Received: 7 March 2023 | Accepted: 3 May 2023

DOI: 10.1111/age.13331

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Whole genome sequencing identifies candidate genes and mutations that can explain diluted and other colour varieties of domestic canaries (*Serinus canaria*)

Samuele Bovo[1] | **Anisa Ribani**[1] | **Valerio Joe Utzeri**[1] | **Valeria Taurisano**[1] | **Gianmaria Bertarini**[2] | **Luca Fontanesi**[1]

[1]Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Bologna, Italy

[2]Federazione Ornicoltori Italiani, Piacenza, Italy

Correspondence
Luca Fontanesi, Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy.
Email: luca.fontanesi@unibo.it

Funding information
Federazione Ornicoltori Italiani; Università di Bologna

Abstract

The domestic canary (*Serinus canaria*) is one of the most common pet birds and has been extensively selected and bred over the last few centuries to constitute many different varieties. Plumage pigmentation is one of the main phenotypic traits that distinguish canary breeds and lines. Feather colours in these birds, similarly to other avian species, are mainly depended on the presence of two major types of pigments: carotenoids and melanins. In this study, we exploited whole genome sequencing (WGS) datasets produced from five canary lines or populations (Black Frosted Yellow, Opal, Onyx, Opal × Onyx and Mogno, some of which carrying different putative dilute alleles), complemented with other WGS datasets retrieved from previous studies, to identify candidate genes that might explain pigmentation variability across canary breeds and varieties. Sequencing data were obtained using a DNA pool-seq approach and genomic data were compared using window-based F_{ST} analyses. We identified signatures of selection in genomic regions harbouring genes involved in carotenoid-derived pigmentation variants (*CYP2J19*, *EDC*, *BCO2* and *SCARB1*), confirming the results reported by previous works, and identified several other signatures of selection in the correspondence of melanogenesis-related genes (*AGRP*, *ASIP*, *DCT*, *EDNRB*, *KITLG*, *MITF*, *MLPH*, *SLC45A2*, *TYRP1* and *ZEB2*). Two putative causative mutations were identified in the *MLPH* gene that may explain the Opal and Onyx dilute mutant alleles. Other signatures of selection were also identified that might explain additional phenotypic differences between the investigated canary populations.

KEYWORDS
avian species, breeding, domestication, feather colour, indel, MLPH, mutation, pigmentation, SNP, songbird

INTRODUCTION

The domestic canary (*Serinus canaria*) is one the most common caged songbirds, currently including three main breeds – song canaries, colour canaries and type canaries – each one characterised by several varieties and derived selected lines. Domestication of canaries started from the native wild populations of the Canary Islands and the islands of the Madeira and Azores archipelagos, in the North Atlantic Ocean (Parsons, 1987). Canaries with an originally grey-green plumage were imported into Europe by Spanish and Portuguese

sailors from the beginning of the fifteenth century, mainly because of their beautiful song (Parsons, 1987). The songbird was considered a luxury bird, frequently present in many European courts, and its breeding became fashionable among the aristocracy. By the early sixteenth, the canary started to be massively imported and traded in many countries, including Italy, Germany and England, where canary breeding (or canariculture) was first established (Parsons, 1987). The transition from grey-green to yellow plumage is thought to represent one of the first evidences of the domestication process of this bird (Birkhead et al., 2004; Parsons, 1987). There is much uncertainty as to when and where the yellow canaries appeared and if they were produced through selection and breeding: it seems that yellow canaries were present in the sixteenth or seventeenth centuries in Spain, Italy, Germany and England (Birkhead et al., 2004; Parsons, 1987; Perez-Beato, 2008). Subsequently, several other genetic variants of plumage colourations appeared in Europe, marking the beginning of coloured canary breeding as a hobby. It is only more recently, in 1930, that the artificial hybridisation of canaries with the red siskin (*Spinus cucullatus*) introduced the red factor into the species (Perez-Beato, 2008). Thus, selective breeding has introduced several other plumage colourations that characterise and distinguish many varieties of coloured canaries.

Feather colours in these birds, similarly to other avian species, mainly depend on the presence of two major types of pigments: carotenoids and melanins (Perez-Beato, 2008). Carotenoids are responsible for the yellow or red/orange ground colours (that also include the white colour) and derive from the metabolic processes of carotenes and xanthophylls, which are obtained from the diet, and their deposition in the integument (Brush, 1990; McGraw, 2006). The identification of genes acting on carotenoids has been quite challenging, considering that the expression of carotenoid-based colouration is strongly affected by the environment as carotenoid precursors are available only from the diet. The genetic basis for red colouration in red factor canaries was elucidated taking advantage of comparative genome analyses between red siskins (*Spinus cucullata*), *Serinus canaria* and their hybrids, the 'red factor' canaries (Lopes et al., 2016). Two genomic regions introgressed into the red factor canaries, derived from *S. cucullata*, are required for the red colouration. One of these regions contains a gene encoding a cytochrome P450 enzyme (*CYP2J19*), that produces a carotenoid ketolase, which is responsible for the transformation of yellow carotenoids to red ketocarotenoids in canaries with red feathers (Lopes et al., 2016). Another introgressed region contains the epidermal differentiation complex (*EDC*), a cluster of genes involved in the development of the integument (Lopes et al., 2016). Beta-carotene oxygenase 2 (*BCO2*) is another key gene in the carotenoid metabolism that has been linked to carotenoid-based colouration in several birds, including canaries, where it mediates sexual dichromatism (Gazda, Araújo, et al., 2020a) and is also responsible for carotenoid-based bare part colouration in Urucum canaries (Gazda, Toomey, et al., 2020b). Scavenger receptors are another important gene family acting on carotenoids as they mediate the cellular uptake of such molecules. A splice donor site mutation in the scavenger receptor B1 (*SCARB1*) gene is responsible for the white recessive phenotype in canaries (Toomey et al., 2017).

Melanin pigments include eumelanin (black/brown pigments) and pheomelanin (yellow/red pigments) that are responsible for dark-coloured varieties. Their distribution has also a role in producing spatial patterns (e.g. stripes). Unlike carotenoids, these pigments are biochemically produced by the canaries themselves as endogenous products of the metabolism, based on their genetic information, and do not have to be acquired from the diet. Both eumelanin and pheomelanin pigments are synthesised and accumulated in melanosomes starting from the hydroxylation of tyrosine catalysed by tyrosinase (*TYR*), followed by several other enzymatic steps (Kondo & Hearing, 2011). The switch between eumelanin and pheomelanin synthesis is regulated by two other key genes: melanocortin 1 receptor (*MC1R*) and its ligand, agouti signaling protein (*ASIP*). Moreover, alongside melanogenic enzymes, many other genes contribute to melanogenesis via affecting melanocyte morphology, development and migration or affecting the melanosomal structure and function (Baxter et al., 2019; Lamorux et al., 2010).

Some dilute melanic forms have been selected in canaries and are combined to constitute different colour varieties, including Agate, Isabelle and Pastel canaries, among several others (Perez-Beato, 2008). Most of the dilute factors have been only phenotypically described and their correct identification is sometimes difficult when they are combined with other modifier genetic factors. The Opal mutation appeared in 1949 in Germany and only subsequently became very popular among fancy canary breeders (Perez-Beato, 2008). The Opal mutation has been described as an extreme dilution factor that affects both types of melanin pigments. This mutation reduces the level of phaeomelanin (brown) and shifts the eumelanin black towards the underside of the feather, together with its general reduction. Onyx is another mutation that appeared in Spain in 1983–1984, which was described as a variant of the Opal mutatation (Perez-Beato, 2008). According to Perez-Beato (2008), the *Opal* locus has the following allele series: the normal dominant allele *L*, the recessive allele *l* responsible for the Opal mutation, and finally the recessive allele *l′* responsible for the Onyx mutation. These two mutated alleles have been considered co-dominant to each other. Mogno is another more recently appeared variety (2012–2013), quite similar to Opal. Opal and Mogno differ as the latter mutant shows the complete absence of phaeomelanin

whereas the black eumelanin, although lowered, is normally localised on the upper side of the feather. There is no agreement between fancy breeders as to whether Mogno is considered another allele of the previous series or a mutant of a distinct locus. Very active debates are underway in the canariculture sector about the genetic basis of these and many other mutations that might be phenotypically distinguishable only by expert breeders. Figure 1 shows canaries carrying these mutated forms.

To shed some light on this interesting debate, we adopted a DNA pool-seq strategy to identify signatures of selection that might refer to the phenotypic difference in the Opal, Onyx and Mogno dilution factors. To fully understand the mechanism at the basis of these colourations, considering also that the individual canaries included in the DNA pools might not be potentially completely uniform for other carotenoid-derived genetic factors and patterns within the same group of investigated canaries, we added to the study a group of canaries with a phenotype similar to the wild type and other canaries derived from an Opal × Onyx cross. Our sequencing data were also integrated with publicly available whole genome resequencing datasets of additional canary varieties and populations (Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017) to further confirm the identified variants and extract additional signatures of selection in genomic regions carrying pigmentation-related genes that would contribute to the explanation of the phenotypic variability in different canary populations and in the Opal, Onyx and Mogno variants. Other genes affecting pigmentation were also identified in addition to those that might explain the targeted diluted phenotypes in the investigated canary varieties and populations.

MATERIALS AND METHODS

Canaries

A total of 31 canaries were included in this study. These birds were from several Italian licensed fancy breeders and included six Black Frosted Yellow canaries (BFY), 10 Opal canaries (OPA; three Black Opal Mosaic Red; one Black Opal White Dominant/Perl and six Black Opal Intensive Red), four Black Onyx White Dominant canaries (ONY), five crossbred Opal × Onyx canaries (OXO; phenotypes similar to two Black Onyx White Dominant, one Black Onyx Mosaic Yellow and two Black Onyx Mosaic Red) and six Mogno canaries (MOG; two Brown Mogno White Dominant, two Black Mogno Mosaic Yellow and two Black Mogno White Dominant/Brown). Figure 1 shows the colour types of the investigated canary varieties putatively carrying dilute genetic factor(s). Birds were not raised or treated in any way for the purpose of this study, therefore no ethical questions were relevant.

DNA samples and sequencing

DNA was extracted from the base or calamus of feathers provided by the fancy breeders. DNA was obtained using the Wizard Genomic DNA Purification kit (Promega Corporation). A DNA pool-seq approach was adopted to obtain genomic information. A total of five DNA pools were constructed using, for each individual included in the pool, equimolar DNA concentration (Table S1). Five genomic libraries of 300–400 bp in size were constructed and sequenced on a BGISeq500 machine, following the provider's protocol. About 56 Gbp sequenced paired-end reads of 150 bp in length were obtained for each DNA pool.

Publicly available whole genome sequencing datasets

Six additional whole genome sequencing (WGS) datasets of *S. canaria* were downloaded from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/) resource. These datasets were produced by previous studies (Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017) applying a DNA pool-seq approach like our approach. Datasets were from six additional canary types including Lipochrome Red (red plumage; LR), Black Intense Red (red plumage; BIR), Gibber Italicus (yellow plumage; GIB), White Recessive (white plumage; WR), one wild population from the Canaries (WT) and the Urucum canary (red bill and legs; red or

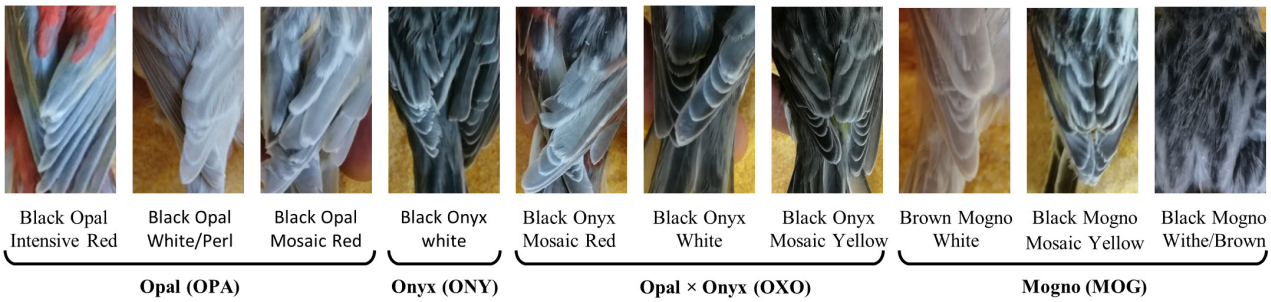

FIGURE 1 Different diluted colour types of the investigated canary varieties sequenced in this study.

yellow plumage; URU). More details on these additional datasets are given in Table S1.

Data quality controls, read mapping and variant detection

Sequenced reads were inspected using FASTQC v.0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) that highlighted very high-quality reads. No other filtering procedures were adopted. Reads were mapped to the *S. canaria* reference genome (GCA_007115625.1; cibio_Scana_2019) with BWA-MEM tool 0.7.17 (Li & Durbin, 2009) with default parameters. PICARD v.2.1.1 (https://broadinstitute.github.io/picard/) was used to remove duplicated reads. Variant calling and filtering were performed with GATK4 (Poplin et al., 2017), HAPLOTYPECALLER (indel-size-to-eliminate-in-ref-model equal to 10) and VARIANTFILTRATION (hard-filter; basic filtering thresholds for SNPs and insertions/deletions or INDELs, as suggested in the manual), respectively. We further retained only bi-allelic variants covered by reads in all the 11 investigated DNA pools (five newly generated datasets and six datasets retrieved from ENA). Allele frequencies were estimated by counting the number of reads supporting the reference and alternative alleles. The VARIANT EFFECT PREDICTOR v.95.0 tool (McLaren et al., 2016) was used to annotate variants based on gene information reported in the cibio_Scana_2019 NCBI's GFF file.

Detection of signatures of selection and annotation

The fixation index (F_{ST}) statistics was used to identify signatures of selection in the analysed canary populations (Bovo et al., 2020). Signatures of selection were evaluated in adjacent genome windows of size 20 kb, for a total of 56 373 computed windows. Only SNPs were evaluated and windows with <20 SNPs (about 2000 windows) were excluded. On average, windows contained 230 ± 132 SNPs (median = 207) in the 11 datasets. The F_{ST} index was computed in each window according to the formula proposed by Karlsson et al. (2007). Briefly, F_{ST} values were calculated for each SNP and then averaged within the window. As not all SNPs segregated in all 11 populations, only SNPs actually segregating and not fixed in the populations under comparison (e.g. DNA pool A vs. DNA pool B) were included in the computation of the averaged F_{ST} value of the windows. In each comparison, windows presenting an F_{ST} value above the 99.9th percentile of the related distribution were considered candidate regions for signatures of selection. All pairwise comparisons between the 11 populations (Table S1) were carried out. The thresholds used to detect outlier windows are reported in Figure S1.

Outlier windows were annotated with BEDTOOLS v.2.17.0 (Quinlan & Hall, 2010) considering the gene features retrieved from the cibio_Scana_2019 NCBI GFF file. The relevance of annotated genes was evaluated through a detailed analysis of the scientific literature. Sequence data were also manually inspected to identify candidate mutations affecting the targeted colour phenotypes. PANTHER-PSEP (position-specific evolutionary preservation) (Tang & Thomas, 2016) was used to predict the impact of missense mutations on the protein function of the selected candidate genes. PANTHER-PSEP calculates the length of time (in millions of years, MY) a given amino acid has been preserved in the lineage, leading to the protein of interest. The longer the preservation time, the greater the likelihood of functional impact (converted to a probability of deleterios effect; Pdel): if the preservation time is greater than 450 MY, the amino acid substitution is 'probably damaging' and is very likely to disrupt the protein function; if the preservation time is between 200 and 450 MY, the amino acid substitution is 'possibly damaging' and may disrupt the protein function; if the preservation time is less than 200 MY, the amino acid substitution is 'probably benign' and is more likely not to disrupt the protein function (Tang & Thomas, 2016).

Pipelines were developed either in PYTHON 3.0 or in R 4.2.0 (R Core Team, 2018).

RESULTS

Sequencing data and detected variants

Sequencing of the five DNA pools produced a total of ~1.9 billion of reads, with an average number of sequenced read pairs per pool equal to ~388.9 million leading to a depth of sequencing of about 46×. Data obtained from ENA had a depth of sequencing that ranged from ~19× to ~30×. The percentage of duplicated reads was around 15 and 6% for in-house produced and ENA-retrieved WGS data, respectively. Sequencing statistics are given in Table S1. A total of 15 057 112 high-quality biallelic variants were retrieved from the analysis of the 11 DNA pools, including 13 070 505 SNPs (87%) and 1 986 607 indels (13%).

F_{ST} signals: Overview of the identified signatures of selection

F_{ST} statistics were applied to detect signatures of selection between two populations using a window-based approach on DNA-pool sequencing datasets. We first compared the wild population (WT) against the other five canary populations derived from the publicly available datasets (Lipochrome Red, LR; Black Intense Red, BIR; Gibber Italicus, GIB; White recessive, WR; and

Urucum, URU) and against our five newly sequenced canary populations (Black Frosted Yellow, BFY; Opal, OPA; Onyx, ONY; Opal × Onyx, OXO; and Mogno, MOG). Manhattan plots of these 10 pairwise comparisons are reported in Figure 2. These two first groups of comparisons were useful to evidence signatures of selection derived from several colour mutants, considering that WT can be regarded as the representative of the ancestral domestic canary (Lopes et al., 2016; Toomey et al., 2017). Some of the signatures of selection that emerged in the comparison between WT and the other publicly available datasets were only in part already evidenced in previous studies (Gazda, Araújo, et al., 2020a; Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017). Then, all other possible pairwise comparisons between populations (n. 45) were carried out, also including those between the newly sequenced populations (BFY, OPA, ONY, OXO and MOG). Figure 3 reports a few Manhattan plots derived from relevant pairwise comparisons that involved some of these later five canary populations. Manhattan plots of all other remaining comparisons are reported in Figure S2.

Many other signatures of selection were identified, including the candidate regions which might explain the targeted dilute factors that determine the Opal, Onyx and Mogno varieties. A summary of the pigmentation related genes detected in all comparisons is provided in Figure 4. Figure 5 shows the allele frequency of the polymorphic sites in the major gene regions identified across all canary populations. Table S2 includes the detailed information of all outlier genomic regions with also the list of annotated genes.

The subsequent paragraphs describe in more details the results that (1) confirmed what was reported by previous studies on pigmentation in canaries, (2) identified additional signatures of selection involving pigmentation genes not directly related to the targeted dilute factors and (3) identified signatures of selection that might explain the targeted diluted phenotypes in the investigated Opal, Onyx and Mogno varieties. In addition, a few other signatures of selection emerged that, however, may contain additional candidate or novel genes involved in feather development and colouration or that might affect other differentiating characteristics between canary varieties.

Pigmentation genes already detected in other studies in canaries

The F_{ST} pairwise analyses that involved WT canaries in contrast with all other 10 canary populations identified in most comparisons at least one of the few candidate genes already described in previous studies in this bird species (Gazda, Araújo, et al., 2020a; Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017, 2022). For example, a signature of selection including the *CYP2J19* gene (positioned at NW_022041653.1:

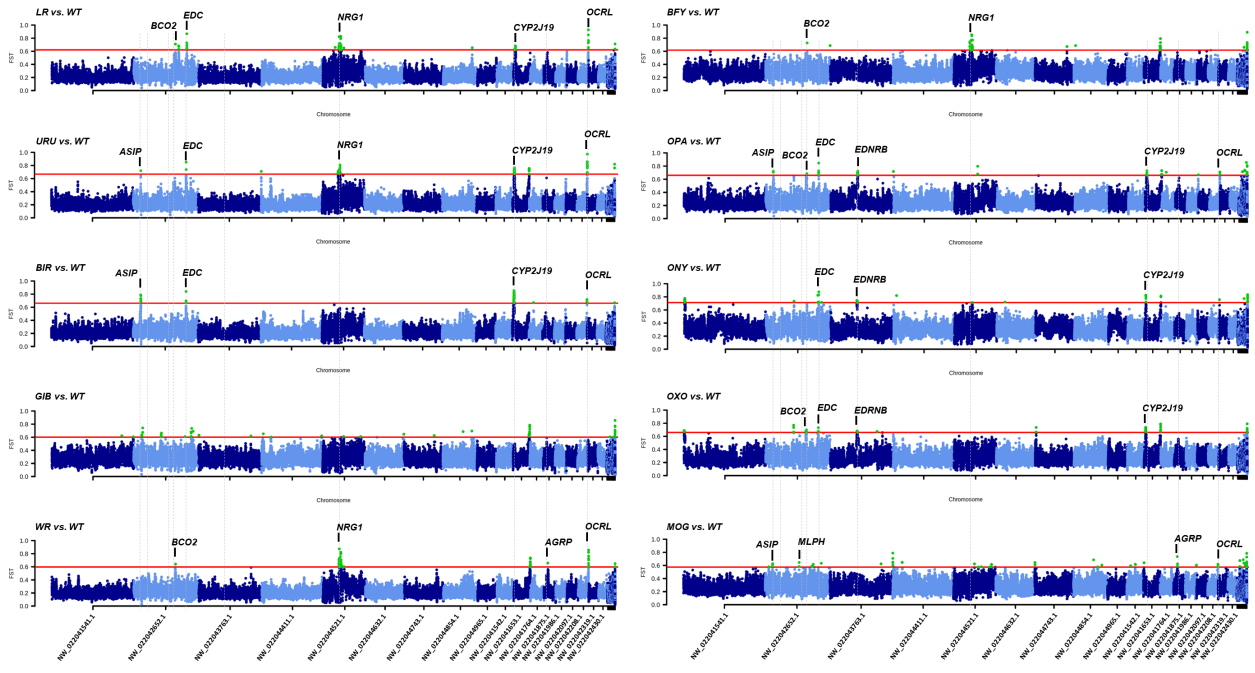

FIGURE 2 Manhattan plots of the pairwise F_{ST} comparisons between whole genome sequencing (WGS) datasets obtained from wild canaries (WT) and WGS datasets obtained from all 10 other canary varieties and populations (on the left, those retrieved from previous studies; on the right, those sequenced in this study). Each dot represents a genomic window. Red lines indicate the threshold in F_{ST} analyses used to highlight signatures of selection (see Figure S1). BFY, Black Frosted Yellow; BIR, Black Intense Red; GIB, Gibber Italicus; LR, Lipochrome Red; MOG, Mogno; OPA, Opal; ONY, Onyx; OXO, Opal × Onyx; WR, White Recessive; WT, wild canaries; URU, Urucum. Chromosome scaffolds are ordered according to their size. Unassembled scaffolds are reported at the right end of each plot.

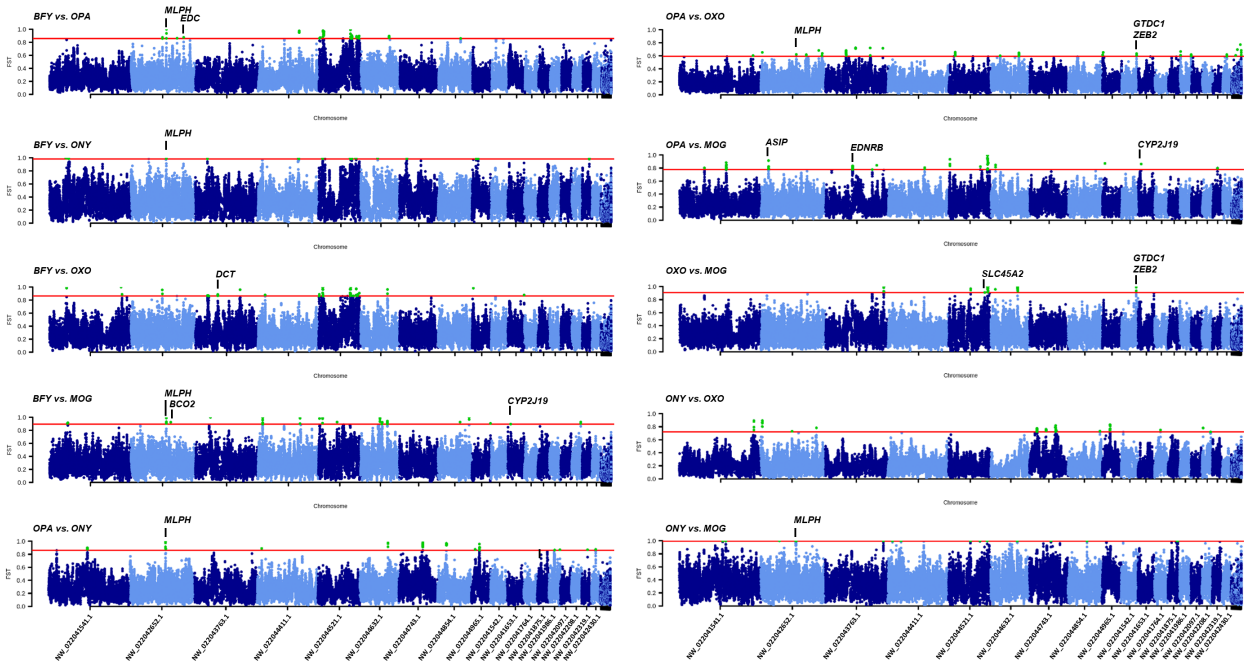

FIGURE 3 Manhattan plots of the pairwise F_{ST} comparisons between whole genome sequencing datasets obtained from varieties and populations sequenced in this study. Each dot represents a genomic window. Red lines indicate the threshold in F_{ST} analyses used to highlight signatures of selection (see Figure S1). BFY, Black Frosted Yellow; MOG, Mogno; OPA, Opal; ONY, Onyx; OXO, Opal × Onyx. Chromosome scaffolds are ordered according to their size. Unassembled scaffolds are reported at the right end of each plot.

3.46–3.48 Mb) was identified in six contrasts (Figure 2): as expected, three were against red varieties for which data were retrieved from ENA (Lipochrome Red, Urucum and Black Intense Red) and three were from the new sequenced populations (Opal population that included nine red or red related canaries; Onyx population that included black Onyx white birds; and the crossbred Opal × Onyx). Another region, containing the *EDC* gene complex (positioned at NW_022042652.1: 98.7–98.8 Mb), was identified in the same six out of 10 pairwise comparisons with the WT group, supporting that both *CYP2J19* and *EDC* are co-involved in determining red pigmentation in canaries (Lopes et al., 2016), as also confirmed by the overlapping pattern of allele frequencies of the two gene regions (Figure 5). Over the 55 pairwise comparisons, *CYP2J19* and *EDC* regions emerged in 20 and 18 of them, respectively, 15 of which highlighted both regions together (Figure 4).

The *BCO2* gene region emerged in five out of 10 comparisons with the WT canaries, including the Lipochrome Red and White Recessive (from ENA) and the Black Frosted Yellow, Opal and Opal × Onyx populations. The comparison between WT and Urucum produced an F_{ST} of ~0.6 in this region (just below the threshold). As expected, this region emerged with the most relevant peak in the comparison between Lipochrome Red and Urucum, where a missense mutation (NW_022042652.1:g.75258192G>A; p.R413H; XP_009096268.1) determines a carotenoid-based pigmentation of bare parts (bright red bills and legs; Gazda, Toomey, et al., 2020b). PANTHER-PSEP analysis supported the functional relevance (probably damaging) of this mutation (Table S3), as already reported (Gazda, Toomey, et al., 2020b). We confirmed the presence of this mutation in the Urucum canaries and we also identified another missense mutation (p.G443S) that was present only in the Gibber Italicus variety (Table S3). The probably damaging role of the p.G443S mutation was supported by the PANTHER-PSEP scores (Table S3).

The *SCARB1* gene (NW_022042652.1: 27.46–27.48 Mb) emerged only in the pairwise analysis between White Recessive and Gibber Italicus, in agreement to Toomey et al. (2017), and in line with the fact that the splice donor site mutation in this gene (NW_022042652.1:g.27474120A>C), that disrupts the carotenoid uptake in white canaries, was fixed in the White Recessive variety (Table S3). However, this mutation was also identified in some of the newly sequenced populations (Table S3): Black Frosted Yellow (frequency = 0.45), Opal (frequency = 0.20) and Mogno (frequency = 0.27).

Other pigmentation genes not directly related to the targeted dilute factors

The F_{ST} pairwise analyses that involved WT canaries in contrast with six other canary populations (including all other populations with datasets derived from ENA and the Black Frosted Yellow populations that we sequenced) and all other pairwise combinations between these seven populations evidenced several other pigmentation genes not identified in previous studies in canaries. As these

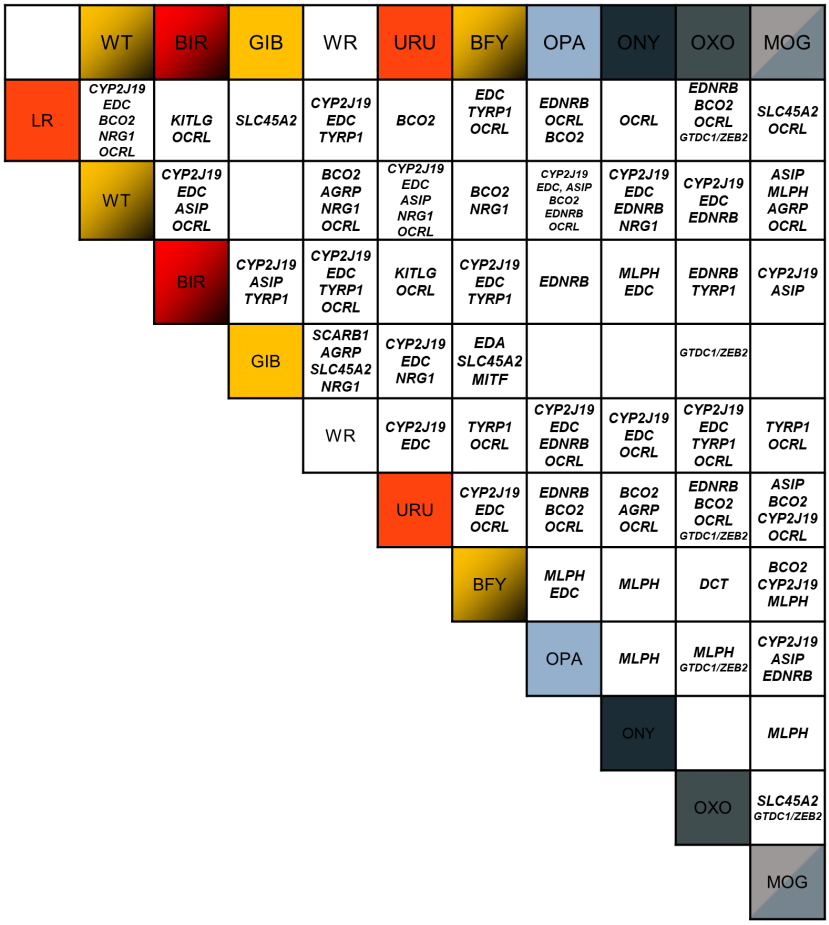

FIGURE 4 Summary of the signatures of selection identified in all pairwise comparisons and including pigmentation related genes. Only regions with a F_{ST} above the related F_{ST} thresholds are reported. Acronyms of the canary breeds and populations are reported in the legend to Figure 2 and in Table S1.

populations have not been selected for any dilute factors [and therefore can be putatively considered as non-dilute carrying factor(s) populations], these genes might not be directly involved in determining the Opal, Onyx and Mogno colour varieties. This question is important as in some pairwise analyses based only on the Opal, Onyx, Opal × Onyx and Mogno populations [defined as putative dilute carrying factor(s) populations] a few emerged genes were the same as those that also emerged in the comparisons between non-dilute carrying factor(s) populations. These results might provide some indirect evidence that would tend to exclude these common genetic factors [i.e. evidenced both in the comparisons involving non-dilute carrying factor(s) populations and dilute carrying factor(s) populations] from being involved in the targeted diluted phenotypes represented by the Opal, Onyx and Mogno lines.

However, based on the applied experimental design we cannot completely exclude that variants in these additional genes could not act as modifiers of the dilute factors, in particular if different mutated haplotype structures and frequencies could genetically characterise one or more newly sequenced populations, a scenario that however cannot be completely disentangled by the F_{ST} window-based scan. This seems the case for the *ASIP* gene (NW_022042652.1:13.98–14.00 Mb) region detected in several pairwise comparisons based on the analysed canary DNA pools. The *ASIP* gene region emerged not only using data derived from ENA (WT vs. Urucum, WT vs. Black Intense Red, and Black Intense Red vs. Gibber Italicus) but also from other five pairwise comparisons that included one or both newly sequenced groups of canaries (WT vs. Opal, WT vs. Mogno, Opal vs. Mogno, Black Intense Red vs. Mogno, and Urucum vs. Mogno). According to the three-way relationships that could be inferred from these comparisons (Figure 4), it seems that more than one functional *ASIP* allele form may be present in the investigated populations: one form may be carried by WT canaries, another form might be prevalent in Urucum and Black Intense Red and two other forms could be present in Opal and Mogno, respectively. In the *ASIP* gene, a missense mutation (p.T63S; 'possibly damaging'; Table S3) that segregated in some of the investigated canary populations (with frequency ranging from 0.10 to 0.52; Table S3) did not explain the signatures of selection that emerged in the comparisons mentioned above.

The region including the agouti-related protein (*AGRP*) gene (NW_022041875.1:7.02–7.04 Mb), which in birds is involved in pheomelanin synthesis in response to oxidative status (Rodríguez-Martínez & Galván, 2020), emerged in two comparisons with WT (i.e. against White Recessive and against Mogno) and in a comparison between Urucum and Onyx. It is, however, worth

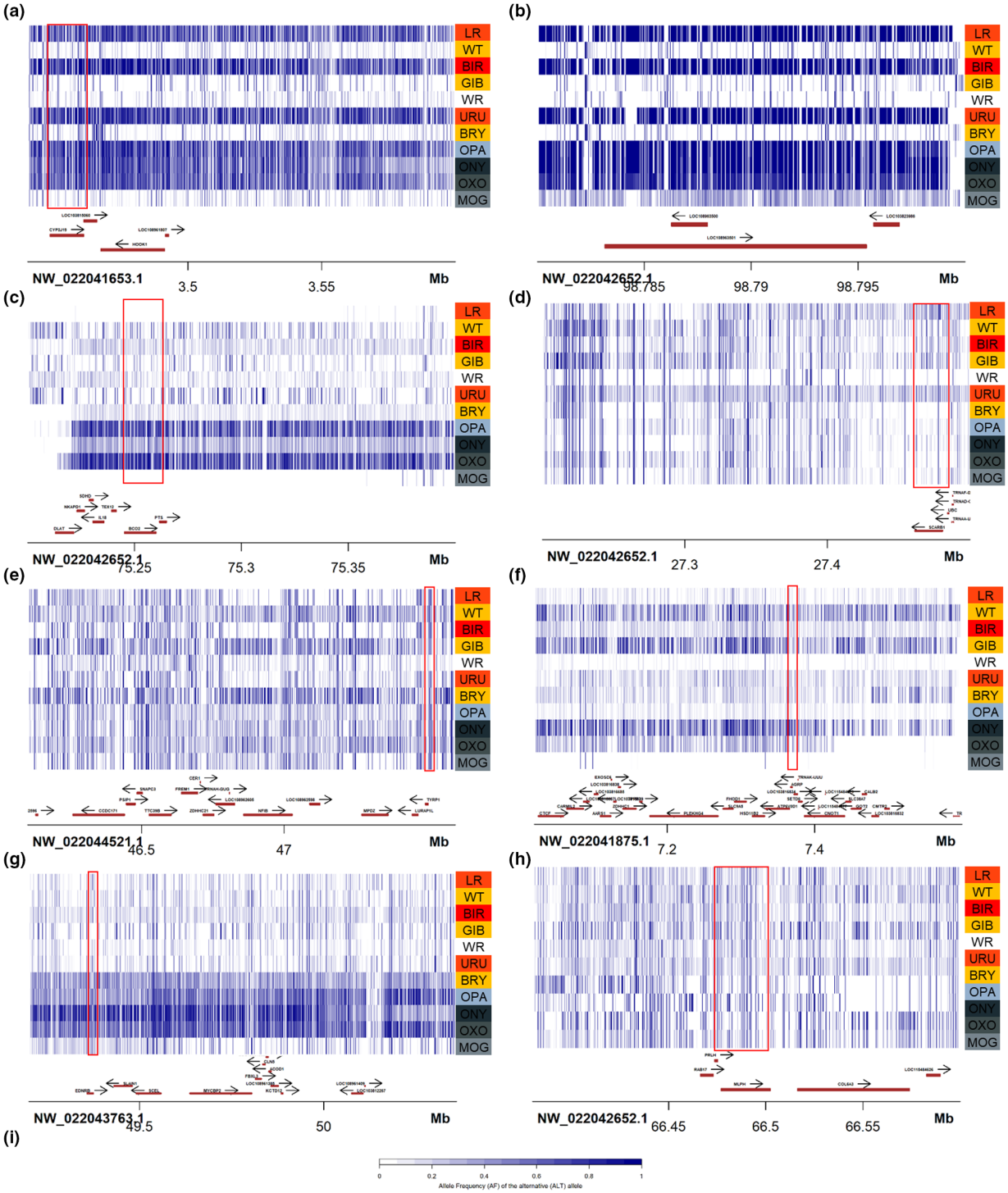

FIGURE 5 Allele frequency patterns obtained from the 11 whole genome sequencing datasets, each from one of the analysed canary breeds/populations, reported for the genomic regions that harbour some pigmentation relevant genes highlighted in this study. Each bar represents an SNP coloured according to the frequency of the alternative allele. The red boxes highlight the gene region: (a) *CYP2J19*, (b) *EDC* gene complex, (c) *BCO2*, (d) *SCARBI*, (e) *TYRP1*, (f) *AGRP*, (g) *EDNRB* and (h) *MLPH*. Acronyms of the different canary breeds/populations are explained in the legend to Figure 2 and in Table S1.

mentioning that in a few other comparisons, this gene region was just below the adopted thresholds (data not shown).

Another region including a well-known gene involved in the melanin biosynthetic pathway, tyrosinase related protein 1 (*TYRP1*), located on scaffold NW_022044521.1:47.49–47.51 Mb, was evidenced in three comparisons based only on ENA-derived datasets (Lipochrome Red vs. White recessive; Gibber Italicus vs. Black Intense Red; and White Recessive vs. Black

Intense Red) or in six comparisons between ENA datasets and newly sequenced varieties (three that comprised Black Frosted Yellow vs. Lipochrome Red, Black Intense Red and White Recessive; two comparisons that comprised White recessive vs. Opal × Onyx and Mogno; and a comparison between Black Intense Red and Opal × Onyx). Inspecting the coding region of the canary reference genome, we noted that the assembled *TYRP1* gene included a frameshift mutation (p.S280fs; XM_018918325.2, XP_018773870.1; Figure S3) that we also identified in the sequenced DNA pools of Gibber Italicus, White Recessive and Urucum varieties but not in all other sequenced populations (Table S3). Therefore, adding this information to the signatures of selection data based on the F_{ST} results, it seems that more than one allele form, probably with regulatory effects (as no other obvious mutations were identified in the coding regions), might affect *TYRP1* function or structure in the sequenced varieties.

One F_{ST} peak, that included the dopachrome tautomerase (*DCT*) gene (NW_022043763.1:42.11–42.14 Mb), which encodes for a key enzyme in melanin biosynthesis, was evidenced just in the comparison between Black Frosted Yellow and Opal × Onyx canaries.

The region that encompassed the solute carrier family 45 member 2 (*SLC45A2*) gene (NW_022044521.1:68.04–68.06 Mb), which encodes a transporter protein that mediates melanin synthesis, emerged in five comparisons: Gibber Italicus vs. Lipochrome Red, White Recessive and Black Intense Red; Mogno vs. Lipochrome Red and Opal × Onyx. No obvious mutations that could explain the observed signatures of selection were identified by inspecting all canary sequencing datasets.

Other signatures of selection appeared in regions that included genes involved in the migration of the early melanocyte precursors: KIT ligand (*KITLG*; NW_022044632.1:31.2–31.3 Mb) that appeared in the F_{ST} analysis of Black Intense Red vs. White Recessive and Urucum varieties; melanocyte inducing transcription factor (*MITF*; NW_022041986.1:4–48-4.59 Mb) identified in the comparison between Black Frosted Yellow and Gibber Italicus; endothelin receptor B (*EDNRB*; NW_022043763.1:49.35–49.38 Mb) that emerged in several comparisons that mainly distinguished two groups of canaries, one based on Black Frosted Yellow, Opal, Onyx and Opal × Onyx and the other that included all remaining varieties, as also evidenced by the distribution of allele frequencies at the polymorphic sites in this region (Figure 5).

Identification of the candidate gene mutations determining some dilute factors in canaries

Several signatures of selection emerged when the dilute varieties (Opal, Onyx, Opal × Onyx crosses and Mogno colours) were compared with the other canary varieties (Figures 3, 4). As mentioned above, most of them also appeared when the datasets retrieved from ENA and derived from other colour varieties were compared with themselves and with the newly sequenced canary populations. One genome region, however, emerged only when dilute varieties were compared with other non-dilute varieties or when newly sequenced dilute populations were contrasted with each other (Figures 2–4). This region contained the melanophilin (*MLPH*) gene (NW_022042652.1:66.48–66.51 Mb). Mutations in this gene, reported in many mammals and in few avian species, are well known to determine abnormal pigment distribution, resulting in diluted phenotypes (Bed'hom et al., 2012; Fontanesi et al., 2014; Ishida et al., 2006; Li et al., 2016; Matesic et al., 2001; Philipp et al., 2005; Vaez et al., 2008; Van Gele et al., 2009). Therefore, *MLPH* is a strong candidate gene whose variability might explain the diluted pigmentation in the investigated canary variants. Interestingly, a signal in the *MLPH* gene region emerged not only when Black Frosted Yellow canaries were compared with all mutated varieties (Opal, Onyx and Mogno), but also when Opal was compared with Onyx and Opal × Onyx and when Onyx was compared with Mogno (Figures 3 and 4), suggesting that more than one *MLPH* allele might be present in these different canary varieties.

By investigating the *MLPH* gene sequence from the generated whole genome sequencing datasets, we identified two putative causative mutations of the dilute canary variants. A missense mutation (NW_022042652.1:g.66497548C>T; Figure 6), causing the p.R111K substitution at the protein level (XM_009088195.3; XP_009086443.1), was fixed in the Onyx canaries and had an estimated frequency of 0.58 in the Opal × Onyx DNA pool (Table S3), close to the expected frequency of 50% based on fact that the crossbred canaries might have received a copy of the mutated allele from the Onyx parent and a wild-type form at this position from the other parental variety. PANTHER-PSEP analysis indicated that this missense mutation is 'probably damaging' (preservation time 456 MY; Pdel 0.57). The second mutation was a deletion of seven nucleotides (NW_022042652.1:g.66493407_66493413del; Figure 6), causing a frameshift of the reading frame and then the constitution of a premature stop codon that truncates the 676 residue wild-type protein at residue 236 (p.A237fs; XP_009086443.1). This deletion was fixed in both Opal and Mogno canaries. Again, in the Opal × Onyx DNA pool, its frequency was close to the expected 1:1 ratio of the two parental alleles (Table S3). This is again consistent with the fact that Opal and Onix carried two different *MLPH* alleles. As Opal and Mogno canaries carried the same allele at the *MLPH* gene, their phenotypic differences might be due to other modifier loci. Among the signatures of selections that emerged in the different comparisons and considering the allele frequency patterns at several pigmentation-related genes that were

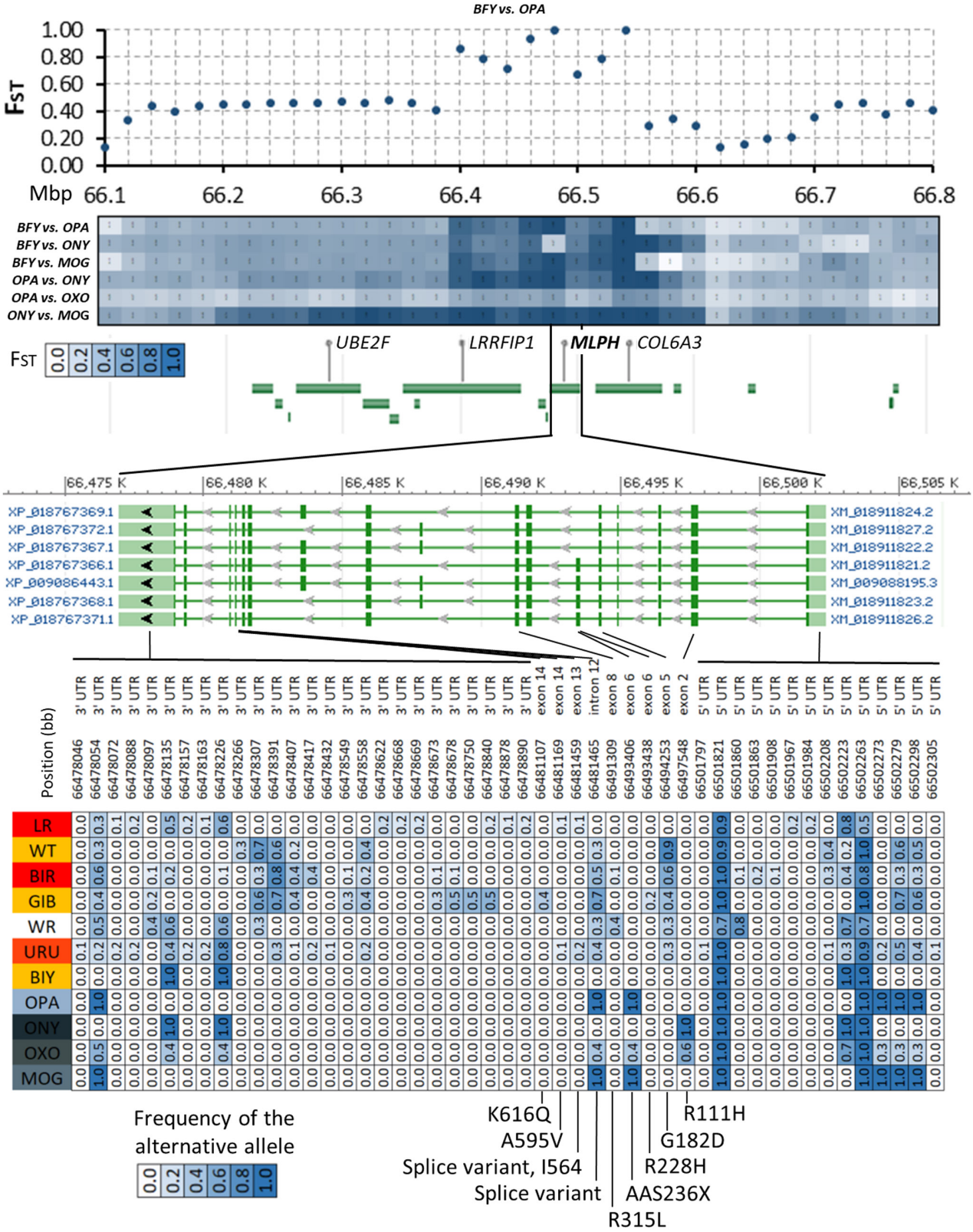

FIGURE 6 Detailed analysis of the *MLPH* gene region including the identified variants and their frequency in the analysed canary breeds/populations. Deduced transcripts and variant positions are referred to the *MLPH* gene annotated in the NW_022042652.1 scaffold. Acronyms of the different canary breeds/populations are explained in the legend to Figure 2 and in Table S1.

identified in our study, Opal and Mogno canaries were different at the *CYP2J19*, *EDC*, *BCO2*, *ASIP*, *MITF* and *EDNRB* gene regions (Figures 3–5, S4).

It is also worth noting that a few other signatures of selection differentiated the three dilute canary varieties, i.e. Opal, Onyx and Mogno (Table S2), suggesting that a few other modifier genes could be involved in defining subtle feather colour shades that might distinguish these canary lines. For example, another signature of selection that emerged in Opal × Onyx vs. Mogno (NW_022041542.1:27.76–27.84 Mb) and several other comparisons (Figure 4) contained the genes (glycosyltransferase-like domain-containing 1, *GTDC1*; and zinc finger E-box binding homeobox 2, *ZEB2*) that are annotated in the rat *Downunder* (*Du*) coat colour locus region, a pigmentation mutation that also causes eye abnormalities and embryonic lethality in homozygous condition (Hieu et al., 2022). In mice, melanocyte-specific *Zeb2* deletion resulted in a congenital loss of hair pigmentation, which suggests an essential role for this gene in melanogenesis (Denecker et al., 2014).

Other major signatures of selection

Several other signatures of selection, which did not contain any annotated known pigmentation-related genes, emerged in some comparisons (Figures 2, 3; Table S2 and Figure S2). These regions could other gene variants that may contribute to explaining the phenotypic differences of the analysed canary breeds and lines, which can also be distinguished for a few other traits, including feather characteristics and behavioural traits.

For example, a region on NW_022042208.1 emerged in 24 out of 55 pairwise comparisons, and in several cases with F_{ST} equal or close to 1.0. This region contained the OCRL inositol polyphosphate-5-phosphatase gene (*OCRL*; NW_022042208.1:19.09–19.12 Mb). In chicken, LOC431648, a pseudogene with high sequence similarity to the mRNA sequence of *OCRL*, has been shown to affect tail formation (Wang et al., 2017).

Another recurrent signature of selection (four out of 10 comparisons with WT and in three others pairwise contrasts; Figure 1, Table S2) was in the correspondence of the neuregulin 1 gene (*NRG1*; NW_022044521.1:33.24–33.35 Mb), which encodes for an estrogen-responsive gene closely correlated with the estrogen-dependent development of the oviduct of chicks and regeneration of the oviduct after moulting (Jeong et al., 2017). The encoded membrane glycoprotein mediates cell–cell signalling and plays critical roles in the growth and development of multiple organ systems and in neurological processes. Disregulation of this gene causes behaviour-related diseases (Babovic et al., 2008).

Ectodysplasin A gene (*EDA*), which is expressed in developing feather placodes in chicken embryos and involved in the feather patterning pathways, was in a signature of selection region (NW_022042208.1:5.25–5.32 Mb) identified by comparing Black Frosted Yellow and Gibber Italicus. *EDA* is considered one of the few genes that distinguish the genome of hybridising warblers (Toews et al., 2016).

DISCUSSION

Since the beginning of the canariculture, selective breeding of *S. canaria* raised for hobby purposes has made it possible to identify and then multiply many colour and feather varieties that, together with other traits, have been fixed in peculiar lines and then used extensively to produce novel varieties through their combinations. These canary lines constitute unique genetic resources that can be exploited to further expand the knowledge of the basic biological processes underlying pigmentation, cell and tissue development and other complex phenotypes that are difficult to dissect in other animal models.

Morphological phenotyping provided by amateur breeders can usually be very helpful to classify the main pigmentation related mutants and their allele series even if, in some cases, the incomplete knowledge on the dominance, pleiotropic and epistatic effects of several loci, the results of repeated introgressions, and also the involvement of modifier loci, could complicate the identification of the genetic factors involved in determining peculiar colour lines or breed varieties. Whole genome sequencing coupled with specific comparative analyses of sequence data obtained from a few canary varieties or crossbred lines has been already very successful in identifying a few genes involved in the production of carotenoid-based colourations (Gazda, Araújo, et al., 2020a; Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017, 2022). Reanalysing these datasets and adding other whole genome sequencing datasets produced from five additional canary varieties, we confirmed the signatures of selection already reported in major genes (*CYP2J19*, *EDC*, *BCO2* and *SCARB1*) and some of the relevant mutations identified by previous studies (Gazda, Araújo, et al., 2020a; Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017, 2022). In the case of *SCARB1*, it seems that the splice donor site mutation that determines the White Recessive phenotype has been spread in several other breeds and lines, providing a marker that can be used to track back the genetic history of some canary populations and the potential admixture events between populations. Similar fingerprints might be evidenced from the multiple signals that appeared in many different varieties in the correspondence of the *CYP2J19*, *EDC* and *BCO2* genes.

It is also worth mentioning here that, according to the fact that several gene regions already described by previous studies (Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017) also emerged from the newly sequenced DNA pools with similar peak heights

and background genetic variability to those derived from the original DNA pools retrieved from ENA from which this information was first identified (Gazda, Toomey, et al., 2020b; Lopes et al., 2016; Toomey et al., 2017), it seems evident that the F_{ST} approach was not biased by the different numbers of canaries of the investigated varieties included in the contrasted DNA pools. Therefore, the sub-optimal sample size that we used to construct the new DNA pools (four to 10 canaries/pool) was sufficient to obtain a good estimation of allele frequencies and then to capture differences between the compared groups of canaries, even if we cannot completely exclude some false positive and false negative results owing to a sampling effect and or to the Wahlund effect. Adequate sample size and sequencing depth are critical for pursuing efficiency in DNA pool-seq for population genomic studies: in these contexts, allele frequency estimates are more robust if the number of individual samples included in the DNA pool is larger and if the depth of sequencing is higher (Schlötterer et al., 2014). Technically speaking, therefore, we counterbalanced, at least in part, the biases of allele frequency estimates derived by the low number of individuals included in the DNA pools with a higher depth of sequencing that we obtained for the newly sequenced DNA pools than for those retrieved from ENA.

In line with these considerations and the discussed evidence and according to the observed general F_{ST} baseline values in all pairwise analyses, it seems that the genome of many canary populations became very homogenised, probably owing to extensive crossbreeding between breeds or varieties and therefore ongoing gene flow. A few exceptions are represented by some clear signals of strong genetic divergence in small genomic regions containing genes involved in the phenotypic peculiarities that are directly selected by fancy breeders. As a proof of concept, for example, some of these signals were those already reported from across-species introgression, and that constituted the 'red factor' canaries, which derived from the hybridisation between red siskins and common canaries (Lopes et al., 2016).

This general scenario (i.e. highly homogenised genome except in relevant selected regions) is also a great advantage when other signatures of selection are searched by comparing data obtained in DNA pools constituted by a few genomes. Exploiting this important aspect, we were also able to substantially enlarge the number of pigmentation-relevant genes identified thus far in *S. canaria*. The newly identified genes are mainly involved in the melanogenesis processes, complementing previous studies that focused on carotenoid-derived pigmentation. However, we cannot exclude the possibility that some of these additional detected regions could be false positives and that other regions exhibiting more subtle signatures have been missed. Some genes (*ASIP*, *TYRP1*, *AGRP*, *EDNRB* and *SLC45A2*) emerged in pairwise comparisons involving different canary varieties, suggesting that, for a few of them, more than one mutated allele occurred independently in different populations. For example, *ASIP* and *TYRP1* emerged in several lines in the comparisons against WT or Black Frosted Yellow canaries that represent wild or wild-like populations as well as in several other pairwise analyses (Figure 4). Additional genes (*DCT*, *KITLG* and *MITF*), that in other avian and mammalian species have been already shown to affect pigmentation (Liu et al., 2022; Ren et al., 2021; Stryjewski & Sorenson, 2017; Sultana et al., 2018), emerged in just one or a few comparisons involving canary DNA pools.

In a variety of avian species, *ASIP* gene mutations have been already recognised to affect pigmentation in several ways, for example in dorsoventral pigmentation patterning and recessive black plumage (like in mammals), in sexual dichromatism and other pigmentation effects (Campagna et al., 2022; Hiragaki et al., 2008;Nadeau et al., 2008; Oribe et al., 2012; Robic et al., 2019). The multiple signatures that we observed in the canary *ASIP* gene region could be due to more than one regulatory mutation, as the frequency deduced from DNA pools of the identified missense mutation (p.T63S) was not consistent with the signals that emerged in 14.5% pairwise comparisons (eight out of 55). A broad spectrum of regulatory regions affecting this gene have already been reported in chicken and Japanese quail (Nadeau et al., 2008; Oribe et al., 2012; Robic et al., 2019) and in several other mammalian species (Fontanesi et al., 2010; Trigo et al., 2021; Vrieling et al., 1994). In our cases, long read sequencing would be much more effective in solving the putative haplotype structures of this region that might exist in different canaries and capture candidate mutations instead of the short-read paired-end sequencing approach on DNA pools that we used. Gene expression studies involving different developmental stages, body regions and tissues might be also needed to support a functional role of this gene in affecting pigmentation variability in different canary lines.

Mutations in the *TYRP*1 gene, affecting plumage colouration, have been already reported in some avian species (Domyan et al., 2014; Li et al., 2019; Nadeau et al., 2007). We also identified a putative functional mutation in the canary *TYRP1* gene (that was also present in the reference genome sequence). This *TYRP1* frameshift mutation may disrupt the encoded enzyme, which might impair the regular production of black eumelanin. This mutation was observed in three canary varieties (Gibber Italicus, White Recessive and Urucum; Table S3). It would be interesting to evaluate the distribution of this frameshift mutation in several other canary colour varieties, in particular in Brown or Cinnamon lines, where it could be involved in determining the type of pigmentation that gives the names to these canary lines (Perez-Beato, 2008).

Frequent signatures of selection also highlighted *AGRP* as another candidate gene that may contribute to

genetically differentiate some of the compared canary lines. *AGRP* could potentially be involved in affecting pigmentation (Rodríguez-Martínez & Galván, 2020) even if its direct biological role in melanogenesis is still not completely clarified *in vivo* (Gluckman & Mundy, 2017). *AGRP* might be also relevant for other physiological mechanisms, according to its well-established role as an endogenous melanocortin receptor antagonist, with high affinity for melanocortin 4 receptor (MC4R) and, in turn, as a regulator of body weight, energy expenditure, feeding and behaviour (Deem et al., 2022; Nijenhuis et al., 2001).

EDNRB emerged as another interesting gene with a role in pigmentation patterns. Other studies in birds have already reported the involvement of an avian-specific paralog of endothelin receptor B (*EDNRB2*) in determining plumage colour variants in Japanese quail, chicken, geese and domestic ducks (Kinoshita et al., 2014; Li et al., 2015; Miwa et al., 2007; Xi et al., 2020). Clear signatures of selection in the *EDNRB* gene region, also with distinct allele frequency patterns (Figure 5), defined two groups of canaries: one group included four out of five of the newly sequenced lines (with the only exception of Mogno) and the other group included Mogno and all other varieties whose datasets were retrieved from previous studies. Allele frequency differences between the two groups were, however, present over a larger chromosome region than that covered by the *EDNRB* gene, suggesting the presence of a high level of linkage disequilibrium all over a few megabase pairs (Figure 5). Therefore, just using sequence data information, in this case, it would be difficult to disentangle the potential involvement of this or other closely annotated genes in contributing to any differences between these two groups of canaries, which, apparently, seem phenotypically quite heterogeneous.

Mutations in the *SLC45A2* gene explain the *S* (*Silver*) locus allele series in chickens, the sex-linked imperfect albinism locus in Japanese quails (Gunnarsson et al., 2007) and the sex-linked recessive dilute (*d*) allele in domestic rock pigeons (Domyan et al., 2014). Diluted effects on the pigmentation of Japanese quails have been also reported by structural mutations in the *ASIP* gene (Robic et al., 2019). Here, the F_{ST} pairwise analyses did not show any consistent signals between the genomic regions carrying these two genes (*SLC45A2* and *ASIP*) and the diluted phenotypes represented by the Opal, Onyx and Mogno varieties. However, the diluted phenotypes clearly matched with the signals that highlighted *MLPH* as the strongest candidate gene explaining a multiple allele series present in the investigated dilute varieties and attributed to the *Opal* locus (Perez-Beato, 2008). Mutations in this gene have been already associated with diluted pigmentations in many other species (Bed'hom et al., 2012; Fontanesi et al., 2014; Ishida et al., 2006; Li et al., 2016; Matesic et al., 2001; Philipp et al., 2005; Vaez et al., 2008; Van Gele et al., 2009). Two putative causative mutations were identified by analysing the coding region of this gene: one frameshift mutation (NW_022042652.1:g.66493407_66493413del), corresponding to the recessive allele *l* at the *Opal* locus, and a missense mutation (p.R111K), corresponding to the second recessive allele *l′* at the same locus. Sequence data clearly indicated that these two alleles emerged independently from wild-type forms, as allele *l′* does not contain the mutation of allele *l*, and vice versa. Even if we cannot formally exclude that other regulatory mutations might be involved in determining the diluted phenotypic effects of the Opal and Onyx mutants, these two exonic mutations, fixed respectively in the two canary lines, can be considered the strongest candidates to act as causative mutations. Anyway, they are markers that can be used to track these two recessive allele forms. The frameshift mutation was also fixed in the Mogno canaries, indicating that this line originated from an Opal mutant (at least considering the *MLPH* gene). Therefore, the diluted effect of Mogno canaries seems directly related to this disrupting mutation. Other regulatory *MLPH* mutations, whose presence in Mogno canaries cannot be excluded by our data analysis, would probably not change the disrupting effect of the NW_022042652.1:g.66493407_66493413del allele. Therefore, it is possible that Mogno phenotype could be due to a peculiar multilocus allele combination, as suggested by several signatures of selection that emerged by comparing Opal and Mogno DNA pools. Further studies are needed to better clarify this issue.

The quite large number of pigmentation-related genes that we highlighted in our study might also be due the heterogeneous composition of some of the newly sequenced DNA pools which included, in addition to the DNA of canaries having all the main target loci, also the DNA of some canaries that might carry other genotypes affecting secondary pigmentation factors. The whole genome sequencing comparative strategy that we applied in this study was also successful in showing other strong signatures of selection, probably not directly related to pigmentation differences between the contrasted canary lines. These additional results might be useful as reverse genetic starting points to reconstruct the biological meaning of fixed or almost fixed genomic regions in some canary lines and explain other unique phenotypic differences. Therefore, mining whole genome sequencing data can disclose unexpected hints that are useful to reconstruct the genetic history of some canary lines.

In conclusion, our study demonstrated that genomic information could be very useful to clarify the genetic mechanisms underlying phenotypic differences of many canary colour varieties and may represent complementary tools to inform fancy breeders in designing more effective plans in their breeding activities.

FUNDING INFORMATION

This study was supported by Federazione Ornicoltori Italiani and by the University of Bologna RFO2021 funds.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest that could have influenced the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data generated in this study are included in this published article and its supplementary information files. The newly produced whole genome sequencing data are available in the European Nucleotide Archive under the project number PRJEB58861.

ORCID

Samuele Bovo https://orcid.org/0000-0002-5712-8211
Anisa Ribani https://orcid.org/0000-0001-6778-1938
Valerio Joe Utzeri https://orcid.org/0000-0001-5320-4216
Valeria Taurisano https://orcid.org/0000-0001-5349-1606
Luca Fontanesi https://orcid.org/0000-0001-7050-3760

REFERENCES

Babovic, D., O'Tuathaigh, C.M., O'Connor, A.M., O'Sullivan, G.J., Tighe, O., Croke, D.T. et al. (2008) Phenotypic characterization of cognition and social behavior in mice with heterozygous versus homozygous deletion of catechol-O-methyltransferase. *Neuroscience*, 155, 1021–1029. Available from: https://doi.org/10.1016/j.neuroscience.2008.07.006

Baxter, L.L., Watkins-Chow, D.E., Pavan, W.J. & Loftus, S.K. (2019) A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell & Melanoma Research*, 32, 348–358. Available from: https://doi.org/10.1111/pcmr.12743

Bed'hom, B., Vaez, M., Coville, J.L., Gourichon, D., Chastel, O., Follett, S. et al. (2012) The lavender plumage colour in Japanese quail is associated with a complex mutation in the region of *MLPH* that is related to differences in growth, feed consumption and body temperature. *BMC Genomics*, 13, 442. Available from: https://doi.org/10.1186/1471-2164-13-442

Birkhead, T.R., Schulze-Hagen, K. & Kinzelbach, R. (2004) Domestication of the canary, *Serinus canaria* – the change from green to yellow. *Archives of Natural History*, 31, 50–56. Available from: https://doi.org/10.3366/anh.2004.31.1.50

Bovo, S., Ribani, A., Muñoz, M., Alves, E., Araujo, J.P., Bozzi, R. et al. (2020) Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. *Genetics Selection Evolution*, 52, 33. Available from: https://doi.org/10.1186/s12711-020-00553-7

Brush, A.H. (1990) Metabolism of carotenoid pigments in birds. *FASEB Journal*, 4, 2969–2977. Available from: https://doi.org/10.1096/fasebj.4.12.2394316

Campagna, L., Mo, Z., Siepel, A. & Uy, J.A.C. (2022) Selective sweeps on different pigmentation genes mediate convergent evolution of Island melanism in two incipient bird species. *PLoS Genetics*, 18, e1010474. Available from: https://doi.org/10.1371/journal.pgen.1010474

Deem, J.D., Faber, C.L. & Morton, G.J. (2022) AgRP neurons: regulators of feeding, energy expenditure, and behavior. *The FEBS Journal*, 289, 2362–2381. Available from: https://doi.org/10.1111/febs.16176

Denecker, G., Vandamme, N., Akay, O., Koludrovic, D., Taminau, J., Lemeire, K. et al. (2014) Identification of a ZEB2-MITF-ZEB1 transcriptional network that controls melanogenesis and melanoma progression. *Cell Death and Differentiation*, 21, 1250–1261. Available from: https://doi.org/10.1038/cdd.2014.44

Domyan, E.T., Guernsey, M.W., Kronenberg, Z., Krishnan, S., Boissy, R.E., Vickrey, A.I. et al. (2014) Epistatic and combinatorial effects of pigmentary gene mutations in the domestic pigeon. *Current Biology*, 24, 459–464. Available from: https://doi.org/10.1016/j.cub.2014.01.020c

Fontanesi, L., Forestier, L., Allain, D., Scotti, E., Beretti, F., Deretz-Picoulet, S. et al. (2010) Characterization of the rabbit agouti signaling protein (*ASIP*) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. *Genomics*, 95, 166–175. Available from: https://doi.org/10.1016/j.ygeno.2009.11.003

Fontanesi, L., Scotti, E., Allain, D. & Dall'Olio, S. (2014) A frameshift mutation in the melanophilin gene causes the dilute coat colour in rabbit (*Oryctolagus cuniculus*) breeds. *Animal Genetics*, 45, 248–255. Available from: https://doi.org/10.1111/age.12104

Gazda, M.A., Araújo, P.M., Lopes, R.J., Toomey, M.B., Andrade, P., Afonso, S. et al. (2020a) A genetic mechanism for sexual dichromatism in birds. *Science*, 368, 1270–1274. Available from: https://doi.org/10.1126/science.aba0803

Gazda, M.A., Toomey, M.B., Araújo, P.M., Lopes, R.J., Afonso, S., Myers, C.A. et al. (2020b) Genetic basis of de novo appearance of carotenoid ornamentation in bare parts of canaries. *Molecular Biology and Evolution*, 37, 1317–1328. Available from: https://doi.org/10.1093/molbev/msaa006

Gluckman, T.L. & Mundy, N.I. (2017) The differential expression of MC1R regulators in dorsal and ventral quail plumages during embryogenesis: implications for plumage pattern formation. *PLoS One*, 12, e0174714. Available from: https://doi.org/10.1371/journal.pone.0174714

Gunnarsson, U., Hellström, A.R., Tixier-Boichard, M., Minvielle, F., Bed'hom, B., Ito, S. et al. (2007) Mutations in SLC45A2 cause plumage color variation in chicken and Japanese quail. *Genetics*, 175, 867–877. Available from: https://doi.org/10.1534/genetics.106.063107

Hieu, H.T., Tanaka, M., Kuwamura, M., Mashimo, T., Serikawa, T. & Kuramoto, T. (2022) The rat *Downunder* (*Du*) coat color mutation is associated with eye anomalies and embryonic lethality and maps to a 3.9-mb region on chromosome 3. *Experimental Animals*, 72, 88–94. Available from: https://doi.org/10.1538/expanim.22-0086

Hiragaki, T., Inoue-Murayama, M., Miwa, M., Fujiwara, A., Mizutani, M., Minvielle, F. et al. (2008) Recessive black is allelic to the yellow plumage locus in Japanese quail and associated with a frameshift deletion in the *ASIP* gene. *Genetics*, 178, 771–775. Available from: https://doi.org/10.1534/genetics.107.077040

Ishida, Y., David, V.A., Eizirik, E., Schäffer, A.A., Neelam, B.A., Roelke, M.E. et al. (2006) A homozygous single-base deletion in *MLPH* causes the dilute coat color phenotype in the domestic cat. *Genomics*, 88, 698–705. Available from: https://doi.org/10.1016/j.ygeno.2006.06.006

Jeong, W., Bae, H., Lim, W., Bazer, F.W. & Song, G. (2017) Differential expression of neuregulin 1 (NRG1) and candidate miRNA regulating NRG1 transcription in the chicken oviduct in response to hormonal changes. *Journal of Animal Science*, 95, 3885–3904. Available from: https://doi.org/10.2527/jas2017.1663

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., Anderson, N. et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature*

Genetics, 39, 1321–1328. Available from: https://doi.org/10.1038/ng.2007.10

Kinoshita, K., Akiyama, T., Mizutani, M., Shinomiya, A., Ishikawa, A., Younis, H.H. et al. (2014) *Endothelin receptor B2* (*EDNRB2*) is responsible for the tyrosinase-independent recessive white (*mo*^{w}) and mottled (*mo*) plumage phenotypes in the chicken. *PLoS One*, 9, e86361. Available from: https://doi.org/10.1371/journal.pone.0086361

Kondo, T. & Hearing, V.J. (2011) Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Review of Dermatology*, 6, 97–108. Available from: https://doi.org/10.1586/edm.10.70

Lamorux, M.L., Delmas, V., Larue, L. & Bennett, D.C. (2010) *The colors of mice: a model genetic network*. Oxford, UK: Wiley-Blackwell.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25, 1754–1760. Available from: https://doi.org/10.1093/bioinformatics/btp324

Li, L., Li, D., Liu, L., Li, S., Feng, Y., Peng, X. et al. (2015) Endothelin receptor B2 (*EDNRB2*) gene is associated with spot plumage pattern in domestic ducks (*Anas platyrhynchos*). *PLoS One*, 10, e0125883. Available from: https://doi.org/10.1371/journal.pone.0125883

Li, W., Sartelet, A., Tamma, N., Coppieters, W., Georges, M. & Charlier, C. (2016) Reverse genetic screen for loss-of-function mutations uncovers a frameshifting deletion in the melanophilin gene accountable for a distinctive coat color in Belgian blue cattle. *Animal Genetics*, 47, 110–113. Available from: https://doi.org/10.1111/age.12383

Li, J., Bed'hom, B., Marthey, S., Valade, M., Dureux, A., Moroldo, M. et al. (2019) A missense mutation in *TYRP1* causes the chocolate plumage color in chicken and alters melanosome structure. *Pigment Cell & Melanoma Research*, 32, 381–390. Available from: https://doi.org/10.1111/pcmr.12753

Liu, H., Wang, J., Hu, J., Wang, L., Guo, Z., Fan, W. et al. (2022) Genome-wide association analysis reveal the genetic reasons affect melanin spot accumulation in beak skin of ducks. *BMC Genomics*, 23, 236. Available from: https://doi.org/10.1186/s12864-022-08444-5

Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J. et al. (2016) Genetic basis for red coloration in birds. *Current Biology*, 26, 1427–1434. Available from: https://doi.org/10.1016/j.cub.2016.03.076

Matesic, L.E., Yip, R., Reuss, A.E., Swing, D.A., O'Sullivan, T.N., Fletcher, C.F. et al. (2001) Mutations in *Mlph*, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. *Proceedings of the National Academy of Sciences of the USA*, 98, 10238–10243. Available from: https://doi.org/10.1073/pnas.181336698

McGraw, K.J. (2006) Mechanics of carotenoid-based coloration. In: Hill, G.E. & McGraw, K.J. (Eds.) *Bird coloration. I. Mechanisms and measurements*. Cambridge, MA, USA: Harvard University Press, pp. 177–242.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A. et al. (2016) The Ensembl variant effect predictor. *Genome Biology*, 17, 122. Available from: https://doi.org/10.1186/s13059-016-0974-4

Miwa, M., Inoue-Murayama, M., Aoki, H., Kunisada, T., Hiragaki, T., Mizutani, M. et al. (2007) Endothelin receptor B2 (*EDNRB2*) is associated with the panda plumage colour mutation in Japanese quail. *Animal Genetics*, 38, 103–108. Available from: https://doi.org/10.1111/j.1365-2052.2007.01568.x

Nadeau, N.J., Mundy, N.I., Gourichon, D. & Minvielle, F. (2007) Association of a single-nucleotide substitution in *TYRP1* with roux in Japanese quail (*Coturnix japonica*). *Animal Genetics*, 38, 609–613. Available from: https://doi.org/10.1111/j.1365-2052.2007.01667.x

Nadeau, N.J., Minvielle, F., Ito, S., Inoue-Murayama, M., Gourichon, D., Follett, S.A. et al. (2008) Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian *ASIP* (*agouti*) gene. *Genetics*, 178, 777–786. Available from: https://doi.org/10.1534/genetics.107.077073

Nijenhuis, W.A.J., Oosterom, J. & Adan, R.A.H. (2001) AgRP (83–132) acts as an inverse agonist on the human-melanocortin-4 receptor. *Molecular Endocrinology*, 15, 164–171. Available from: https://doi.org/10.1210/mend.15.1.0578

Oribe, E., Fukao, A., Yoshihara, C., Mendori, M., Rosal, K.G., Takahashi, S. et al. (2012) Conserved distal promoter of the agouti signaling protein (*ASIP*) gene controls sexual dichromatism in chickens. *General and Comparative Endocrinology*, 177, 231–237. Available from: https://doi.org/10.1016/j.ygcen.2012.04.016

Parsons, J.J. (1987) The origin and dispersal of the domesticated canary. *Journal of Cultural Geography*, 7, 19–33. Available from: https://doi.org/10.1080/08873638709478505

Perez-Beato, O. (2008) *Fundamentals of color genetics in canaries*. Pittsburgh, Pennsylvania, USA: RoseDog Books.

Philipp, U., Hamann, H., Mecklenburg, L., Nishino, S., Mignot, E., Günzel-Apel, A.R. et al. (2005) Polymorphisms within the canine *MLPH* gene are associated with dilute coat color in dogs. *BMC Genetics*, 6, 34. Available from: https://doi.org/10.1186/1471-2156-6-34

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A. et al. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178. Available from: https://doi.org/10.1101/201178

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. Available from: https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2018) *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org/

Ren, S., Lyu, G., Irwin, D.M., Liu, X., Feng, C., Luo, R. et al. (2021) Pooled sequencing analysis of geese (*Anser cygnoides*) reveals genomic variations associated with feather color. *Frontiers in Genetics*, 12, 650013. Available from: https://doi.org/10.3389/fgene.2021.650013

Robic, A., Morisson, M., Leroux, S., Gourichon, D., Vignal, A., Thebault, N. et al. (2019) Two new structural mutations in the 5′ region of the *ASIP* gene cause diluted feather color phenotypes in Japanese quail. *Genetics Selection Evolution*, 51, 12. Available from: https://doi.org/10.1186/s12711-019-0458-6

Rodríguez-Martínez, S. & Galván, I. (2020) A source of exogenous oxidative stress improves oxidative status and favors pheomelanin synthesis in zebra finches. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 228, 108667. Available from: https://doi.org/10.1016/j.cbpc.2019.108667

Schlötterer, C., Tobler, R., Kofler, R. & Nolte, V. (2014) Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Reviews. Genetics*, 15, 749–763. Available from: https://doi.org/10.1038/nrg3803

Stryjewski, K.F. & Sorenson, M.D. (2017) Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution*, 1, 1912–1922. Available from: https://doi.org/10.1038/s41559-017-0364-7

Sultana, H., Seo, D., Choi, N.R., Bhuiyan, M.S.A., Lee, S.H., Heo, K.N. et al. (2018) Identification of polymorphisms in *MITF* and *DCT* genes and their associations with plumage colors in Asian duck breeds. *Asian-Australasian Journal of Animal Sciences*, 31, 180–188. Available from: https://doi.org/10.5713/ajas.17.0298

Tang, H. & Thomas, P.D. (2016) PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, 32, 2230–2232. Available from: https://doi.org/10.1093/bioinformatics/btw222

Toews, D.P., Taylor, S.A., Vallender, R., Brelsford, A., Butcher, B.G., Messer, P.W. et al. (2016) Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*,

26, 2313–2318. Available from: https://doi.org/10.1016/j.cub.2016.06.034

Toomey, M.B., Lopes, R.J., Araújo, P.M., Johnson, J.D., Gazda, M.A., Afonso, S. et al. (2017) High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proceedings of the National Academy of Sciences of the USA*, 114, 5219–5224. Available from: https://doi.org/10.1073/pnas.1700751114

Toomey, M.B., Marques, C.I., Araújo, P.M., Huang, D., Zhong, S., Liu, Y. et al. (2022) A mechanism for red coloration in vertebrates. *Current Biology*, 32, 4201–4214.e12. Available from: https://doi.org/10.1016/j.cub.2022.08.013

Trigo, B.B., Utsunomiya, A.T.H., Fortunato, A.A.A.D., Milanesi, M., Torrecilha, R.B.P., Lamb, H. et al. (2021) Variants at the *ASIP* locus contribute to coat color darkening in Nellore cattle. *Genetics Selectin Evolution*, 53, 40. Available from: https://doi.org/10.1186/s12711-021-00633-2

Vaez, M., Follett, S.A., Bed'Hom, B., Gourichon, D., Tixier-Boichard, M. & Burke, T. (2008) A single point-mutation within the melanophilin gene causes the lavender plumage colour dilution phenotype in the chicken. *BMC Genetics*, 9, 7. Available from: https://doi.org/10.1186/1471-2156-9-7

Van Gele, M., Dynoodt, P. & Lambert, J. (2009) Griscelli syndrome: a model system to study vesicular trafficking. *Pigment Cell & Melanoma Research*, 22, 268–282. Available from: https://doi.org/10.1111/j.1755-148X.2009.00558.x

Vrieling, H., Duhl, D.M., Milla, S.E., Miller, K.A. & Barsh, G.S. (1994) Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proceedings of the National Academy of Sciences of the USA*, 91, 5667–5671. Available from: https://doi.org/10.1073/pnas.91.12.5667

Wang, Q., Pi, J., Pan, A., Shen, J. & Qu, L. (2017) A novel sex-linked mutant affecting tail formation in Hongshan chicken. *Scientific Reports*, 7, 10079. Available from: https://doi.org/10.1038/s41598-017-10943-5

Xi, Y., Wang, L., Liu, H., Ma, S., Li, Y., Li, L. et al. (2020) A 14-bp insertion in endothelin receptor B-like (*EDNRB2*) is associated with white plumage in Chinese geese. *BMC Genomics*, 21, 162. Available from: https://doi.org/10.1186/s12864-020-6562-8

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bovo, S., Ribani, A., Utzeri, V.J., Taurisano, V., Bertarini, G. & Fontanesi, L. (2023) Whole genome sequencing identifies candidate genes and mutations that can explain diluted and other colour varieties of domestic canaries (*Serinus canaria*). *Animal Genetics*, 00, 1–16. Available from: https://doi.org/10.1111/age.13331