bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Title page**
2
3 **Classification**: BIOLOGICAL SCIENCES: Evolution
4
5 Repeated divergent selection on pigmentation genes in a rapid finch radiation driven by sexual
6 selection
7
8 **Short title:** Repeated divergent selection in an avian radiation
9
10 Leonardo Campagna[a,b,1], Márcio Repenning[c], Luis Fabio Silveira[d], Carla Suertegaray Fontana[c],
11 Pablo L Tubaro[e], Irby J Lovette[a,b].
12
13 [a]Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, 159 Sapsucker
14 Woods Road, Ithaca, NY, USA, 14850.
15
16 [b]Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca,
17 NY, USA, 14853.
18
19 [c]Laboratório de Ornitologia, Museu de Ciências e Tecnologia, Pontifícia Universidade Católica
20 do Rio Grande do Sul (PUCRS), Av. Ipiranga 6681, 90619-900, Porto Alegre, RS. Brazil.
21
22 [e]División de Ornitología, Museo Argentino de Ciencias Naturales ''Bernardino Rivadavia''
23 (MACN - CONICET), Av. Ángel Gallardo 470, Ciudad de Buenos Aires, C1405DJR Buenos
24 Aires, Argentina.
25
26 [d]Seção de Aves, Museu de Zoologia, Universidade de São Paulo (MZUSP), Caixa Postal 42.494,
27 CEP 04218-970, São Paulo, SP, Brasil.
28
29 [1]To whom correspondence should be addressed. Leonardo Campagna. Cornell Laboratory of
30 Ornithology, 159 Sapsucker Woods Road, Ithaca, NY, USA, 14850. Phone:1-607-339-1892.
31 Email: leocampagna@gmail.com.
32
33 Key Words: Capuchino Seedeaters; Genomic Scan; Melanogenesis; Neotropics.
34
35

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36 **Abstract**
37 The search for molecular targets of selection is leading to a better understanding of how
38 evolution shapes biological diversity. Instances of recent and rapid speciation are suitable for
39 associating phenotypes with their causal genotypes, because gene flow may homogenize areas of
40 the genome that are not under divergent selection. Locating differentiated genomic regions
41 among taxa allows us to test associations between the genes in these regions and their
42 contributions to phenotypic diversity. Here we study a rapid radiation of nine sympatric bird
43 species known as southern capuchino seedeaters, which are strikingly differentiated in sexually
44 selected characters of male plumage and song. We sequenced the genomes of 72 individuals
45 representing a diverse set of species and associated phenotypes to seearch for differentiated
46 genomic regions. We asked what genes are harbored in divergent regions and to what extent has
47 selection on the same targets shaped phenotypic diversity across different lineages. Capuchinos
48 show differences in a small proportion of their genomes, yet selection has acted independently
49 on the same targets during the groups' radiation. Many divergence peaks contain genes involved
50 in the melanogenesis pathway, with the strongest signal originating from a regulatory region
51 upstream of the gene coding for the Agouti-signaling protein. Across all divergence peaks, the
52 most differentiated areas are similarly likely regulatory. Our findings are consistent with
53 selection acting on the same genomic regions in different lineages to shape the evolution of cis-
54 regulatory elements, which control how more conserved genes are expressed and thereby
55 generate diversity in sexually selected traits.
56

57 **Introduction**

58

59 Understanding the processes that shape biological diversity at the molecular level is a central
60 goal of evolutionary biology. Studies of non-model organisms with divergent traits can be
61 powerful systems in which to discover the genetic basis of distinct phenotypes. In some cases the
62 same genes generate similar phenotypes independently in different taxa (e.g., wing color patterns
63 in butterflies (1)). Although some phenotypic differences are caused by mutations in coding
64 regions of causal genes (2-4), others arise through selection on areas that regulate the expression
65 of these genes (e.g., morphological and coloration differences in African cichlids (4)).
66 Additionally, macro-mutations such as chromosomal inversions can suppress local
67 recombination leading to the formation of supergenes, which allow genes to co-evolve and
68 produce complex traits (e.g., plumage coloration and mating behavior in Ruffs and White-
69 throated Sparrows (5-7)). Despite this wealth of knowledge, connecting the evolution of
70 phenotype to the genetic mechanisms that generate reproductive isolation and ultimately
71 speciation remains challenging in most systems. One of the best understood cases of the genetics
72 of speciation in animals comes from Darwin's finches, where the morphological traits that are
73 under selection have been identified (8), and both the molecular mechanisms that generate those
74 traits (9-11), and the effect of trait variation on reproductive isolation are known (12, 13).
75 From a genomic perspective, Darwin's finches have offered two key advantages for
76 researchers searching for the molecular basis of the phenotypes that distinguish these birds. First,
77 the finches have speciated recently, which translates into a relatively low background genomic
78 differentiation. Those few areas of the genome that are highly divergent among species contain
79 candidate loci that may have shaped the evolution of the adaptive radiation, or are still in action,
80 even in the face of gene flow (10, 11). Secondly, there are many different species in the radiation
81 with comparable divergence times, which leads to similarly low background genomic

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

82 differentiation across multiple possible comparisons. This allows researchers to compare the
83 genomes of more than one pair of forms with similar differences in phenotype, and assess the
84 degree to which molecular evolution has happened in parallel (11).
85 The study of additional biological systems that share these tractable attributes, but which
86 have been driven by forces other than natural selection on foraging-related phenotypes, can
87 provide further insights into the genomics of traits that may lead to speciation. Here we focus on
88 a group of finch-like birds from continental South America known as the southern capuchino
89 seedeater radiation (14, 15). Capuchinos share many characteristics with Darwin's finches (16),
90 yet differ in that they seem to have diversified primarily via sexual selection on plumage traits
91 that are likely melanin-based, rather than via natural selection on foraging-related traits (14, 15).
92 Capuchino seedeaters belong to the genus *Sporophila* and are in the same family as Darwin's
93 finches (Thraupidae) (17), and both radiations show comparable speciation rates that are much
94 greater than all other groups within that large family (17). The southern capuchinos are nine
95 predominantly sympatric species that occur in Neotropical grasslands (Fig. 1A): *S. bouvreuil*
96 (bou), *S. pileata* (pil), *S. cinnamomea* (cin), S. *ruficollis* (ruf), *S. melanogaster* (mel), *S.*
97 *nigrorufa* (nig), *S. palustris* (pal), *S. hypochroma* (hypoch), and *S. hypoxantha* (hypox).
98 Capuchinos are sexually dimorphic, and males from different species differ in secondary sexual
99 characters such as the plumage coloration patterns they use to attract mates and the songs they
100 use to defend their territories (14). Males defend their territories during simulated intrusions of
101 conspecifics, but not from sympatric male capuchinos of other species (18). The species in the
102 group are otherwise indistinguishable morphologically (19, 20) (and very similar ecologically
103 (21, 22)), to the extent that females and juveniles lacking male secondary sexual characters
104 cannot be identified to species even in the hand (14, 15). Despite their phenotypic diversity in
105 male plumage, southern capuchinos show extremely low levels of genetic differentiation (14),
106 and except for *S. bouvreuil* cannot be assigned reliably to species even using thousands of
107 genome-wide SNP markers (15). This genetic homogeneity is a result of the groups' recent
108 origin during the Pleistocene, and likely the product of both incomplete lineage sorting and
109 ongoing genetic admixture (14, 15). The apparent genetic homogeneity among southern
110 capuchinos, despite their distinct phenotypes that are maintained in sympatry, led us to
111 hypothesize that these species differences in male plumage may be the result of strong selection
112 at a few key loci. We therefore sequenced the genomes of the nine southern capuchinos with the
113 objective of locating such loci and to test whether the same targets of selection have shaped
114 phenotypic diversity independently across different species.
115
116 **Results**
117
118 Individual capuchinos clustered by species in a PCA derived from 11.5 million SNPs, yet the
119 percentage of variation explained by the first two principal components was low, suggesting that
120 a small proportion of these SNPs could be driving the pattern (Fig. 1B, Fig. S1). To search for
121 divergent areas of the genome we compared F_{ST} values for non-overlapping 25 kb windows
122 across the ten possible pairwise comparisons of five species (those for which our sample sizes
123 were larger). The mean differentiation across all windows/comparisons was low (mean F_{ST} =
124 0.008 and SD = 0.015 across ~43.5 thousand windows), yet we found a number of divergence
125 peaks with highly elevated F_{ST} with respect to this low background. For example, Fig. 1C shows
126 two Manhattan plots; the upper graph (nig vs. mel) had the largest number of elevated windows
127 (0.3%) among all comparisons and involved two allopatric species with small ranges. The

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

128 bottom comparison (hypox vs. pal) had an order of magnitude fewer (0.03%) and compared two
129 species with highly overlapping ranges. All other pairwise comparisons had a number of
130 divergence peaks that ranged between the extremes shown in Fig. 1C (Fig. S2). We identified a
131 total of 25 divergence peaks with elevated F_{ST} (>0.2) that are candidate targets of selection
132 driving species differences among capuchinos (Table 1); these peaks range in width from 25 to
133 840 kb (average ~243 kb, inset in Fig. 1C). SNPs from these divergence peaks alone can be used
134 to assign individuals to species in a PCA (inset in Fig. 1B). We found 99 SNPs that were fixed
135 (F_{ST}=1) in at least one comparison across all pairwise combinations of five capuchinos; these
136 represented 65 different sites in the genome. Because our sample sizes were low for the
137 remaining capuchino species (bou, cin, hypoch, ruf), we did not use them to identify divergence
138 peaks. A PCA using the SNPs from under the peaks showed some overlap between taxa when
139 we included all nine species (Fig. S1, mainly between ruf/hypox and bou/pil). It is therefore
140 possible that there are some additional undetected areas of the genome that are involved in
141 capuchino seedeater differentiation.
142 Next we asked if the same divergence peaks were involved in differentiation across
143 multiple combinations of capuchino species in ways that imply independent patterns of selection
144 on the same loci. The left panels in Fig. 2 show examples of divergence peaks (5 kb windows)
145 with the ten different pairwise comparisons overlaid. Although no single area of the genome (i.e.,
146 differentiation peak) was present in all comparisons, many are found in multiple comparisons.
147 For example, the peak in Fig. 2A was present in nine out of the ten comparisons, many of which
148 involved independent pairs of species (e.g., nig vs. mel and pil vs. hypox). Other divergence
149 peaks were less ubiquitous, yet are present across multiple pairs of species (left panels in Fig. 2,
150 Fig. S3, and Fig. S4). To better understand the nature of the differences among species within the
151 divergence peaks, we conducted PCAs with the SNPs from each of these areas separately. Fig.
152 1B shows four clusters of haplotypes in the region under the most common peak in our dataset.
153 Other divergence peaks varied in the species they were present and the extent to which they
154 could be used to diagnose species in PCAs (center panels in Fig. 2, Fig. S3, and Fig. S4).
155 We identified a total of 246 gene models within these divergent areas of the genome (an
156 average of 10 per divergence peak), 156 of which matched genes of known functions in other
157 species. We performed an enrichment analyses to understand if genes in this list were
158 predominantly involved in certain pathways. The most prominent hit was the melanogenesis
159 pathway (KEGG pathway analysis, p= 2.0E-3). We found nine melanogenesis genes in eight
160 different divergence peaks (Table 1). The peak containing the gene coding for the Agouti-
161 signaling protein (ASIP) had the largest number of highly divergent SNPs in our dataset: 30% of
162 all observed SNPs with F_{ST}>0.85 and 58% of all SNPs with F_{ST}=1 (Fig. 2C). Accordingly, this
163 was the peak that showed the greatest increase in absolute sequence divergence (measured using
164 the Dxy statistic) when comparing the region under the peak to the areas on the same scaffold
165 outside of the peak (Fig. S5). Other peaks contained a smaller number of these highly divergent
166 SNPs (Fig. 2, Table 1, Fig. S3, Fig. S4). The peaks containing melanogenesis genes accumulated
167 60% of SNPs with F_{ST}>0.85 and 63% of fixed SNPs (across all pairwise comparisons). Fig. 2
168 and Fig. S3 show the eight divergence peaks containing melanogenesis genes, ranked from top to
169 bottom by the number of highly divergent SNPs (F_{ST}>0.85) present in these peaks (summarized
170 in Table 1). The right panels in these figures indicate the position of these SNPs with respect to
171 the closest gene models. We also identified three peaks that together accumulated 30% of the
172 observed highly divergent SNPs and did not contain melanogenesis genes (Fig. S4), however one
173 of these peaks contained the gene HERC2. An intron within this gene functions as an enhancer

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

174 which regulates the expression of the OCA2 pigmentation gene in humans, involved in
175 controlling eye, hair and skin color (23). The remaining peaks in Fig. S4 did not contain genes
176 that could be easily associated to plumage or other phenotypes (Fig. S4). The remaining 14 peaks
177 (Table 1) accounted for only 10% of the observed highly divergent SNPs. Finally, because the
178 Melanocortin 1 Receptor (MC1R) is known to affect plumage coloration in many bird species
179 and interact directly with ASIP (24), we asked if this gene showed divergence among capuchino
180 seedeaters and had been overlooked in our analysis. MC1R was present in our reference genome
181 assembly, yet did not show differences in any of the pairwise comparison across our five
182 capuchino species (Fig. S6).
183 Nearly all the fixed sites we observed (99%) were located in non-coding areas of the
184 genome. The most common peak in our dataset concentrated 58% of these fixed differences
185 within several thousand kb up and downstream of the ASIP gene (Fig. 2C). We found that these
186 areas contained positions that were highly conserved across the genomes of distantly related
187 birds (Turkey, Chicken, Budgerigar and Zebra finch), comparable in their levels of conservation
188 to certain positions on the exons of ASIP (Fig. S7). It is therefore likely that these regions
189 contain cis-regulatory elements that control the expression of ASIP, and a similar situation may
190 be true for the other differentiated regions found in close proximity to genes.
191
192 **Discussion**
193
194 Despite their striking differences in male plumage, southern capuchino seedeaters are
195 differentiated only in a small proportion of their genomes. The identity of these rare
196 differentiated genomic regions differs somewhat among capuchinos, yet in many cases the same
197 divergent regions are present in comparisons across many pairs of species. This convergence of
198 differentiation across multiple independent pairs of species implies that selection has acted
199 repeatedly in different lineages on the same genomic targets to shape phenotypes.
200 Many of the most highly differentiated areas of the capuchino genome contain genes that
201 are part of the melanogenesis pathway. The area upstream of the gene coding for the Agouti-
202 signaling protein is the most ubiquitous peak, showing the strongest signal of differentiation.
203 More generally, we observe narrow divergence peaks in different components of the
204 melanogenesis pathway that are genetically unlinked. Plumage coloration is generally important
205 for reproductive isolation in birds (25), and differences in genes that control melanin-based
206 variation in plumage have been found in different pairs of incipient avian taxa (e.g., carrion and
207 hooded crows (26), flycatchers of the Solomon Islands (27), blue-winged and golden-winged
208 warblers (28)). The differences we observe in capuchinos are mostly in non-coding areas of the
209 genome, therefore our findings are consistent with the evolution of cis-regulatory elements.
210 These regulatory elements may vary by species, yet control the expression of the same set of
211 genes, generating the strikingly different phenotypes we observe in the capuchinos. In particular,
212 the regulation of the expression of melanogenesis genes may lead to pigmentation differences
213 across the plumage patches (e.g., throat vs. back) that appear to vary somewhat modularly across
214 capuchino species (Fig. 1A).
215 Three factors could have contributed to the rapid evolution of phenotypic diversity in the
216 capuchinos. First, a very large effective population size was inferred for the ancestor of the
217 radiation (15). The amount of genetic variation a population can sustain is proportional to its
218 size, with large populations providing more possible substrates for rapid evolution from standing
219 genetic variation to take place (29). Additionally, differences could have accumulated among

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

220 species in allopatry and eventually been exchanged via hybridization, leading to novel
221 phenotypes as has been described for *Heliconius* butterflies (30). In particular, the differentiation
222 and exchange via hybridization of regulatory elements that control the expression of more
223 conserved genes has been found to drive phenotypic diversity in *Heliconius* (31). A similar
224 situation could have contributed to phenotypic diversity in the capuchinos, which also show
225 modular variation in their plumage, with the same patches having different colors depending on
226 the species (e.g., throat, back or belly can be black, white, cinnamon or rufous in different
227 species; Fig. 1A). Finally, 10 out of the 25 divergence peaks that we identified are located on the
228 Z chromosome. Sex chromosomes are known to evolve faster than autosomes because of their
229 smaller effective population size (32). Moreover, sex chromosomes may also be particularly
230 relevant to speciation (33), as suggested by Haldane's rule, where the heterogametic sex is
231 commonly inviable or missing from hybrid crosses (34). For both reasons divergence in genes
232 located on sex chromosomes could have facilitated rapid evolution in capuchinos.
233 We have identified genomic substrates that lead to phenotypic diversity in capuchinos,
234 differences that are likely relevant to mate recognition and eventually reproductive isolation in
235 these strongly sexually dimorphic species. Because the most divergent areas across the genomes
236 of capuchinos contain pigmentation genes, this leads to the question of whether we have found
237 the genes responsible for maintaining these lineages as separate species. The pigmentation genes
238 are linked to other loci for which the connection between genotype and phenotype is harder to
239 make, and we do not understand the contribution of these additional genes to species differences.
240 It is possible that prezygotic isolation via mate choice is strong enough to maintain the
241 phenotypic integrity of capuchino species even though many species breed in local-scale
242 sympatry. However, we also cannot at this point discard the possibility that postzygotic
243 incompatibilities exist between species, perhaps associated with these same divergent regions of
244 the genome. As natural selection has shaped the beaks of finches in the Galapagos, leading to the
245 generation of biological diversity, our study suggests sexual selection may have shaped the
246 plumage and songs of male capuchinos, generating yet another extraordinary rapid radiation of
247 finch-like birds.
248
249 **Materials and Methods**
250
251 **Reference genome assembly and annotation**. We combined short-read Illumina data
252 (estimated depth of coverage of 112x) with long-read Pacific Biosciences sequencing (estimated
253 depth of coverage of 9.1x) to assemble the genome of a *S. hypoxantha* individual using
254 ALLPATHS-LG (35) and PBJelly (36). The total length of the assembly was 1.17 Gb distributed
255 in 5120 scaffolds, with an N50 of 8.7 Mb and 4.8% Ns. Our reference genome contained a single
256 copy of 90% of a set of conserved vertebrate genes used to assess assembly quality (37). We
257 annotated the *S. hypoxantha* genome with MAKER (38), which produced a total of 14,667 gene
258 models (75.5% of the 19,437 genes annotated in the Zebra Finch genome). We searched for
259 annotations of interest in the UniProt database (http://www.uniprot.org/) and identified enriched
260 pathways using DAVID v6.7 (39). See SI Materials and Methods for details.
261
262 **Population level sequencing and variant discovery**. We re-sequenced the genomes of 72
263 individuals from nine capuchino species. We included 12 individuals per species for nig, pil,
264 mel, hypox, and pal, and 3 individuals per species for bou, cin, hypoch, and ruf. After quality
265 filtering we aligned individually barcoded samples to the reference genome using Bowtie2 (40).

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

266 Because *S. hypoxantha* shows similar genetic distances to all other members of the clade (14),
267 we did not observe a species effect on mapping efficiency or quality. We called SNPs following
268 the GATK best practices (41). After filtering low quality variants we retained 11,530,110 SNPs
269 (referred to as 11.5 M SNPs), genotyped for 72 individuals of the nine southern capuchino
270 species (mean depth of coverage per species: nig 3.8x, pil 4.1x, mel 4.8x, hypox 4.3x, pal 5.7x,
271 bou 7.0x, cin 4.1x, hypoch 5.9x, ruf 5.2x). See SI Materials and Methods for details.
272
273 **Population genomic analyses**. We calculated F_{ST} values for non-overlapping 25 kb and 5 kb
274 windows, and for individual SNPs using VCFtools (42). These statistics were calculated for the
275 ten different pairwise comparisons across the five capuchinos with larger sample sizes (nig, pil,
276 mel, hypox, and pal). We identified divergence peaks using the average F_{ST} value for 25 kb
277 windows, discarding regions with less than two windows and windows with less than 10 SNPs.
278 The species with smaller sample sizes were not used to identify divergence peaks. We selected
279 regions that showed an F_{ST} value elevated above 0.2 (between 12 and 13 standard deviations
280 above the overall F_{ST} mean). We only retained candidate regions that had at least one individual
281 SNP with an F_{ST} of 0.85 or higher, which we considered a putative target of selection. In total,
282 we identified 25 divergent regions using these criteria. See SI Materials and Methods for details
283
284 **Acknowledgements**
285
286 This project was funded by NSF grant DEB – 1555754 to IJL, by PICT 2014-2154 (ANPCyT,
287 Argentina) to PLT and by a Genomics Scholar fellowship from the Center for Vertebrate
288 Genomics (Cornell University) to LC. We thank the members of the Fuller Evolutionary Biology
289 lab group (particularly B Butcher, D Toews, S Taylor, N Mason, S Aguillon, J Walsh-Emond, P
290 Dean-Coe, N Hofmeister, J Berv and B Van Doren), C Dardia, J Lewis, D Lijtmaer, and R
291 Razavi for support and feedback on this project. We thank government agencies of Brazil
292 (CNPq, FAPESP, ICMBio, SISBIO 36881-1, and CEMAVE 361788) and Argentina for grants
293 and permits. For tissue loans, we are indebted to the Museo Argentino de Ciencias Naturales
294 "Bernardino Rivadavia" (MACN, CONICET, Argentina), University of Kansas Natural History
295 Museum (USA) and Museu de Ciências e Tecnologia of the Pontifícia Universidade Católica do
296 Rio Grande do Sul (Brazil). Illustrations in Fig. 1A were obtained with permission from (43).
297
298 **References**
299
300 1. Martin A, et al. (2012) Diversification of complex butterfly wing patterns by repeated
301 regulatory evolution of a Wnt ligand. *Proc Natl Acad Sci USA*: 109(31):12632-12637.
302
303 2. Linnen CR, et al. (2013) Adaptive evolution of multiple traits through multiple mutations at a
304 single gene. *Science*. 339(6125):1312-1316.
305
306 3. Theron E, Hawkins K, Bermingham E, Ricklefs RE, Mundy NI (2001) The molecular basis of
307 an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is
308 perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Curr*
309 *Biol* 11(8):550-557.
310

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

311 4. Brawand D, et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish.
312 *Nature* 513(7518):375-381.
313
314 5. Küpper C, et al. (2016) A supergene determines highly divergent male reproductive morphs in
315 the ruff. *Nature Genet* 48(1):79–83.
316
317 6. Lamichhaney S, et al. (2016) Structural genomic changes underlie alternative reproductive
318 strategies in the ruff (*Philomachus pugnax*). *Nature Genet* 48(1):84-88.
319
320 7. Tuttle EM, et al. (2016) Divergence and degradation of a sex chromosome-like supergene.
321 *Curr Biol* 26(3):344–350.
322
323 8. Boag PT, Grant PR (1981) Intense natural selection in a population of Darwin's finches
324 (Geospizinae) in the Galapagos. *Science* 214(4516):82-85.
325
326 9. Abzhanov A, et al. (2006) The calmodulin pathway and evolution of elongated beak
327 morphology in Darwin's finches. *Nature* 442(7102):563-567.
328
329 10. Lamichhaney S, et al. (2015) Evolution of Darwin's finches and their beaks revealed by
330 genome sequencing. *Nature* 518(7539):371-375.
331
332 11. Lamichhaney S, et al. (2016) A beak size locus in Darwin's finches facilitated character
333 displacement during a drought. *Science* 352(6284):470-474.
334
335 12. Podos J (2001) Correlated evolution of morphology and vocal signal structure in Darwin's
336 finches. *Nature* 409(6817):185-188.
337
338 13. Ratcliffe LM, Grant PR (1985) Species recognition in Darwin's finches (Geospiza, Gould).
339 III. Male responses to playback of different song types, dialects and heterospecific songs. *Anim*
340 *Behav* 33(1):290-307.
341
342 14. Campagna L, et al. (2012) Rapid phenotypic evolution during incipient speciation in a
343 continental avian radiation. *Proc R Soc Lond [Biol]* 279(1734):1847-56.
344
345 15. Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ (2015) Distinguishing noise from
346 signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Mol Ecol*
347 24(16):4238-4251.
348
349 16. Grant PR, Grant BR (2014) 40 years of evolution: Darwin's finches on Daphne Major island.
350 (Princeton University Press, New Jersey), pp 304.
351
352 17. Burns KJ, et al. (2014) Phylogenetics and diversification of tanagers (Passeriformes:
353 Thraupidae), the largest radiation of Neotropical songbirds. *Molec Phylogenet Evol* 75:41-77.
354
355 18. Benites P, Campagna L, Tubaro PL (2015) Song‐based species discrimination in a rapid
356 Neotropical radiation of grassland seedeaters. *J Avian Biol* 46(1):55-62.

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

357
358 19. Meyer de Schauensee R. (1952) A review of the genus *Sporophila*. *Proc Acad Nat Sci Phil*
359 54: 153-198.
360
361 20. Machado É, Silveira LF (2011) Plumage variability and taxonomy of the Capped Seedeater
362 *Sporophila bouvreuil* (Aves: Passeriformes: Emberizidae). Zootaxa 2781:49-62.
363
364 21. Areta JI, Repenning M (2011) Systematics of the Tawny-bellied Seedeater (*Sporophila*
365 *hypoxantha*). I. Geographic variation, ecology, and evolution of vocalizations. *Condor*
366 113(3):664-677.
367
368 22. Di Giacomo AS, et al. (2010) Landscape associations of globally threatened grassland birds
369 in the Aguapey river Important Bird Area, Corrientes, Argentina. *Bird Conserv Int* 20(01):62-73.
370
371 23. Visser M, Kayser M, Palstra RJ (2012) HERC2 rs12913832 modulates human pigmentation
372 by attenuating chromatin-loop formation between a long-range enhancer and the OCA2
373 promoter. *Genome Res* 22(3):446-455.
374
375 24. Mundy NI (2005) A window on the genetics of evolution: MC1R and plumage colouration in
376 birds. *Proc R Soc Lond [Biol]* 272(1573):1633-1640.
377
378 25. Price, T (2007) Speciation in birds. (Roberts and Company, Colorado), 1-470 pp.
379
380 26. Poelstra JW, et al. (2014) The genomic landscape underlying phenotypic integrity in the face
381 of gene flow in crows. *Science* 344(6190):1410-1414.
382
383 27. Uy JA, et al. (2016) Mutations in different pigmentation genes are associated with parallel
384 melanism in island flycatchers. *Proc R Soc Lond [Biol]* 283(1834):20160731.
385
386 28. Toews DP, et al. (2016) Plumage Genes and Little Else Distinguish the Genomes of
387 Hybridizing Warblers. *Curr Biol* 26:2313–2318.
388
389 29. Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends Ecol Evol*
390 23:38-44.
391
392 30. Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of
393 mimicry adaptations among species. *Nature* 487(7405):94-98.
394
395 31. Wallbank RW, et al. (2016) Evolutionary novelty in a butterfly wing pattern through
396 enhancer shuffling. *PLoS Biol* 14(1):e1002353.
397
398 32. Charlesworth B (2001) The effect of life-history and mode of inheritance on neutral genetic
399 variability. *Genet Res* 77(02):153-166.
400
401 33. Charlesworth B, Coyne JA, Barton NH (1987) The relative rates of evolution of sex
402 chromosomes and autosomes. *Am Nat* 1:113-146.

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

403
404 34. Haldane JB (1922) Sex ratio and unisexual sterility in hybrid animals. *J Genet* 12(2):101-
405 109.
406
407 35. Gnerre S, et al. (2011) High-quality draft assemblies of mammalian genomes from massively
408 parallel sequence data. *Proc Natl Acad Sci USA* 108(4):1513-1518.
409
410 36. English AC, et al. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS
411 long-read sequencing technology. *PloS one* 7(11):e47768.
412
413 37. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO:
414 assessing genome assembly and annotation completeness with single-copy orthologs.
415 *Bioinformatics*:btv351.
416
417 38. Cantarel BL, et al. (2008) MAKER: an easy-to-use annotation pipeline designed for
418 emerging model organism genomes. *Genome Res* 18(1):188-196.
419
420 39. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large
421 gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44-57.
422
423 40. Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*
424 9:357-359.
425
426 41. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for
427 analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
428
429 42. Danecek P et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156-
430 2158.
431
432 43. del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E (2016) Handbook of the Birds of
433 the World Alive (Lynx Edicions, Barcelona).
434
435 **Figure Legends**
436
437 Figure 1. Genomic landscapes in southern capuchino seedeaters. (A) Map indicating the extent
438 of range overlap in the nine species; note that up to six species breed sympatrically in north-
439 eastern Argentina. The range of each species is outlined by dashed lines, with colors matching
440 species names. The schematic phylogeny shows the relationships among species obtained from
441 (14) and (15) (see text for name abbreviations). (B) PCA including 56 individuals of five species
442 genotyped at ~11.5 million SNPs, and a second PCA (60 individuals) using SNPs from
443 divergence peaks alone. Four outlier individuals were omitted from the first PCA (see Fig. S1
444 and SI Materials and Methods for details). (C) Manhattan plots for nig vs. mel (top) and hypox
445 vs. pal (bottom); 12 individuals per species. Each circle indicates the mean F_{ST} value for all the
446 SNPs within a non-overlapping 25 kb window. Scaffolds in the reference genome were sorted by
447 decreasing size and are indicated by alternating colors. The threshold for calling divergent
448 windows is indicated by the dashed red line, and the percentage of total elevated windows is

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

449 noted next to each comparison. The inset is a histogram showing the width distribution (in kb)
450 for the 25 divergence peaks we identified.
451
452 Figure 2. Repeated selection on pigmentation genes in different capuchino species. (A) The
453 divergence peak on scaffold 252, which mapped to chromosome 20 in the Zebra Finch. The ten
454 possible pairwise comparisons across five capuchino species are overlaid (see color-code legend
455 to identify specific comparisons). Each circle is the mean F_{ST} value for all SNPs within a non-
456 overlapping 5 kb window. (B) PCA for 60 individuals of 5 species using the SNPs from under
457 the peak in (A); see the color-code in the legend to identify the species. (C) F_{ST} and genomic
458 location of individual SNPs with values of 0.85 and higher, color-coded by pairwise comparison
459 as in (A). The positions of genes that are close to these highly divergent SNPs are indicated by
460 arrows. Names in red note genes involved in the melanogenesis pathway. (D, E, and F) As
461 above, for the divergence peak on scaffold 412. (G, H, and I) As above, for the divergence peak
462 on scaffold 404. (J, K, and L) As above, for the divergence peak on scaffold 257. (K and L) The
463 top plot corresponds to the peak labeled "A" and the bottom one to the peak labeled "B" in (J).
464 Annotations with question marks did not match genes with known names.

Table 1. Areas of the genome that are highly differentiated in capuchino seedeaters.

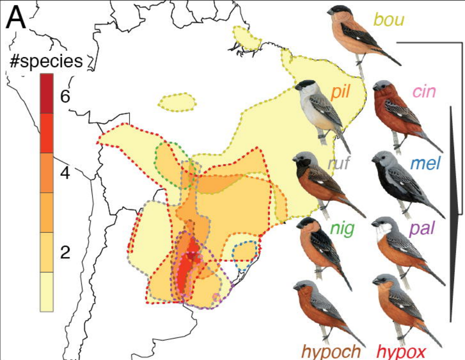
Scaffold	Chromosome	Peak size (kb)	Highest F_{ST} for 5 kb window	SNPs with F_{ST}>0.85 (=1)[1]	Melanogenesis gene	Function of melanogenesis gene[2]	Total #genes (known function)[3]	Figure
252	20	90	0.83	383(57)	ASIP	Induces melanocytes to synthesize pheomelanin (yellow) instead of eumelanin (black/brown)	6(4)	2
762	11	35	0.63	178(22)	-	-	9(9)	S4
412	1A	205	0.70	113	KITL	Stimulates melanocyte proliferation	1(1)	2
308	Unknown	140	0.65	112(14)	-	-	4(4)	S4
430	1	85	0.48	105	-	-	4(1)	S4
404	Z	765	0.60	92(5)	SLC45A2	Transports substances needed for melanin synthesis	21(20)	2
257 (A)	Z	500	0.53	70	TYRP1	Enzyme important for melanin biosynthesis	7(6)	2
1717	4	30	0.78	70	CAMK2D	Cell communication	9(7)	S3
3622	1	385	0.48	37	-	-	44(17)	-
257 (B)	Z	840	0.54	29	MLANA	Plays a role in melanosome biogenesis	28(20)	2
567	2	260	0.65	23(1)	-	-	8(4)	-
579	1	670	0.33	13	-	-	16(13)	-
1954	5	75	0.57	11	-	-	3(2)	-
59	15	85	0.35	8	-	-	6(2)	-
118	2	25	0.61	7	-	-	3(2)	-
257	Z	265	0.39	6	-	-	11(6)	-
404	Z	195	0.55	5	-	-	3(3)	-
257	Z	370	0.53	4	-	-	23(10)	-
766	4	65	0.53	4	-	-	6(6)	-
637	Z	40	0.45	4	-	-	0	-
1635	6	30	0.46	3	-	-	1(1)	-
257	Z	45	0.31	2	-	-	16(15)	-
263	Z	535	0.34	2	MYO5A	Actin-based motor protein involved in melanosome transport	24(11)	S3
791	1	50	0.40	1	TYR or DCT[4]	TYR: Enzyme involved in converting tyrosine to melanin. DCT: Regulates eumelanin and phaeomelanin levels.	4(3)	S3
637	Z	300	0.30	1	-		0	-

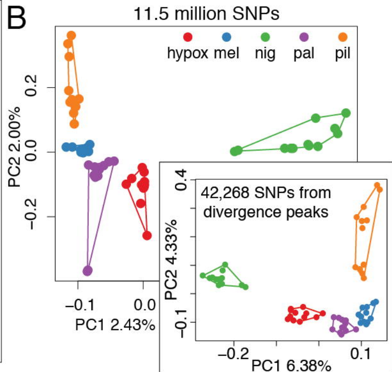
[1]Scaffolds are ranked from top to bottom by the number of highly divergent SNPs (F_{ST}>0.85).
[2]Information from https://david.ncifcrf.gov/ and www.genecards.org.

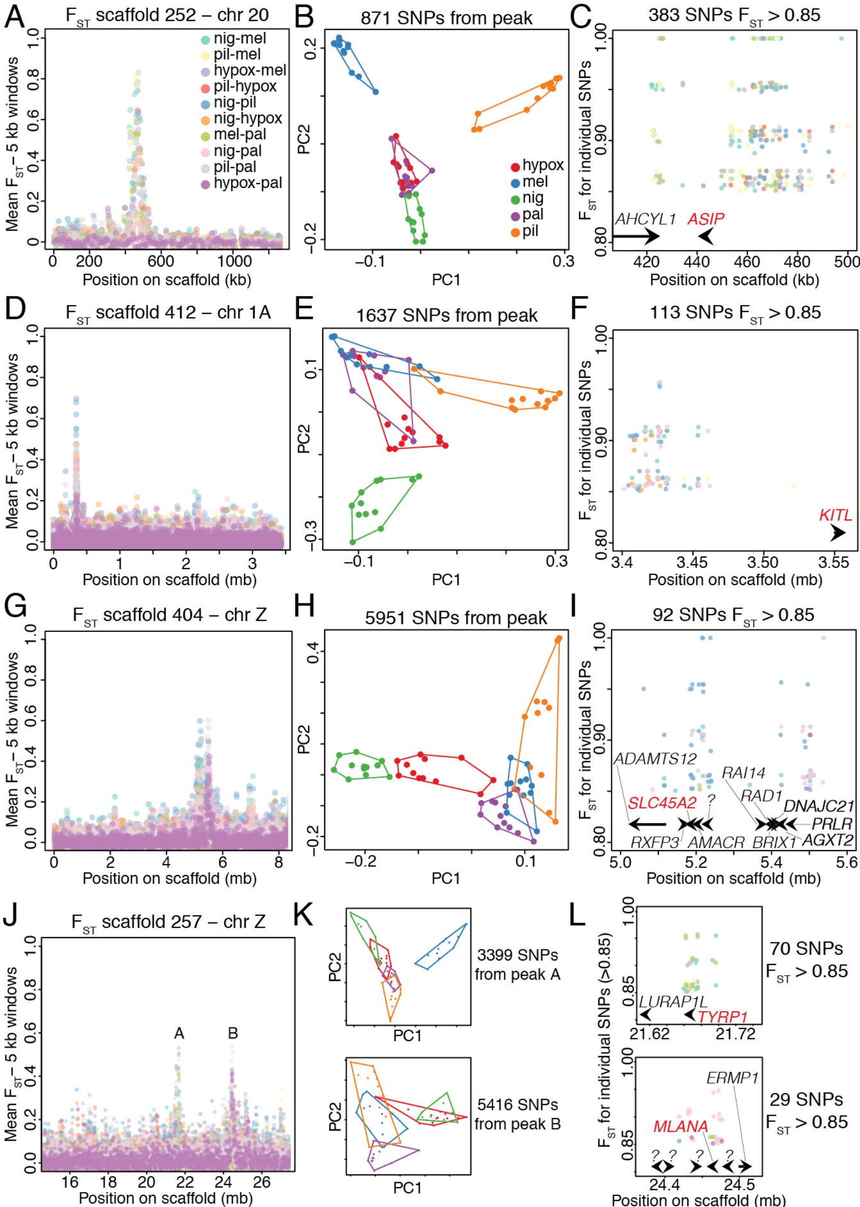
bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[3]Total of 257 annotations of which 246 were unique and 11 were predicted more than once. Approximately 63% (156) of the unique annotations matched a record in the UniProt database with a known name and function.
[4]Annotation predicted based on protein similarity of both TYR and DCT, which in the Zebra Finch are on chromosome 1.

bioRxiv preprint doi: https://doi.org/10.1101/075713; this version posted September 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


A
#species
6
4
2
bou
pil
cin
ruf
mel
nig
pal
hypoch
hypox


B
11.5 million SNPs
hypox mel nig pal pil
PC2 2.00%
0.2
0.0
−0.2
−0.1
0.0
PC1 2.43%
42,268 SNPs from divergence peaks
PC2 4.33%
0.4
−0.1
−0.2
0.1
PC1 6.38%
C
nig vs. mel (0.3%)
hypox vs. pal (0.03%)
Mean F_{ST} 25 kb windows
1
0.8
0.6
0.4
0.2
0
Freq.
12
0
400
800
kb
Scaffolds ordered by size


A
F_ST scaffold 252 – chr 20
nig-mel
pil-mel
hypox-mel
pil-hypox
nig-pil
nig-hypox
mel-pal
nig-pal
pil-pal
hypox-pal
Mean F_ST – 5 kb windows
Position on scaffold (kb)
B
871 SNPs from peak
hypox
mel
nig
pal
pil
PC1
PC2
C
383 SNPs F_ST > 0.85
AHCYL1
ASIP
F_ST for individual SNPs
Position on scaffold (kb)
D
F_ST scaffold 412 – chr 1A
Mean F_ST – 5 kb windows
Position on scaffold (mb)
E
1637 SNPs from peak
PC1
PC2
F
113 SNPs F_ST > 0.85
KITL
F_ST for individual SNPs
Position on scaffold (mb)
G
F_ST scaffold 404 – chr Z
Mean F_ST – 5 kb windows
Position on scaffold (mb)
H
5951 SNPs from peak
PC1
PC2
I
92 SNPs F_ST > 0.85
ADAMTS12
SLC45A2
RXFP3
AMACR
RAI14
RAD1
DNAJC21
PRLR
BRIX1
AGXT2
F_ST for individual SNPs
Position on scaffold (mb)
J
F_ST scaffold 257 – chr Z
A
B
Mean F_ST – 5 kb windows
Position on scaffold (mb)
K
3399 SNPs from peak A
5416 SNPs from peak B
PC1
PC2
L
70 SNPs F_ST > 0.85
LURAP1L
TYRP1
29 SNPs F_ST > 0.85
ERMP1
MLANA
F_ST for individual SNPs (>0.85)
Position on scaffold (mb)