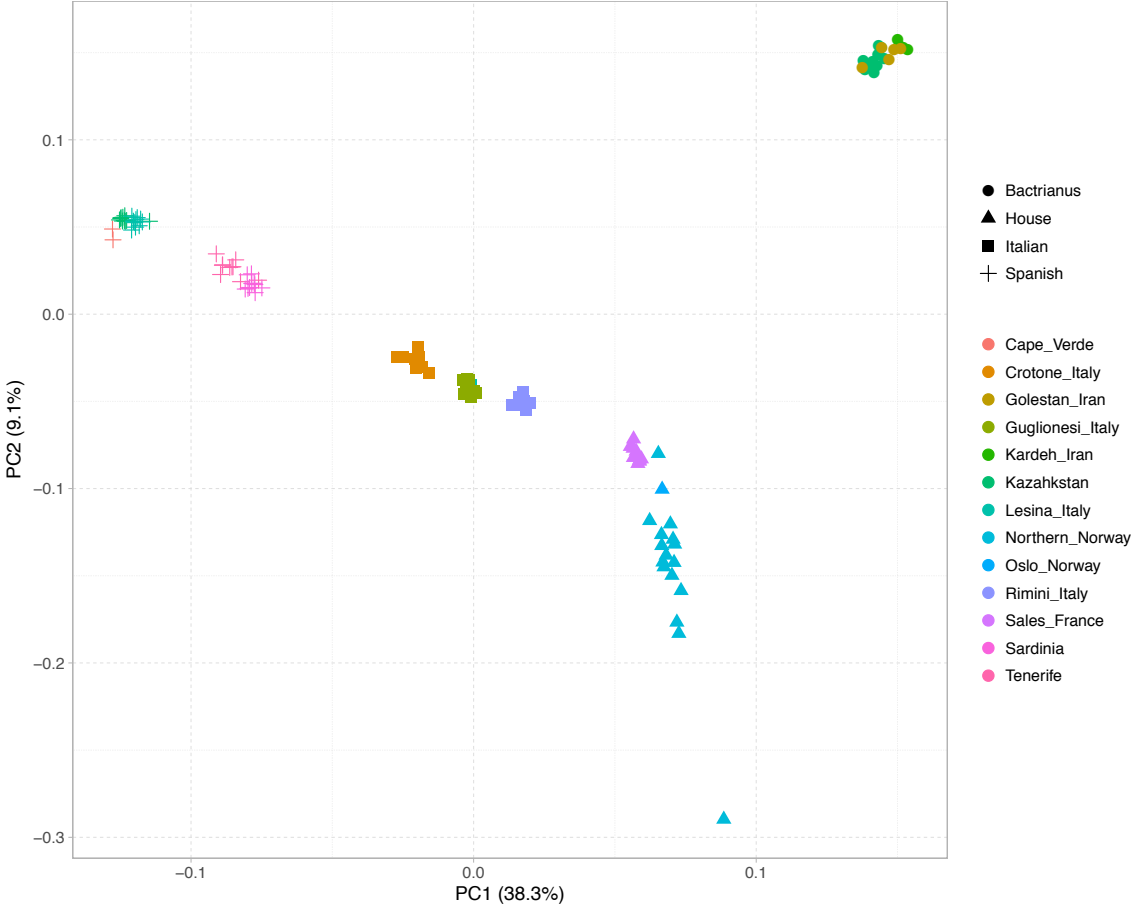


Figure S2: Cross-validation of ADMIXTURE runs showing highest support for K = 2 & 3.

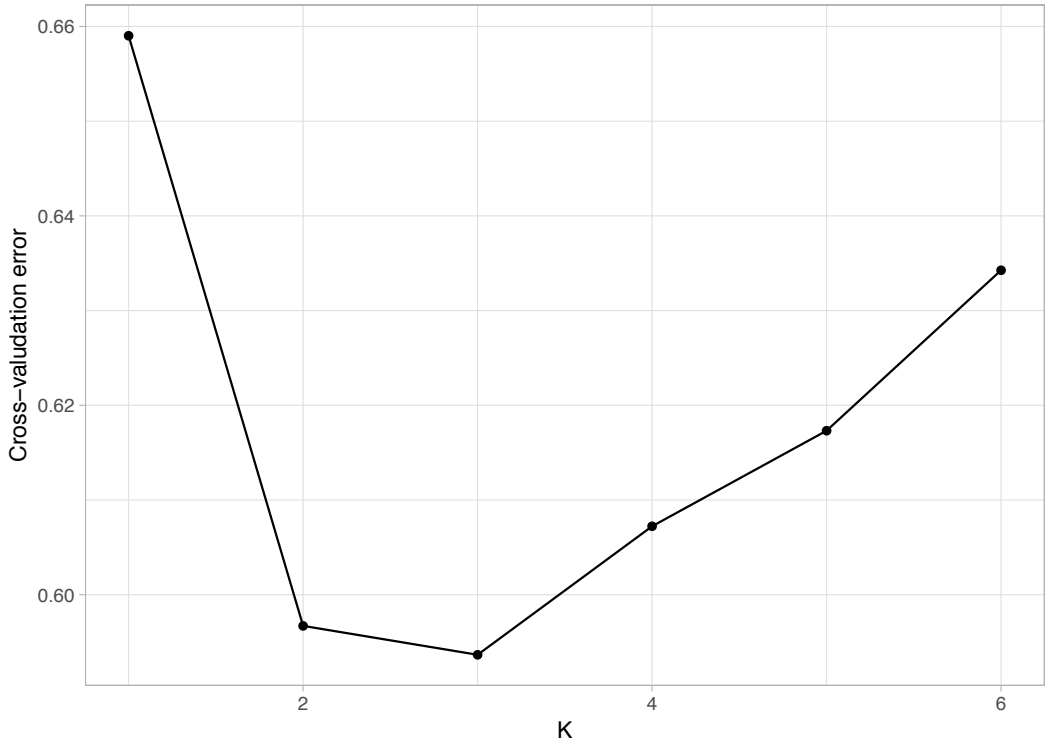


Figure S3: Boxplots of pairwise genome-wide F_{ST} estimates from 100 Kb sliding windows with a 25 Kb step.

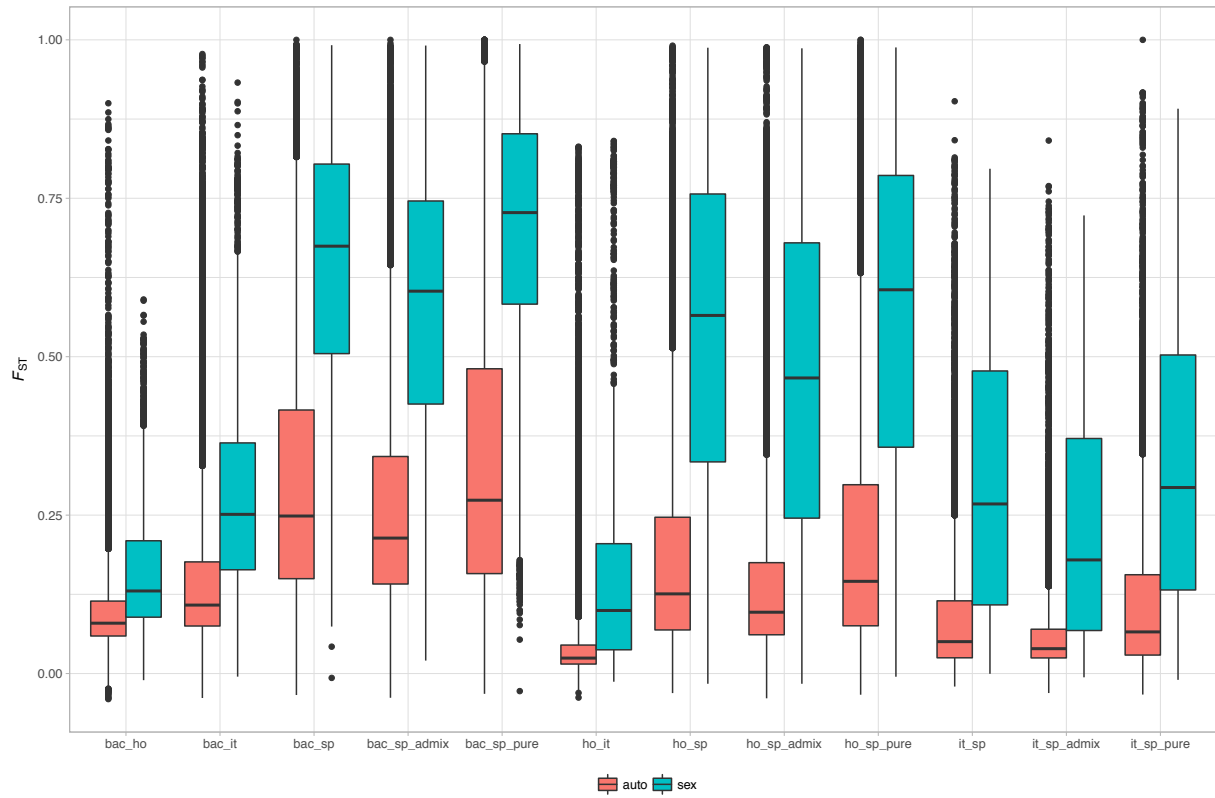


Figure S4: Boxplots of pairwise genome-wide d_{XY} (a) and f_d (b) estimates from 100 Kb sliding windows with a 25 Kb step.

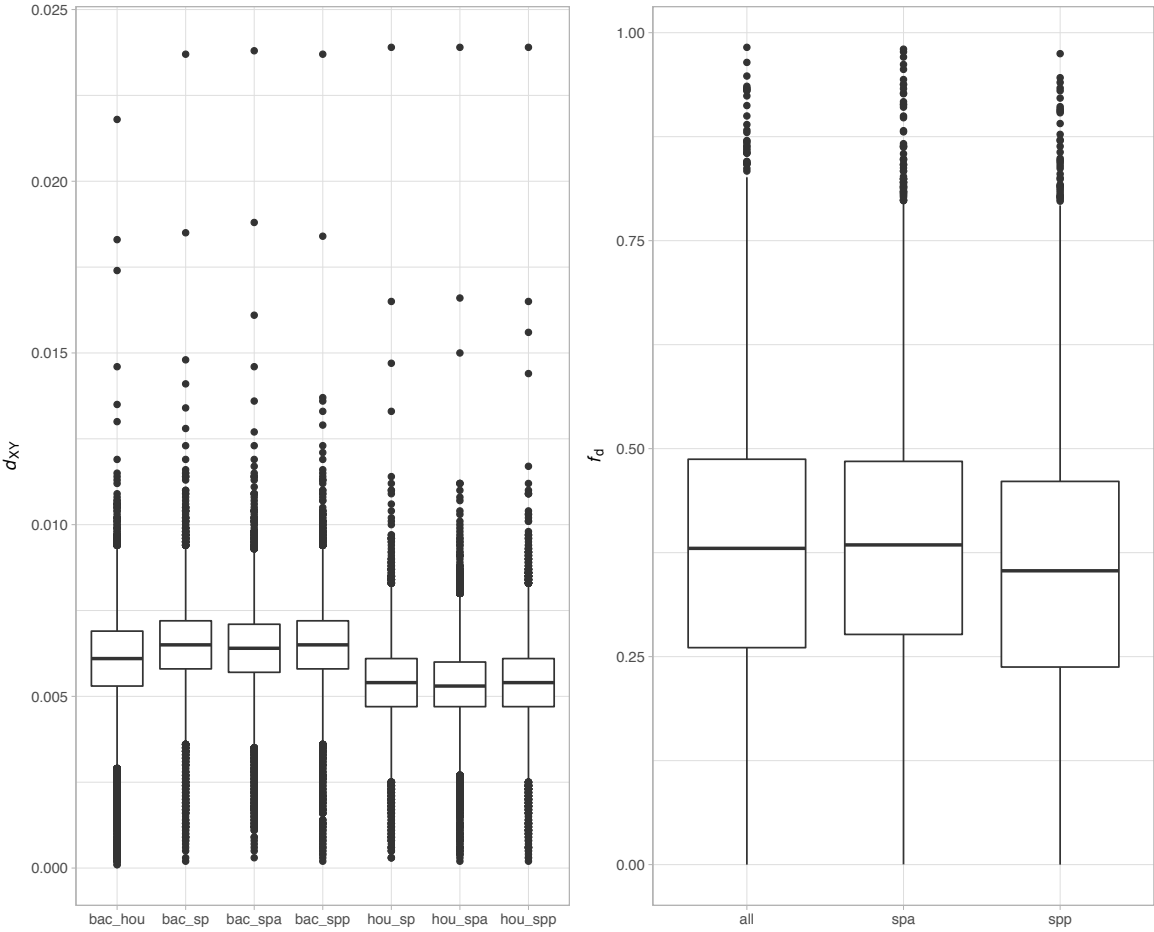


Figure S5: Demographic models tested using site-frequency spectrum methods.

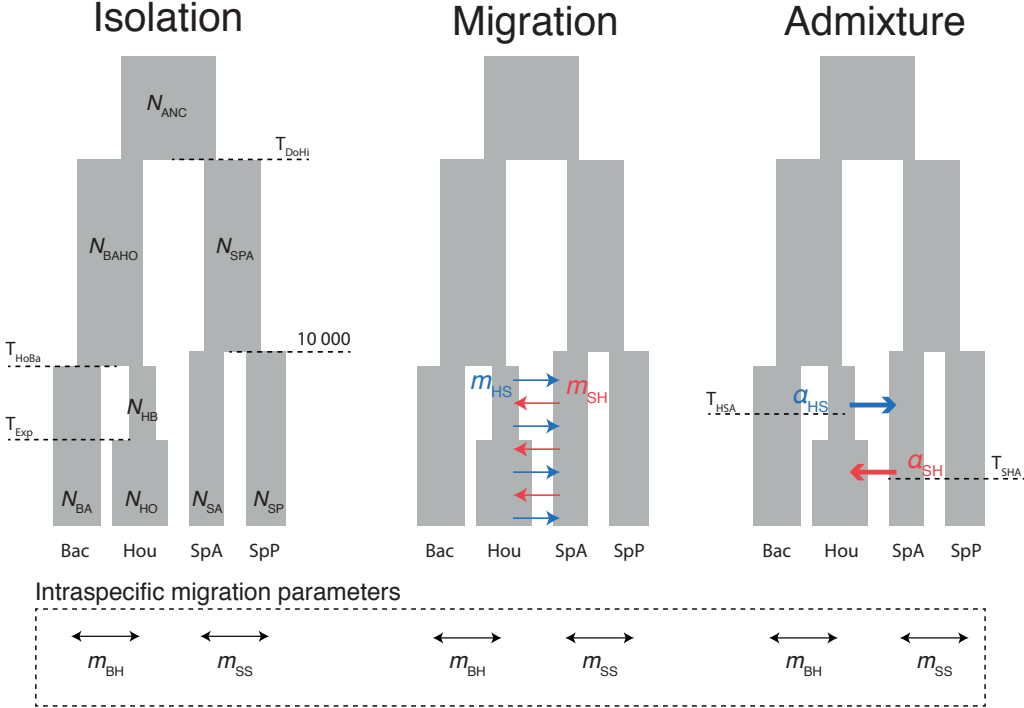


Figure S6: Log-likelihood distributions for demographic models.

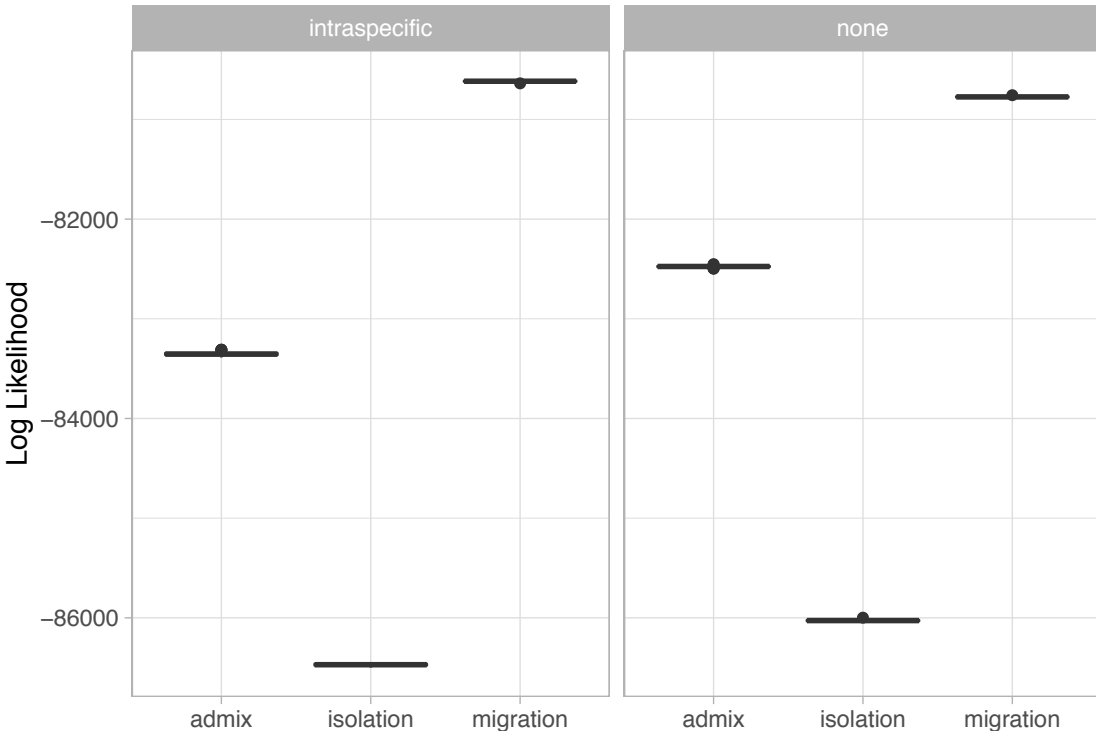


Figure S7: Manhattan plot of iHS across the house sparrow genome. The full dataset has been downsampled to 20% (176,701 SNPs) in order to aid visualisation.

