To Tame a Songbird: The Genomics of the Domestication Syndrome in a Songbird Model Species

Madza Farias-Virgens

madzayf@gmail.com

University of Washington   https://orcid.org/0000-0002-3764-1827

David Peede

Brown University

Terrence Deacon

University of California Berkeley   https://orcid.org/0000-0002-8672-5461

Kazuo Okanoya

University of Tokyo

Stephanie White

University of California Los Angeles

Emilia Huerta-Sanchez

Brown University

Article

Keywords:

Additional Declarations: There is **NO** Competing Interest.

---

1 **To Tame a Songbird: The Genomics of the Domestication Syndrome in a Songbird Model**
2 **Species**

3 Madza Farias-Virgens*[1][†], David Peede[2,3,4], Terrence Deacon[5], Kazuo Okanoya[6], Stephanie A.
4 White[7], Emilia Huerta-Sanchez[2,3]
5 *Corresponding author: madzafv@uw.edu

6 [1]Interdepartmental Graduate Program in Molecular, Cellular and Integrative Physiology,

7 University of California Los Angeles; [†]Moved to Department of Biology, University of Washington

8 [2]Department of Ecology, Evolution, and Organismal and Evolutionary Biology, Brown University

9 [3]Center for Computational Molecular Biology, Brown University

10 [4]Institute at Brown for Environment and Society, Brown University

11 [5]Department of Anthropology, University of California Berkeley

12 [6]Graduate School of Arts and Sciences, The University of Tokyo & RIKEN-Brain Science Institute

13 [7]Department of Integrative Biology and Physiology, University of California Los Angeles

14

Abstract

Many domesticated animals share a syndromic phenotype marked by a suite of traits that include more variable patterns of coloration, reduced stress, aggression, and altered risk-taking and exploratory behaviors relative to their wild counterparts. Roughly 150 years after Darwin's pioneering insight into this phenomenon, reasonable progress has been made in understanding the evolutionary and biological basis of the so-called domesticated phenotype in mammals. However, the extent to which these processes are paralleled in non-mammalian domesticates is scant. Here, we address this knowledge gap by investigating the genetic basis of the domesticated phenotype in the Bengalese finch, a songbird frequently found in pet shops and a popular animal model in the study of learned vocal behaviors. Using whole-genome sequencing and population genomic approaches, we identify strain-specific selection signals in the BF and its wild munia ancestor. Our findings suggest that, like in mammals, the evolution of the domestication syndrome in avian species involves a shift in the selective regime, capable of altering brain circuits favoring the dynamic modulation of motivation and reward sensitivity over overall augmented aggression and stress responses.

Introduction

Heightened stress and aggressive responses are significant traits of adaptive value for wild animals, manifested as impulsive reactions to perceived threats, often associated with high arousal states [1]. Domesticated animals show lower levels of reactive aggression than their wild counterparts and more consistently exhibit risk-taking behaviors in contexts where they are highly motivated, following spontaneous learned associations and/or as reinforced by domestication practices [2]. These changes are possible due to the attenuation of environmental

2

37   sources of selection commonly found in the wild and often accompanied by enhanced selective

38   pressures imposed by close socialization with individuals of the same or distinct species in the

39   domestic setting [3].

40   Here, we study one such domestication case in an under-examined order, the passerine lineage.

41   Beginning ~250YBP, white-backed munias (WBM; *Lonchura striata*), an extant wild songbird

42   readily found throughout East Asia, were brought to Japan by aviculturists [4]. Shortly after the

43   first century of domestication, these artificially bred WBMs started to exhibit piebald (i.e.,

44   patchy) patterns of plumage coloration, marking the origins of today's Bengalese finches (BF;

45   *Lonchura striata domestica*) [5] (Fig. 1A). Since then, BFs have become popular cage birds, known

46   for their easy socialization and willingness to foster offspring, including those of other songbird

47   species [6]. In addition to confirming BFs' attenuated stress response, evidenced by decreased

48   reactive aggression and neophobia relative to their munia ancestors [7-9], studies show that

49   adult BFs' song retains a greater degree of variability in the ordering of vocal elements than

50   exhibited by WBM [10, 11]. These traits evolved in concert in the BF, characterizing a prototypical

51   case of the domestication syndrome in the avian clade.

52   While we have amassed considerable knowledge about the behavioral, physiological, and

53   morphological differences between domesticated and wild mammals [12] and posed specific

54   evolutionary hypotheses for how those changes came about [13, 14], the extent to which this is

55   paralleled in the non-mammalian realm is under-studied. In this work, we address this gap in

56   knowledge by investigating the genetic basis of the domesticated phenotype in the BF. Using

57   whole-genome sequencing data and population genomic approaches, we identified strain-

58   specific and overlapping selection signals in the domesticate BF and its wild munia ancestor and

59   interrogated annotated genes within those regions of interest (ROIs). Selection signals in the

60   WBM population are associated with morphological and behavioral traits relating to stress and

61   aggression. These selection signals are weaker or absent in the domesticated BF, which instead

62   shows signals related to reward aspects of social interactions and feeding. In accordance with

63   comparisons drawn between other domesticated species and their wild counterparts [15], our

64   findings suggest that the evolution of the domestication syndrome in the BF likely involved the

65   attenuation of selective forces associated with hormonal regulation of stress and aggression and

66   the intensification of ones leading to dynamic modulation of motivation and reward sensitivity.

67   **Methods**

68   Sample collection

69   Blood samples were collected from 15 BFs and 15 WBMs housed in the Okanoya Lab at RIKEN-

70   Brain Science Institute (RIKEN-BSI, Japan). BF source colonies in the Okanoya lab originated from

71   and are continuously supplied by various breeders in Japan and diligently maintained per

72   standard bird breeder practices to avoid undesired effects of inbreeding. WBMs were imported

73   from multiple wild colonies from different locations in Taiwan and bred in Okanoya Lab's aviaries,

74   following similar practices as for BFs (~5 generations at the time of sample collection for this

75   study). To avoid immediate relatedness biases, we selected individuals with no less than two

76   degrees related to one another within each BF and WBM population. Blood samples were

77   extracted by puncturing the brachial vein and collecting the resulting droplet in a capillary tube,

78   a well-suited method for sampling small birds with minimal injury and distress. Samples were

79   stored at -80ºC at RIKEN-BSI until transported to UC Berkeley under a material transfer

80   agreement and with USDA-APHIS approval (#128913). Housing conditions and procedures

81   conformed to RIKEN's Animal Care and Use Committee.

82   Whole genome library preparation and sequencing

83   Total DNA was extracted from individual blood samples using a DNAeasy kit (Qiagen,

84   Germantown, MD) following protocols at the Evolutionary Genetics Laboratory at UC Berkeley.

85   DNA quality was visually inspected by gel electrophoresis, and DNA concentrations were

86   measured using NanoDrop spectrophotometry and a Qubit dsDNA BR Assay Kit (Thermo Fisher

87   Scientific, Waltham, MA). Samples showing signs of degradation or yielding DNA concentrations

88   insufficient for library preparation (< 50ng/microL) were discarded. Whole genomic double-

89   stranded DNA from the remaining 13 BF and 12 WBM DNA samples were fragmented to a 350bp

90   average size using a Bioruptor sonicator (Diagenode-Hologic, Denville, NJ). Sheared samples were

91   visualized by gel electrophoresis to verify size distribution, and DNA concentrations were re-

92   measured using a Qubit dsDNA BR Assay Kit. All 25 samples passed this quality control step and

93   were used to prepare whole-genome libraries using a KAPA Hyper Prep PCR-free Kit (Hoffmann-

94   La Roche, Basel, Switzerland) and TruSeq adapters (Illumina, San Diego, CA). Library sizing

95   information was obtained by electrophoresis using an Agilent 2100 Bioanalyzer system (Agilent

96   Technologies, Santa Clara, CA). The 11 BF and 11 WBM libraries passing this quality control step

97   were sequenced on 7 lanes in a 4000 HiSeq system (Illumina, San Diego, CA) in two replicates at

98   UC Berkeley's QB3 sequencing facility.

99   Sequencing data preprocessing and quality control

100   The resulting paired-end sequencing reads (150bp) underwent quality-control procedures using

101   the software package *readcleaner* (github.com/tplinderoth/ngsQC), which includes trimming

102   adapters using *cutadapt* v1.10; pair-end read merging using *pear* v0.9.10; filtering contaminants

103   identified through alignment to the human (GRCh38) and *E. coli* (ASM584v2) genomes using

104   *bowtie2* v2.2.9 [16], and generates final read quality reports using *fastqc* v0.11.4 [17] (Supp. Text

105   1).

106   Clean reads were aligned to the recently published, high-coverage zebra finch (ZF; *Taeniopygia*

107   *guttata*) genome sequenced as part of the Vertebrate Genomes Project (bTaeGut2.pat.W.v2-

108   Pacific Biosciences, Menlo Park, CA) using the *Burrows-Wheeler Alignment* (BWA) tool. Genome

109   indexes for each reference genome were generated using *BWA* v0.7.17 [18] and *SAMtools* v.10

110   [19], along with sequence dictionaries from *Picard Tools* v2.0.1 (Picard Tools:

111   CreateSequenceDictionary), and aligned using appropriate read group information (Supp. Text

112   1).

113   After alignment, duplicate reads were marked (Picard Tools: MarkDuplicates), and an additional

114   step of local indel realignment was performed using *GATK* v3.8.0 [20]. Indels were called from all

115   sampled individuals (GATK: HaplotypeCaller) and used to create lists of indel sites for each

116   sampled individual separately (GATK: RealignerTargetCreator), which were then used as targets

117   in the realignment (GATK: IndelRealigner) (Supp. Text 1).

118   A subsequent filtering step was performed to remove paralogous or repetitive sequence regions

119   that confound short-read mapping using *ngsParalog* (github.com/tplinderoth/ngsParalog). This

120   step involved calling single nucleotide polymorphisms (SNPs) in each BF and WBM population

121   separately using *Analysis of Next Generation Sequencing Data* (ANGSD) [21], a software package

122   that calculates various population genetic summary statistics from genotype likelihoods (Supp.

123 Text 1). The union of the identified variant positions (angsd:-*dovcf 1*) in each population served

124 in estimating per-site paralog-log likelihood ratios. Likelihood ratios for each site were then

125 compared between BF and WBM, and the highest values were retained. The -log likelihood ratio

126 is asymptotically distributed as a 50:50 mixture of a chi-square with one degree of freedom and

127 a chi-square distribution with zero degrees of freedom. Accordingly, the cutoff above which a

128 given site would be considered paralogous and excluded from future analyses was calculated as

129 $0.5 + 0.5 * qchisq(p, df = 1) = 1 - W$, where $W = 0.05$ significance level, and corrected

130 for multiple hypothesis testing using the total number of sites.

131 After excluding potential paralogous or repetitive regions, we calculated depth counts in each

132 remainder position for each BF and WBM individual (angsd: *-doCounts 1, -dumpCounts 2)* (Supp.

133 Text 1). The sequencing data rendered an average per-site coverage depth per individual of 8x.

134 To avoid spurious results from missing data, sites covered less than 9 out of 11 individuals in each

135 population were further filtered from the data. Additionally, sites with average depth across

136 individuals of > 20x (~2SD+) in each population were filtered out, as they could be sequencing

137 artifacts or represent paralogous and repetitive regions bypassed by the prior filtering. Finally,

138 among the remaining sites, only diallelic sites were inferred from in downstream analyses.

139 Population genomic analyses

140 Using the *GATK* model within *ANGSD,* we calculated genotype likelihoods and per-site allele

141 frequencies (Supp. Text 1). Genotype likelihoods from sequencing reads were calculated as the

142 probability of the observed sequencing data *X* given the three different possible genotypes $g =$

143 0,1,2 representing the number of copies of the minor allele in a site for an individual, $P(X|G =$

144 $g)$. The allele frequency at a given site in the population was used in calculating the prior

7

145   genotype probability of the same site for an individual, $P(G_{is} is|p_s)$, which was calculated under

146   the assumption of Hardy-Weinberg Equilibrium (HWE) for the three different possible genotypes

147      $g = 0,1,2$ as $P(G_{is} = 0|p_s) = (1 - p_s)^2$, $P(G_{is} = 1|p_s) = 2p_s(1 - p_s)$ and $P(G_{is} = 2|p_s) =$

148   p_s^2), and used to define posterior genotype probabilities using Bayes' theorem.

149   To perform *Principal Component Analysis* (PCA) of the genetic relationship among individuals in

150   our populations, we used a heuristic approach implemented in *PCAngsd* [22] (Supp. Text 1) that

151   does not rely on the assumption of conditional independence between sampled individuals and,

152   in this way, can model missing data with the inferred population structure. *PCAngsd* replaces

153   population allele frequencies with individual allele frequencies calculated from posterior

154   expectations of genotypes to update prior information in calculating posterior genotype

155   probabilities. The individual allele frequencies are computed through a *Singular Value*

156   *Decomposition* method, where genotypes are reconstructed using the top principal components

157   and serve to inform prior genotype probabilities, assuming HWE in calculating posterior genotype

158   probabilities and updating genotype expectations. This procedure is iterated until individual

159   allele frequencies converge. Finally, the covariance matrix is calculated within a probabilistic

160   framework by summing over individual's genotype weighted by the joint posterior probability of

161   individual allele frequencies. Eigendecomposition of the updated estimated covariance matrix is

162   then performed to obtain the principal components.

163   Per site allelic frequencies estimated for each population separately were used to estimate the

164   proportion of alleles at specific frequencies in each population, with ancestral states as defined

165   in the ZF genome – 1-dimensional *Site Frequency Spectrum* (1D-SFS), as well as at a particular

166   frequency in one population and at a different frequency in the other population – 2-dimensional

167 SFS (2D-SFS) (Supp. Text 1). Accordingly, considering a set of *n* aligned DNA sequences, the SFS is

168 defined as the vector that describes the frequency of *k = 1 to n* derived alleles; $\eta \;=\; (\eta_k)_k \;=$

169 $0, ..., n$, where η_k is the number of sites with *k* derived alleles.

170 A series of statistics summarizing the variation or distribution of alleles within the dataset was

171 calculated using *ANGSD realSFS*. The 2D-SFS served as the prior in subsequent per site *Fixation*

172 *index* (Fst) calculations, a commonly used population genetics statistic that measures genetic

173 variance *between* relative to *within* populations. Per-site Fst estimates were calculated using

174 Bathia *et al* [23] derivation of Hudson's Fst estimator, as implemented in *ANGSD realSFS fst*. In

175 the analysis of multiple SNPs comprising genomic regions, Fst was calculated as the ratio of the

176 sum of per site Bathia et al. estimations of the variance *within* and *between* populations. This

177 approach allowed us to calculate weighted Fst between our BF and WBM populations in non-

178 overlapping windows spanning 10kb across the genome, as well as across all positions within a

179 given gene's full-length and within exonic and intronic annotated boundaries (Supp. Text 1 and

180 2; Supp. Data 1).

181 Unfolded genome-wide 1D-SFSs were used as priors in calculating various estimators describing

182 the genetic diversity in each BF and WBM population separately, using *ANGSD realSFS saf2theta*

183 (Supp. Text 1). These estimators measure the ratio between the nucleotide diversity observed in

184 a population and the expected under a neutral evolution model. We calculated *Waterson's theta*

185 (tW), *Pairwise theta* (tP), and *Tajima's D* (TajD) estimators of genetic diversity [24, 25] (Supp. Text

186 2) for each BF and WBM population in non-overlapping windows spanning 10kb across the

187 genome, as well as across all positions within a given gene's full-length and within exonic and

188 intronic annotated boundaries (Supp. Data 2).

189   Selection of a sequence variant associated with an individual's increased chances of survival and

190   reproduction frequently leads to surrounding variants also increasing in frequency within the

191   population due to linkage disequilibrium (i.e., selective sweep) [26]. An event like this will cause

192   TajD to become negative, as there will be relatively more high-frequency variants than expected

193   under a neutral model, thus rendering the expectation of tP smaller than the expectation of tW.

194   Therefore, one of the general signatures of selection is a loss of nucleotide diversity around a

195   beneficial genetic variant. This effect might result from positive selection during the sweep-up of

196   the newly advantageous variant, or purifying selection, resulting in the purging of deleterious

197   variants surrounding the neutral one (i.e., background selection). Positive values of *Tajima's D*

198   arise from an excess of intermediate frequency alleles and can result from population

199   bottlenecks, structure, and/or balancing selection.

200   We performed selection scans using *SweepFinder2* [27], an implementation of a method that

201   performs a composite likelihood ratio test for positive selection (Supp. Text 1). In this method,

202   the likelihood of the null hypothesis is calculated for each population from their respective

203   genome-wide 1D-SFS, and the likelihood of the alternative hypothesis is calculated from a model

204   in which a recent selective sweep alters the neutral spectrum. *SweepFinder2* selection scans were

205   run in non-overlapping windows spanning 10kb each (Supp. Data 3).

206   **Results**

207   Differentiation and variation along Bengalese finch and white-backed munia genomes

208   The PCA of all BF and WBM individuals shows a clear separation between the two populations

209   along PC1 (Fig. 1B). This distinction is marked by a greater proportion of low-frequency genetic

210   variants in the wild WBM population. In contrast, variants at higher-to-fixed frequencies prevail

211   in the domesticated BF, indicating an overall loss of genetic variability within their population

212   compared to their wild counterparts (Fig. 1C).

213   Differences in allelic frequencies between BF and WBM are unevenly distributed across windows

214   along individual chromosomes, allowing for the identification of local signals of differentiation

215   between the two songbirds, as measured by the canonical *Fixation index* (Fst) (Fig. 2). While the

216   genetic differentiation of windows within individual autosomes is close to the autosomal average

217   (10kb Windowed Mean Fst +/- SD ~ 0.2±0.08), the sex chromosome Z (ChrZ) shows considerably

218   higher levels of differentiation (10kb Windowed Mean Fst +/- SD ~ 0.3 ±0.15). The higher

219   differentiation of ChrZ may be partly attributed to this chromosome's smaller effective

220   population size due to its heterogametic mode of transmission (males: ZZ; females: ZW) [28].

221   We also report genomic windows and sets of genes for which observed variability deviates from

222   expected under neutrality in each population, as evidenced by *Tajima's D* (TajD), *Waterson's*

223   *theta* (tW), and *Pairwise theta* (tP) summary statistics (Fig. 2). Genetic diversity in WBMs deviates

224   less across their genome (WBM autosomal 10kb Windowed Mean TajD +/- SD 0.3±0.43, var =

225   0.19; WBM ChrZ 10kb Windowed Mean TajD +/- SD 0.2±0.60, var = 0.36), whereas in the BF, long

226   stretches of the genome show either considerable loss or gain of variability (BF autosomal 10kb

227   Windowed Mean TajD +/- SD 0.5±0.78, var = 0.61; BF ChrZ 10kb Windowed Mean TajD +/- SD

228   0.5±0.82, var = 0.67) (Fig. 2). As with many domesticates, these genome-wide observations are

229   likely due to the heightened effects of genetic drift, set off by the primary population bottleneck

230   that marks the major domestication event [29] and furthered by domestication practices of

231   breeding small BF populations in captivity.

Regions of interest (ROIs) in the Bengalese finch and the white-backed munia genomes

Our sweep scans allowed the detection of strain-specific signals in the domesticate BF and its wild munia ancestor (Fig. 3). Subsequently, we defined genomic regions of interest (ROIs) as the contiguous segments containing the top 5% of windows with a higher composite likelihood ratio (CLR) in each BF and WBM (Fig. 3; Supp. Table 1). Six ROIs were defined encompassing sweep signals found in four autosomes as exclusive to the BF population: **BF1-Chr1**, **BF2-Chr1**, **BF3-Chr1A**, **BF4-Chr2**, **BF5-Chr2**, **BF6-Chr4** (Supp. Fig. 1-12). The genetic differentiation between BF and WBM in these regions was particularly higher than found for the rest of the autosome genome, as shown by their greater-than-autosomal average Fst values (Fst ~ 0.4±0.09, greater than 97% of autosomal windows). TajD values for windows spanning each BF ROI were generally lower for the BF population (10kb Windowed Mean TajD +/- SD ~ -1.0±0.5, lower than 95% of the autosomal windows) than found in the WBM population (10kb Windowed Mean TajD ~ 0.2±0.43; lower than 60% of the autosomal windows), with tW (BF tW ~ 18.7±7.54, lower than 88% of the autosomal windows vs. WBM tW ~33.7±9.44, lower than 54% of the autosomal windows) and tP (BF tP ~ 13.9±5.95, lower than 95% of the autosomal windows vs. WBM tP ~ 35.9±10.40, lower than 56% of the autosomal windows) following the same trend of lower values in the BF, consistent with their greater loss of variability in those genomic segments in the BF. These results suggest that BF ROIs might have experienced substantial changes due to recent or ongoing positive selection specific to the BF population, resulting in higher differentiation with the WBM population.

In the WBM population, we defined four genomic ROIs encompassing sweep signals found in four autosomes (Fig. 3; Supp. Table 1). Among those, only one signal was exclusive to the WBM

254 population, **WBM4-Chr3**, while overlapping signals were also detected in the BF population at

255 **WBM1-Chr1**, **WBM2-Chr1A**, and **WBM3-Chr2** (Supp. Fig. 1-12). In general, the genetic

256 differentiation of these ROIs is not particularly higher than found in the rest of the autosome

257 genome, as shown by their near-average values of Fst (10kb Windowed Mean Fst +/- SD ~

258 0.20±0.09, higher than 53% of the autosomal windows). These aspects suggest that selective

259 sweeps in these regions may have occurred in the ancestral WBM population. Alternatively, these

260 overlapping signals could be false positives, detected by *SweepFinder2* in both populations due

261 to their shared demographic history. Both these scenarios can influence gene function in these

262 regions, with their differing degrees of genetic differentiation resulting from population-specific

263 factors post-divergence, including continued gene flow into and between BF populations, and

264 the different environmental conditions and consequent selective pressures ancestral and derived

265 populations are uniquely subjected to.

266 TajD values for windows overlapping ROIs **WBM1-Chr1**, **WBM2-Chr1A**, and **WBM3-Chr2** were

267 lower for the WBM relative to the BF population (WBM TajD ~ -0.4±0.58, lower than 94% of the

268 autosomal windows vs. BF TajD ~ 0.1±0.63, lower than 73% of the autosomal windows),

269 consistent with a greater loss of relative variability in the given genomic region in the WBM

270 population. However, **WBM4-Chr3** ROI presented a different pattern of genetic variation,

271 whereby the WBM population showed an exclusive sweep signal but higher TajD values than the

272 BF population. Closer inspection reveals that TajD values in the BF population decay (BF TajD ~ -

273 2.0±0.20) and gradually reverse toward less negative values along **WBM4-Chr3**, while TajD values

274 remain closer to 0 in the WBM population (WBM TajD ~ -0.3±0.63) (Supp. Fig. 1-12). The decrease

275 in TajD values in the BF population is accompanied by increases in tW (BF tW ~ 36.5±6.26, greater

276   than 72% of the autosomal windows) and decreases in tP (BF tP ~ 18.9±3.49, lower than 89% of

277   the autosomal windows) suggesting the accumulation of many rare variants rather than common

278   variants in the region in the BF; while in the WBM both tW and tP values remain proportionally

279   lower (WBM tW ~ 5.8±2.22 and tP ~ 5.4±2.56, both lower than 99% of the autosomal windows).

280   This region is associated with a common fragile site prone to mutate and frequently undergoes

281   large structural rearrangements in mammals and birds, within which breakpoints locate to the

282   Parkinson's disease (PD) associated gene, **PRKN** [30]. Accordingly, our findings of greater

283   variability in the BF population on a region otherwise showing an increased proportion of higher

284   frequency variants in the WBM may relate to increased genomic instability at **PRKN** in the BF

285   population.

286   As for **WBM4-Chr3** ROI, **BF7-Chr4** ROI presented a unique pattern of genetic variation, whereby

287   the BF population showed a stronger sweep signal but higher TajD values than the WBM

288   population. (BF TajD ~ 0.16±0.72; WBM TajD ~ -0.62±0.65) (Supp. Fig. 1-12). Accordingly, tW and

289   tP values in the BF deviate less from each other along **BF7-Chr4** (BF tW ~ 5.9±3.45, BF tP ~

290   6.4±4.28), while WBM shows a more skewed distribution of tW and tP values (WBM tW ~

291   10.2±5.29, WBM tP ~ 8.9±5.49). One possible explanation for the genetic pattern we observed is

292   that complex demographic or selective events after a genetic sweep helped shape it. Previous

293   findings support the idea that the BF population experienced a recent bottleneck followed by a

294   rapid expansion [29]. We suggest that the more intense genetic sweep signal in the BF population

295   could result from advantageous alleles quickly rising in frequency during the population growth,

296   while common domestication practices involving the interbreeding of small captive populations,

297   together with changes in selective pressures due to domestication, may have caused selection

298   for genetic diversity to increase at the same or nearby loci experiencing a sweep.

299   Sweep signals within ChrZ cover a broad genomic area, including a major peak shared by both

300   WBM and BF, **BF8-ChrZ**, flanked by regions showing increased CLR values only in the BF (Supp.

301   Fig. 1-12). **BF8-ChrZ** shows a high degree of differentiation between BFs and WBMs (Fst ~

302   0.5±0.17), but similarly negative TajD values (BF TajD -0.3±0.80; WBM TajD -0.1±0.87), suggesting

303   that while both populations have experienced a selective sweep, the specific alleles under

304   selection might differ, leading to distinct genetic patterns in each population.

305   Our subsequent investigation focused on the genes annotated within these ROIs, enabling us to

306   identify candidates accounting for the reported phenotypic distinctions between BFs and WBMs

307   (Table 1). A number of these genes exhibit selection signals in domesticated avian and

308   mammalian species, spanning various indigenous and enhanced breeds worldwide (Supp. Table

309   1).

310   Selection signals in the WBM population are associated with hormonal regulation of stress and

311   aggression.

312   Like other domesticates, BFs exhibit less stress than their wild counterparts, as shown by lower

313   levels of cortisol measured in fecal samples collected throughout the day; and lower reactive

314   aggression, as demonstrated by less frequent biting responses when provoked with a stick

315   attached to a piezo-electric sensor [9, 30]. In mammals and birds, stress and aggression are

316   closely linked to hormonal regulation in the body, primarily mediated by cortisol, adrenaline, and

317   testosterone [31]. As a result, genes associated with hormone function are likely to contribute to

318   variations in stress and aggression between domesticated and wild animals [32]. Following this

319   rationale and considering previous genetic associations, we have identified genes implicated in

320   reactive aggression and stress, exhibiting selection signals in wild munias: **TBX19** (**WBM1-Chr1**),

321   **AKR1D1** (**WBM2-Chr1A**), **MARCHF11** (**WBM3-Chr2**). Noticeably, **MARCHF11** shows high

322   divergence in 11 *Lonchura* munia species radiated throughout Australia and Papua New Guinea

323   [33]; specifically, it shows high differentiation in four pairwise comparisons between sympatric

324   *Lonchura* species with minimal genome-wide divergence, therefore possibly involved in

325   generating adaptive diversity in munias in the wild.

326   We have also identified a selection signal exclusive to the WBM population in a region of

327   importance for craniofacial morphology, for which changes are associated with domestication in

328   mammals [34, 35]. The sweep signal within **WBM4-Chr3** ROI includes the gene **SMOC2**, which

329   codes for a protein related to matrix assembly and cell adhesiveness, expressed in cartilage and

330   bone during development and in the adult skeleton. In mammals, this gene product is implicated

331   in craniofacial dysmorphism, specifically brachycephaly (i.e., shortening of the anteroposterior

332   skull length and widening along the cranial mediolateral axis) and brachygnathism (i.e.,

333   underbite) [34-36]. The role of **SMOC2** in craniofacial morphology appears to be preserved in

334   birds, as it figures among genes showing significant SNP association to beak shape and with a

335   high degree of local linkage disequilibrium [37]. Though direct morphological comparisons

336   between BF and WBM are yet to be explored, the absence in the BF of selective signals found in

337   their munia ancestor at **SMOC2** could indicate craniofacial differences between the two. This

338   finding is consistent with controlled observations that WBMs show stronger biting to provocation

339   than BFs, as differences in beak dimensions and/or head width are the major predictors of bite

340   force in finches [38]. Differences in bite force between BF and WBM, and possibly associated

16

341 craniofacial allometric scaling, could derive from the lack of selective pressures such as physical

342 endurance against aggression and foraging harder and more varied materials and food in the

343 wild.

344 Selection signals in the BF population locate to regions of importance to color variation,

345 neurodevelopment, and dopaminergic neuromodulation of behavior.

346 Eye and coat color tend to be fixed traits within wild species, with some variation related to age

347 and sexual dichromatism [39]. In contrast, individual color variation in the adult stage is

348 characteristic of many domestic species [40]. In line with this, plumage pigmentation is a

349 noticeable difference between BFs and their munia ancestors. While WBMs stereotypically have

350 dark-brown heads, tails, and wing plumage, which morphs into more visibly barred feathers on a

351 darker to lighter gradient towards their flank and rump, BFs' plumage assumes the most varied

352 combination of tones, with variable body distribution, including dark brown (chocolate), lighter

353 brown (chestnut), yellowish brown tan (fawn), pale yellow (crèmino) and white (albino) morphs.

354 We found selection signals specific to the BF at the **HERC2-OCA2-GABRB3** gene cluster in Chr1

355 (**BF1-Chr1**). **OCA2** encodes the melanosomal transmembrane protein P, involved in the trafficking

356 and processing of tyrosinase, a catalyzer of crucial steps in the melanogenesis pathway, whose

357 absence halts melanin production [41]. **OCA2** expression is regulated by an element located

358 within an intronic region of its contiguous upstream neighbor **HERC2**, where variants are major

359 determinants of blue/brown iris coloration and are associated with differences in skin

360 pigmentation in humans [42]. Mutations in **OCA2**'s mouse orthologue (i.e., the pink-eyed dilution

361 locus), causing the lack of functional P protein, lead to a pink-eyed and light fur phenotype [41].

362 Contiguously downstream of **OCA2** is the gene **GABRB3**, which encodes the β3 subunit of GABAA

363 receptors. Consistent with its major role in neural function and development, mutations in this

364 gene have been frequently associated with the neurodevelopmental aspects of Angelman and

365 Prader-Willi syndromes (AS and PWS, respectively) [43]. However, a potential contribution of

366 **GABRB3** to ocular function and development is suspected, as the deletion of **GABRB3** alone

367 causes a nearly complete loss of retinal pigmentation due to atrophied melanosomes [44]. It has

368 been, therefore, proposed that **GABRB3** or elements within the gene might regulate **OCA2**

369 expression, and impairment of this function may indirectly cause ocular hypopigmentation and

370 visual defects, as seen in PWS in AS. Historical accounts indicate that piebald plumage coloration

371 spontaneously emerged ~140YBP and was picked up by Japanese breeders, who further

372 developed many BF color morphs [4]. Noticeably, BF domestication was marked by significant

373 breeding efforts for full white morphs, leading to the development of the black-eyed and pink-

374 eyed albino BF strains [4]. Our findings raise the possibility of **HERC2-OCA2-GABRG3** gene cluster

375 involvement in BF plumage and eye coloration, which is especially relevant in the context of the

376 development of the pink-eyed albino and other light-colored BF strains.

377 Several lines of evidence demonstrate BFs have evolved more complex vocal abilities than WBMs:

378 adult BF song retains a greater degree of variability in the ordering of vocal elements than

379 exhibited by WBM; BFs exposed to multiple tutors compose their songs from a combination of

380 excerpts from the different tutor's songs, in contrast to their wild ancestors, who latch onto and

381 copy the song of only one tutor; moreover, in cross-fostering experiments, BFs learn their WBM

382 foster parent's song more efficiently than WBMs learn their BF foster parent's song [10, 11, 45].

383 Several genes located within the specified BF ROIs exert a significant influence on the regulation

384 of cell guidance and connectivity, particularly relevant in neurodevelopment, and thus potentially

385 contributing to the behavioral signatures observed in the BF: **COL4A2** and **COL4A1** (**BF2-Chr1**),

386 **WNT7B** (**BF3-Chr1A), CDH9** (**BF4-Chr2**), **SLIT2** (**BF6-Chr4**). We also uncovered clear signals of

387 selection exclusive to the BF population in regions including genes capable of eliciting changes in

388 behavior through their involvement in neuromodulation, particularly in relation to a direct

389 association with dopaminergic transmission: **DDC (BF5-Chr2).** Additionally, we detected major

390 selection signals in the BF in regions enclosing two other genes involved in synaptic transmission

391 and related to the molecular etiology of Parkinson's disease (PD): **SCNA (BF7-Chr4)** and **SNCAIP**

392 **(BF8-ChrZ)** (Table 1).

393 **Discussion**

394 *Decreased Stress and Reactive Aggression*

395 Our study found intensified selection in the WBM relative to BF on regions comprising genes

396 linked to the hormonal regulation of stress and aggressive responses. The gene **TBX19** (**WBM1-**

397 **Chr1**) encodes a transcription factor that marks pituitary cell lines that will later express pro-

398 opiomelanocortin (POMC), the peptide that elicits adrenocorticotropic hormone (ACTH) release

399 [46]. Changes in **TBX19** are suggested to underlie tameness and timidity in Chinese indigenous

400 pigs [47]. Though **MARCHF11** (**WBM3-Chr2**) function in the pituitary remains unexplored,

401 emerging findings suggest that its brain expression is regulated by social and stress-related cues

402 tied to reproduction [48]. Key to this process is vasopressin, a hormone synthesized within the

403 hypothalamus and stored in the posterior pituitary. Vasopressin is released to target brain

404 regions, where it has an indirect inhibitory effect on gonadotropin-releasing hormone through

405 the potentiation of ACTH and stimulation of cortisol secretion [49]. **MARCHF11** expression is

406 sexually dimorphic in midbrain vasopressin-responsive neurons, which are activated by prosocial

407   stimuli (e.g., the presence of a female mouse, leading to mating or huddling) but not by

408   antagonistic social stimuli (e.g., the sight of a male mouse, leading to fighting or drawing back)

409   [48]. In songbirds, neurons expressing the avian homolog of the mammalian vasopressin,

410   vasotocin, exert effects on sex-specific behaviors, pair bonding, gregariousness, and aggression

411   [50, 51]. These observations implicate **MARCHF11** in sex-specific reproductive physiology and

412   behavior that discriminates between stress conditions, as regulated by vasopressin/vasotocin.

413   **AKR1D1** (**WBM2-Chr1A**) encodes a steroid A-ring reductase, 5β-reductase, which participates in

414   bile acid synthesis and regulates hormonal action in the liver, including glucocorticoid-mediated

415   effects on carbohydrate and lipid metabolism associated with stress responses [52]. The enzyme

416   5β-reductase is highly prevalent in the brain and pituitary of birds, where its primary function is

417   to inactivate testosterone [53]. In oscine species, steroid metabolism in the brain is linked to

418   structural changes, particularly in the neural circuits that govern song production and

419   modifications in sexual and aggressive behaviors [54]. Altogether, our findings indicate

420   heightened selection pressure on genes associated with the hypothalamic-pituitary-adrenal

421   (HPA) axis function in WBMs relative to their domesticated descendant BFs. This phenomenon

422   likely contributes to the observed differences in aggressive behavior and stress responses

423   between these two groups, aligning with prior studies in mammalian domesticates [15].

424   *Reduced neophobia in feeding contexts*

425   Controlled observations reveal that BF exhibit decreased levels of neophobia in feeding contexts

426   relative to their wild counterparts, as demonstrated by their lower latency in approaching the

427   food cup in the presence of a foreign object [8]. This result can also be explained by differences

428   in food motivation between BF and WBM, to which our discovery of evolutionary changes in

429   regions encompassing genes implicated in reward aspects of food intake may relate**.** Changes in

430   **SNCA (BF7-Chr4)** mRNA and protein levels are associated with reward aspects of food intake and

431   social interactions in mammals [55, 56]. *In vivo* studies suggest that the **SNCA** gene product,

432   alpha-synuclein, regulates dopamine (DA) availability in response to the direct action of

433   metabolic hormones in DA centers or downstream from homeostatic signaling in the

434   hypothalamus [56]. Similarly, **SNCAIP (BF8-ChrZ)** gene product**,** synphilin-1, is implicated in food

435   intake and fat deposition in mammals [57-59] and in hyperphagia, first characterized in

436   transgenic mice overexpressing this gene predominantly in neurons, which unexpectedly

437   manifested obesity resulting from increased food intake, and in the absence of PD-like symptoms

438   [60]. Signatures of selection at **SNCAIP** have been detected in broiler chicken lines divergently

439   bred for abdominal fat content, which supports synphilin-1's relation to food intake and adiposity

440   as being conserved in avian species [61].

441   *Increased flexibility in birdsong learning and practice*

442   Birdsong is regulated by dopaminergic innervation originating in the brain's motivational centers

443   and projecting to specialized telencephalic nuclei, including the striatal region known as Area X

444   and cortical vocal control nuclei, such as HVC (used as a proper name) and the robust nucleus of

445   the arcopallium (RA) [62]. Our study found intensified selection in the BF relative to WBM on

446   regions comprising genes tightly linked to dopaminergic transmission, thus capable of affecting

447   brain circuits highly relevant for song learning and production. The regulation of DA levels within

448   song nuclei relates to corresponding changes in the expression of enzymes involved in DA

449   biosynthesis, like **DDC (BF5-Chr2),** coding for dopa decarboxylase [62]. In male songbirds, **DDC**

450   mRNA levels in the ventral tegmental area (VTA), the origin of the dopaminergic pathways, are

451   negatively correlated with the amount of contact received from a female partner [103]. The gene

452   **SNCA (BF7-Chr4)** negatively regulates DA availability within song control nuclei, where it is

453   constitutively downregulated at both mRNA and protein levels [63, 64]. It has been proposed

454   that overall low levels of alpha-synuclein in these regions may serve as a protective mechanism

455   against protein aggregation and consequent motor deficits during aging [65]. **SNCA** expression is

456   developmentally downregulated during the critical period for song learning within RA and Area

457   X's adjacent efferent nucleus, the lateral magnocellular nucleus of the anterior nidopallium

458   (LMAN) [64, 66]. In the adult songbird brain, alpha-synuclein is differentially regulated within

459   Area X between vocal practice and performance, with increased levels scaling to the time spent

460   singing alone for practice but not during singing directed to females [67]. Though less is known

461   about synphilin-1's function within the song circuit, **SNCAIP (BF8-ChrZ)** mRNA expression is

462   significantly downregulated within Area X during song practice [68]. This result is consistent with

463   synphilin-1's role in regulating the ubiquitin-mediated degradation of alpha-synuclein [69] and

464   suggests this interaction contributes to the modulation of song variability during practice.

465   We also found signals of selection present only in the BF population in genes regulating axon

466   guidance and connectivity that contribute to the differentiation and specialization of the song

467   control circuit, **COL4A1** and **COL4A2 (BF2-Chr1)**, and **SLIT2 (BF6-Chr4)**. **COL4A2** and **COL4A1** gene

468   products show differential expression between song nuclei in male songbird brains and

469   corresponding vestigial reminiscences in female brains in species where song is sexually

470   dimorphic [70]. **COL4A1** expression is particularly enriched in HVC, where it peaks during

471   sensorimotor learning (i.e., 45 days post-hatch in zebra finches) [71]. Although less is known

472   about the specific role of **SLIT2** within the song circuitry, the down-regulation of its paralog SLIT1

473   in song nuclei RA and HVC and in the brainstem motor nucleus nXIIts is suggested to facilitate the

474   targeting of ROBO-expressing terminals and guide the development of neuronal projections from

475   HVC to RA, as well as direct projections from RA to nXIIts, a connection critical for learned

476   vocalizations [72].

477   **Conclusion**

478   Results from various avenues of inquiry indicate that a typical recipe to domestication includes

479   the attenuation of selective forces for augmented reactive aggression and the intensification of

480   ones leading to dynamic modulation of motivation and reward sensitivity. Our findings suggest

481   that the same principles apply to tame a songbird. Major selective sweeps in the BF comprise

482   genes essential for DA synthesis, availability, and response in the songbird brain, thus highly

483   relevant in the specialization and function of brain nuclei dedicated to song. Our scans also reveal

484   selection signals specific to the BF in genes causally linked to pigmentation differences in other

485   animals, which could similarly explain BF color morphs.

502   **Author Contributions**

503   Designed research: MFV, KO, TD; performed research: MFV; analyzed data: MFV, DP, EHS;

504   contributed to writing: MFV, SAW, DP, EHS, KO, TD.

505   **Competing Interests statement**

506   The authors declare that they have no known competing financial interests or personal

507   relationships that could influence the work reported in this paper.

508   **Data availability statement**

509   The raw sequencing data generated during the current study is available in the SRA NCBI

510   repository under accession code [TBD]. Other datasets generated and/or analyzed are available

511   in the GitHub repository, https://github.com/madzafv/SongbirdDomesticationGenomics. The

512   corresponding author can provide any additional information required to reproduce the results

513   reported in this study upon reasonable request.
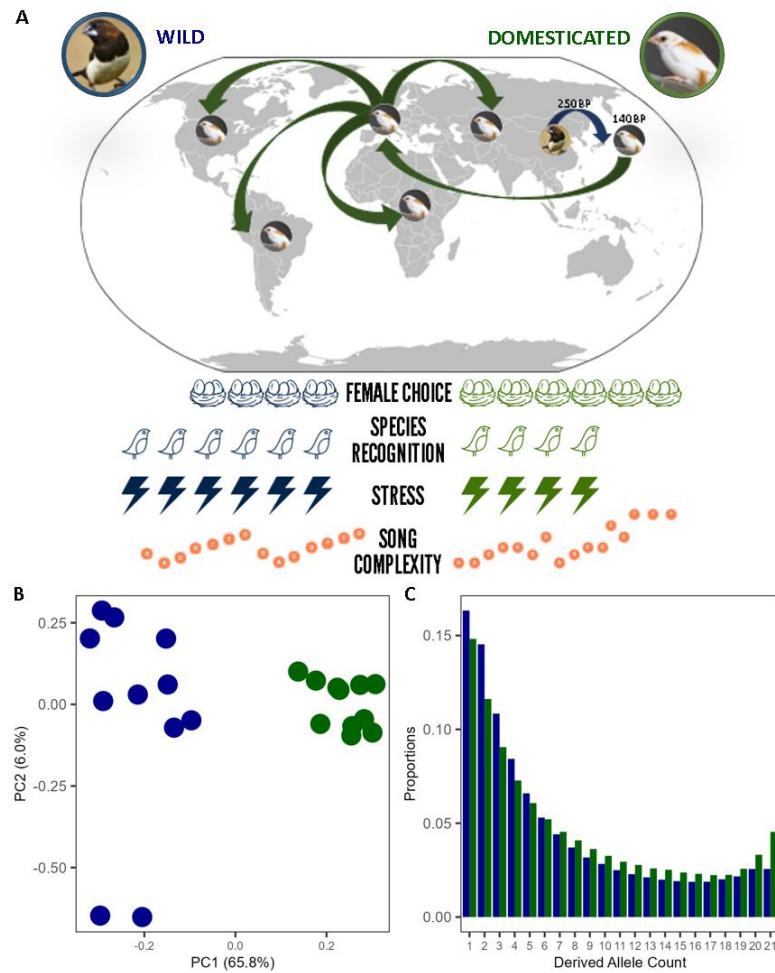
Figures, Tables, and Legends

Fig. 1. The WBM (left; wild; blue) and the BF (right; domesticated; green).(A) Environmental variables and their relationship with stress and birdsong complexity in the wild and domesticated scenarios. (B) PCA of whole genomic genotype likelihoods in WBM and BF. (C) BF and WBM unfolded Site Frequency Spectrums.
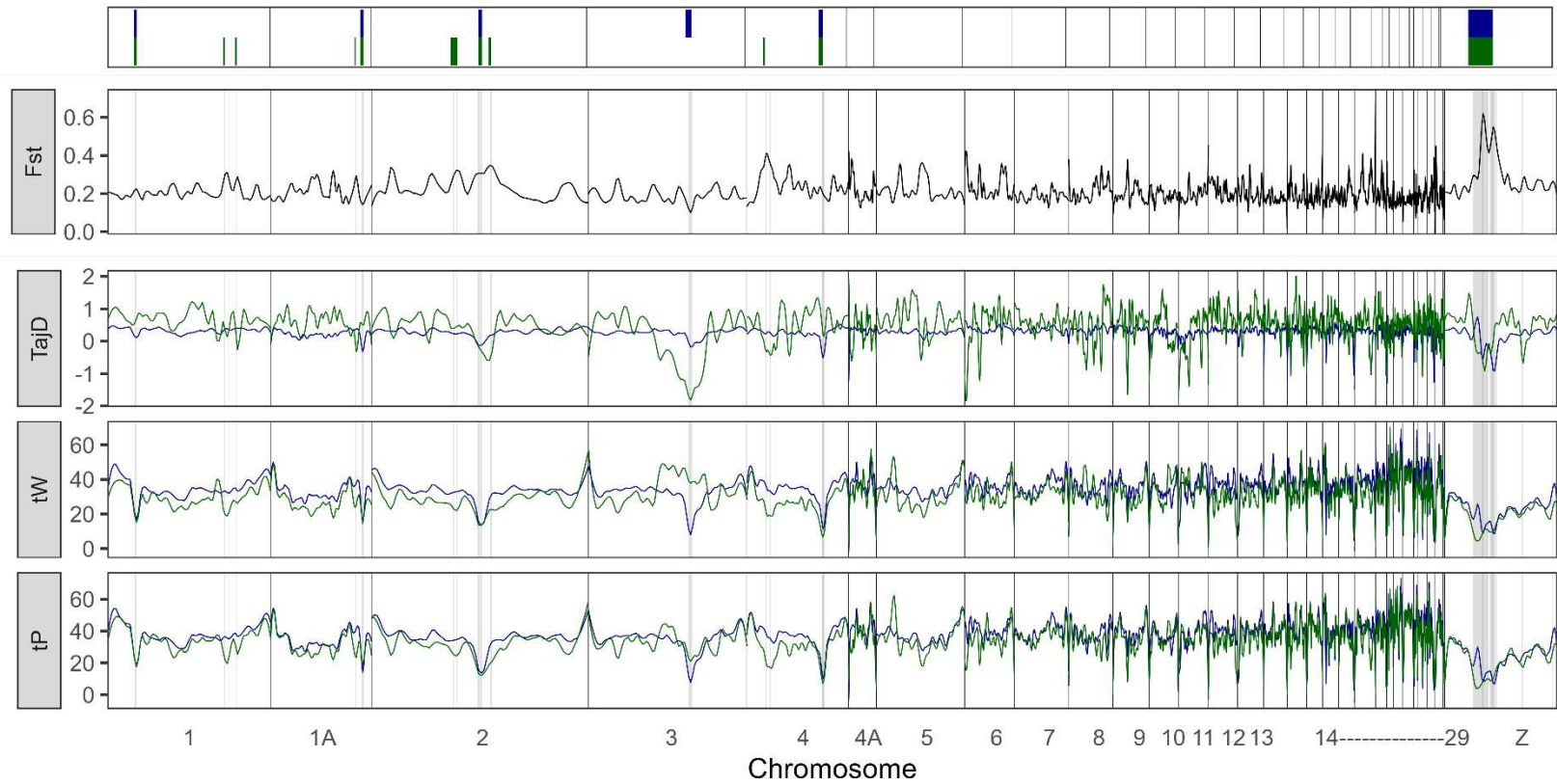
Fig. 2. Genetic differentiation and variability in WBM (wild; blue) and BF (domesticated; green) populations. (**top to bottom**) Schematic representation of defined ROIs enclosing sweep signals discussed in this study; 10kb window measures of differentiation between the two populations, calculated as the averaged Fst across all positions within the window; 10kb window estimations of nucleotide diversity, calculated as the Tajima's D (TajD), Waterson's theta (tW) and Pairwise theta (tW) across all positions within the window.
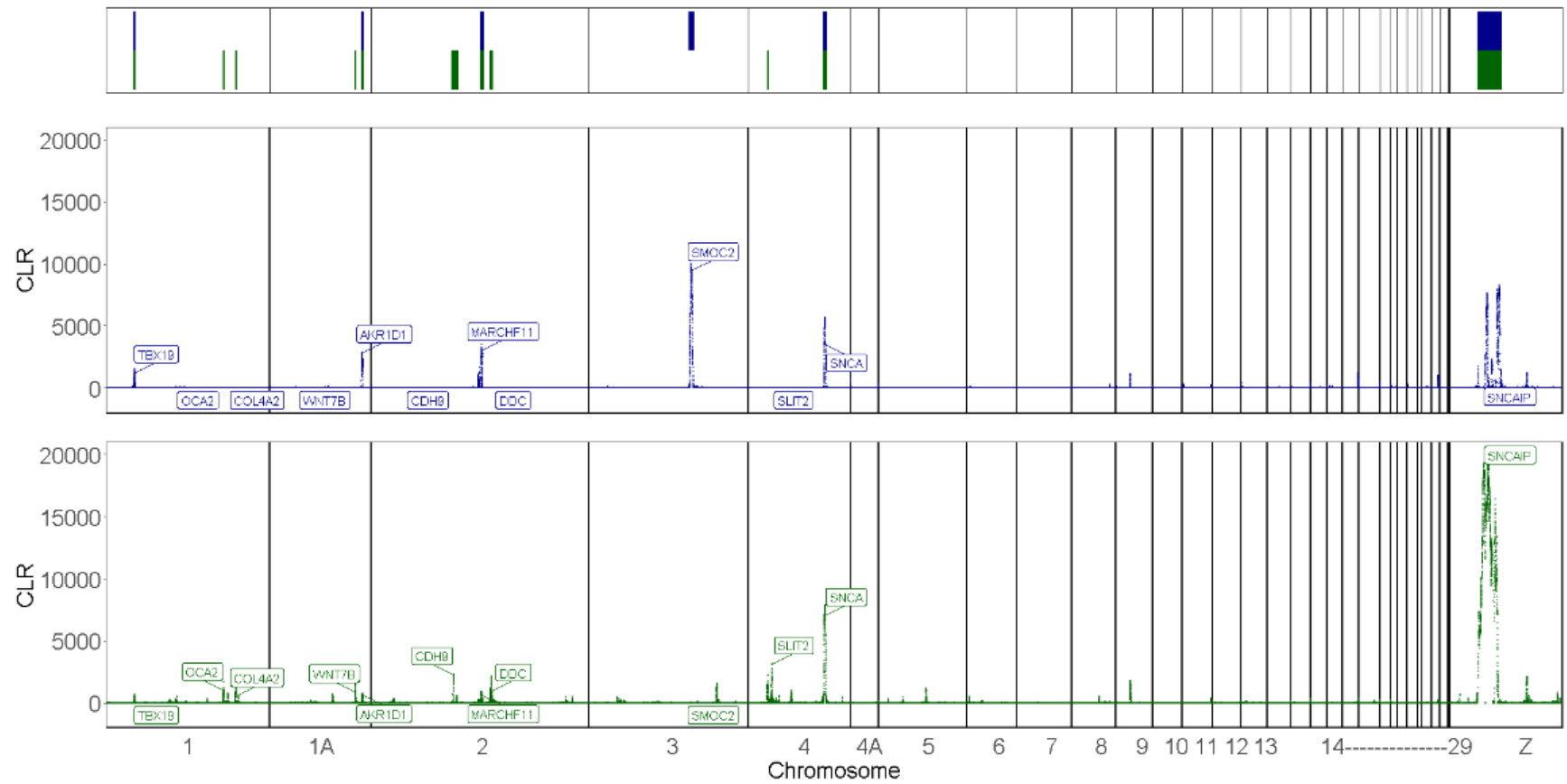
Fig. 3. Selection signals in WBM (wild; blue) and BF (domesticated; green) populations. (**top to bottom**) Schematic representation of defined ROIs enclosing sweep signals discussed in this study; selection signals in WBM and BF populations. Each data point represents a window's composite likelihood of a selective sweep (CLR). Labels highlight the genes within ROIs referred to in the main text.

Table 1. Candidate genes for the domestication syndrome within ROIs

ROI	Sweep Signal		Genes	Gene product; function*
	BF	WBM		
BF1-Chr1			HERC2, OCA2, GABRG3	P protein; melanogenesis [41]
BF2-Chr1			COL4A2, COL4A1	Collagen alpha-2 and alpha-1 (IV) chains; cell adhesion and signaling [73]
BF3-Chr1A			WNT7B	Protein Wnt-7b; neural crest specification and neural cell dendritic arborization [74]
BF4-Chr2			CDH9	Cadherin-9; synapse assembly [75]
BF5-Chr2			DDC	Dopa decarboxylase; dopamine biosynthesis [76]
BF6-Chr4			SLIT2	Slit homolog 2 protein: neural cell axon guidance [77]
BF7-Chr4			SNCA	Alpha-synuclein; synaptic transmission [78]
BF8-ChrZ			SNCAIP	Alpha-synuclein interacting protein-1; synaptic transmission [79]
WBM1-Chr1			TBX19	T-box transcription factor TBX19; pituitary gland development [46]
WBM2-Chr1A			AKR1D1	Aldo-keto reductase family 1 member D1; steroid metabolism [52]
WBM3-Chr2			MARCHF11	E3 ubiquitin-protein ligase MARCHF11; cell differentiation [80]
WBM4-Chr3			SMOC2	SPARC-related modular calcium-binding protein 2; craniofacial morphology [34]

*Noted function of relevance for this study

References

1.   Nimmo, D.G., et al., *Animal movements in fire‑prone landscapes.* Biological Reviews, 2019. **94**(3): p. 981-998.
2.   Moretti, L., et al., *The influence of relationships on neophobia and exploration in wolves and dogs.* Anim Behav, 2015. **107**: p. 159-173.
3.   Price, E.O., *Behavioral development in animals undergoing domestication.* Applied Animal Behaviour Science, 1999. **65**(3): p. 245-271.
4.   Svanberg, I., *Towards a cultural history of the Bengalese Finch (Lonchura domestica).* Der Zoologische Garten, 2008. **77**(5-6): p. 334-344.
5.   Okanoya, K., *The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax.* Ann N Y Acad Sci, 2004. **1016**: p. 724-35.
6.   Dalrymple, P., *The Society Finch: a Wonderful Worker.* AFA Watchbird, 1998. **25**(3): p. 43-43.
7.   Suzuki, K., et al., *Complex Song Development and Stress Hormone Levels in the Bengalese Finch.* Avian Biology Research, 2014. **7**(1): p. 10-17.
8.   Suzuki, K., et al., *Effects of domestication on neophobia: A comparison between the domesticated Bengalese finch (Lonchura striata var. domestica) and its wild ancestor, the white-rumped munia (Lonchura striata).* Behavioural Processes, 2021. **193**.
9.   Suzuki, K. and K. Okanoya, *Domestication effects on aggressiveness: Comparison of biting motivation and bite force between wild and domesticated finches.* Behavioural Processes, 2021. **193**.
10.   Takahasi, M., H. Yamada, and K. Okanoya, *Statistical and Prosodic Cues for Song Segmentation Learning by Bengalese Finches (Lonchura striata var. domestica).* Ethology, 2010. **116**(6): p. 481-489.
11.   Honda, E. and K. Okanoya, *Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica).* Zoological Science, 1999. **16**(2): p. 319-326.
12.   Sánchez-Villagra, M., *The process of animal domestication.* 2022: Princeton University Press.
13.   O'Rourke, T., et al., *Capturing the Effects of Domestication on Vocal Learning Complexity.* Trends Cogn Sci, 2021.
14.   Wilkins, A.S., R.W. Wrangham, and W.T. Fitch, *The "Domestication Syndrome" in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics.* Genetics, 2014. **197**(3): p. 795-808.
15.   Hekman, J.P., et al., *Anterior Pituitary Transcriptome Suggests Differences in ACTH Release in Tame and Aggressive Foxes.* G3 (Bethesda), 2018. **8**(3): p. 859-873.
16.   Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
17.   Andrews, S., *FastQC: a quality control tool for high throughput sequence data.* 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
18.   Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

19.   Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

20.   McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

21.   Korneliussen, T.S., A. Albrechtsen, and R. Nielsen, *ANGSD: Analysis of Next Generation Sequencing Data.* BMC Bioinformatics, 2014. **15**(1): p. 356.

22.   Meisner, J. and A. Albrechtsen, *Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data.* Genetics, 2018. **210**(2): p. 719-731.

23.   Bhatia, G., et al., *Estimating and interpreting FST: the impact of rare variants.* Genome Res, 2013. **23**(9): p. 1514-21.

24.   Watterson, G.A., *On the number of segregating sites in genetical models without recombination.* Theor Popul Biol, 1975. **7**(2): p. 256-76.

25.   Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.* Genetics, 1989. **123**(3): p. 585-95.

26.   Stephan, W., *Selective Sweeps.* Genetics, 2019. **211**(1): p. 5-13.

27.   DeGiorgio, M., et al., *SweepFinder2: increased sensitivity, robustness and flexibility.* Bioinformatics, 2016. **32**(12): p. 1895-7.

28.   Wang, Z., et al., *Temporal genomic evolution of bird sex chromosomes.* BMC Evol Biol, 2014. **14**: p. 250.

29.   Lu, C.W., C.T. Yao, and C.M. Hung, *Domestication obscures genomic estimates of population history.* Molecular Ecology, 2022. **31**(3): p. 752-766.

30.   Suzuki, K., et al., *Decreased fecal corticosterone levels due to domestication: A comparison between the white‑backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica) with a suggestion for complex song evolution.* Journal of Experimental Zoology Part A: Ecological Genetics and Physiology, 2012. **317**(9): p. 561-570.

31.   Lischinsky, J.E. and D. Lin, *Neural mechanisms of aggression across species.* Nat Neurosci, 2020. **23**(11): p. 1317-1328.

32.   Fallahsharoudi, A., et al., *Domestication Effects on Stress Induced Steroid Secretion and Adrenal Gene Expression in Chickens.* Sci Rep, 2015. **5**: p. 15345.

33.   Stryjewski, K.F. and M.D. Sorenson, *Mosaic genome evolution in a recent and rapid avian radiation.* Nature Ecology & Evolution, 2017. **1**(12): p. 1912-1922.

34.   Marchant, T.W., et al., *Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2.* Curr Biol, 2017. **27**(11): p. 1573-1584.e6.

35.   Bannasch, D., et al., *Localization of Canine Brachycephaly Using an Across Breed Mapping Approach.* PLoS ONE, 2010. **5**(3): p. e9632.

36.   Quilez, J., et al., *A selective sweep of >8 Mb on chromosome 26 in the Boxer genome.* BMC Genomics, 2011. **12**(1): p. 339.

37.   Lawson, L.P. and K. Petren, *The adaptive genomic landscape of beak morphology in Darwin's finches.* Mol Ecol, 2017. **26**(19): p. 4978-4989.

38.   Herrel, A., et al., *Evolution of bite force in Darwin's finches: a key role for head width.* J Evol Biol, 2005. **18**(3): p. 669-75.

39. Negro, J.J., M. Carmen Blázquez, and I. Galván, *Intraspecific eye color variability in birds and mammals: a recent evolutionary event exclusive to humans and domestic animals.* Frontiers in Zoology, 2017. **14**(1).

40. Kowalski, E. and R. Bellone, *Investigation of HERC2 and OCA2 SNP for iris color variation in Puerto Rican Paso Fino horses.* Journal of Equine Veterinary Science, 2011. **5**(31): p. 319.

41. Brilliant, M.H., *The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH.* Pigment Cell Res, 2001. **14**(2): p. 86-93.

42. Suarez, P., K. Baumer, and D. Hall, *Further insight into the global variability of the OCA2-HERC2 locus for human pigmentation from multiallelic markers.* Scientific reports, 2021. **11**(1): p. 1-13.

43. Quaio, C.R., et al., *A clinical follow-up of 35 Brazilian patients with Prader-Willi syndrome.* Clinics (Sao Paulo), 2012. **67**(8): p. 917-21.

44. Delahanty, R.J., et al., *Beyond Epilepsy and Autism: Disruption of GABRB3 Causes Ocular Hypopigmentation.* Cell Rep, 2016. **17**(12): p. 3115-3124.

45. Takahasi, M. and K. Okanoya, *Song learning in wild and domesticated strains of white‑rumped munia, lonchura striata, compared by cross‑fostering procedures: domestication increases song variability by decreasing strain‑specific bias.* Ethology, 2010. **116**(5): p. 396-405.

46. Liu, J., et al., *Tbx19,, a tissue-selective regulator of POMC gene expression.* Proceedings of the National Academy of Sciences, 2001. **98**(15): p. 8674-8679.

47. Wang, X., et al., *Whole-genome sequence analysis reveals selection signatures for important economic traits in Xiang pigs.* Scientific Reports, 2022. **12**(1).

48. Patel, T.N., et al., *Social Interactions Increase Activation of Vasopressin-Responsive Neurons in the Dorsal Raphe.* Neuroscience, 2022. **495**: p. 25-46.

49. Goodson, J.L., A.M. Kelly, and M.A. Kingsbury, *Evolving nonapeptide mechanisms of gregariousness and social diversity in birds.* Horm Behav, 2012. **61**(3): p. 239-50.

50. Maney, D.L. and J.L. Goodson, *Neurogenomic mechanisms of aggression in songbirds.* Adv Genet, 2011. **75**: p. 83-119.

51. Crino, O.L., et al., *Evidence for condition mediated trade-offs between the HPA- and HPG-axes in the wild zebra finch.* General and Comparative Endocrinology, 2018. **259**: p. 189-198.

52. Di Costanzo, L., et al., *Structure and catalytic mechanism of human steroid 5beta-reductase (AKR1D1).* Mol Cell Endocrinol, 2009. **301**(1-2): p. 191-8.

53. Schlinger, B.A. and G.V. Callard, *A comparison of aromatase, 5α-, and 5β-reductase activities in the brain and pituitary of male and female quail (C. c. japonica).* Journal of Experimental Zoology, 1987/05/01. **242**(2).

54. Schlinger, B.A., *Steroids in the Avian Brain: Heterogeneity across Space and Time.* J Ornithol, 2015. **156**(Suppl 1): p. 419-424.

55. Smagin, D.A., et al., *Altered Expression of Genes Associated with Major Neurotransmitter Systems in the Reward-Related Brain Regions of Mice with Positive Fighting Experience.* Int J Mol Sci, 2022. **23**(21).

56. Cuvelier, E., et al., *Overexpression of Wild-Type Human Alpha-Synuclein Causes Metabolism Abnormalities in Thy1-aSYN Transgenic Mice.* Front Mol Neurosci, 2018. **11**: p. 321.

57. Emamgholi Begli, H., et al., *Genomic dissection and prediction of feed intake and residual feed intake traits using a longitudinal model in F2 chickens.* Animal, 2018. **12**(9): p. 1792-1798.

58. Li, H., et al., *Genome-Wide Association Studies for Flesh Color and Intramuscular Fat in (Duroc x Landrace x Large White) Crossbred Commercial Pigs.* Genes (Basel), 2022. **13**(11).

59. Abied, A., et al., *Genome Divergence and Dynamics in the Thin-Tailed Desert Sheep From Sudan.* Front Genet, 2021. **12**: p. 659507.

60. Li, X., et al., *A novel obesity model: synphilin-1-induced hyperphagia and obesity in mice.* Int J Obes (Lond), 2012. **36**(9): p. 1215-21.

61. Wang, T., et al., *Analysis of selection signatures on the Z chromosome of bidirectional selection broiler lines for the assessment of abdominal fat content.* BMC Genom Data, 2021. **22**(1): p. 18.

62. Gale, S.D. and D.J. Perkel, *Anatomy of a songbird basal ganglia circuit essential for vocal learning and plasticity.* J Chem Neuroanat, 2010. **39**(2): p. 124-31.

63. Kato, M. and K. Okanoya, *Molecular characterization of the song control nucleus HVC in Bengalese finch brain.* Brain Research, 2010. **1360**: p. 56-76.

64. George, J.M., et al., *Characterization of a novel protein regulated during the critical period for song learning in the zebra finch.* Neuron, 1995. **15**(2): p. 361-72.

65. Nevue, A.A., et al., *Molecular specializations of deep cortical layer analogs in songbirds.* Scientific Reports, 2020. **10**(1).

66. Jin, H. and D.F. Clayton, *Synelfin regulation during the critical period for song learning in normal and isolated juvenile zebra finches.* Neurobiol Learn Mem, 1997. **68**(3): p. 271-84.

67. So, L.Y., S.J. Munger, and J.E. Miller, *Social context-dependent singing alters molecular markers of dopaminergic and glutamatergic signaling in finch basal ganglia Area X.* Behav Brain Res, 2019. **360**: p. 103-112.

68. Hilliard, T., Austin, et al., *Molecular Microcircuitry Underlies Functional Specification in a Basal Ganglia Circuit Dedicated to Vocal Learning.* Neuron, 2012. **73**(3): p. 537-552.

69. Alvarez-Castelao, B. and J.G. Castano, *Synphilin-1 inhibits alpha-synuclein degradation by the proteasome.* Cell Mol Life Sci, 2011. **68**(15): p. 2643-54.

70. Lovell, P.V., et al., *Birdsong "Transcriptomics": Neurochemical Specializations of the Oscine Song System.* PLoS ONE, 2008. **3**(10): p. e3440.

71. Shi, Z., et al., *Dynamic transcriptome landscape in the song nucleus HVC between juvenile and adult zebra finches.* Adv Genet (Hoboken), 2021. **2**(1): p. e10035.

72. Wang, R., et al., *Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners.* J Comp Neurol, 2015. **523**(6): p. 892-906.

73. Meuwissen, M.E., et al., *The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature.* Genet Med, 2015. **17**(11): p. 843-53.

74. Ji, Y., et al., *Wnt Signaling in Neural Crest Ontogenesis and Oncogenesis.* Cells, 2019. **8**(10): p. 1173.

75.   Yamagata, M., X. Duan, and J.R. Sanes, *Cadherins Interact With Synaptic Organizers to Promote Synaptic Differentiation.* Front Mol Neurosci, 2018. **11**: p. 142.

76.   Meiser, J., D. Weindl, and K. Hiller, *Complexity of dopamine metabolism.* Cell Commun Signal, 2013. **11**(1): p. 34.

77.   Bagri, A., et al., *Slit proteins prevent midline crossing and determine the dorsoventral position of major axonal pathways in the mammalian forebrain.* Neuron, 2002. **33**(2): p. 233-48.

78.   Siddiqui, I.J., N. Pervaiz, and A.A. Abbasi, *The Parkinson Disease gene SNCA: Evolutionary and structural insights with pathological implication.* Sci Rep, 2016. **6**: p. 24475.

79.   Kruger, R., *The role of synphilin-1 in synaptic function and protein degradation.* Cell Tissue Res, 2004. **318**(1): p. 195-9.

80.   Morokuma, Y., et al., *MARCH-XI, a novel transmembrane ubiquitin ligase implicated in ubiquitin-dependent protein sorting in developing spermatids.* J Biol Chem, 2007. **282**(34): p. 24806-15.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supp.Fig112.docx
- Supp.Table1.docx
- Supp.Text1.docx
- Supp.Text2.docx
- Supp.Data1.xlsx
- Supp.Data2.xlsx
- Supp.Data3.xlsx