bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Selection and local adaptation in capuchin monkeys revealed through fluorescence-activated cell sorting of feces (fecalFACS)

Joseph D. Orkin[1,2*], Michael J. Montague[3], Daniela Tejada-Martinez[4,5], Marc de Manuel[2], Javier del Campo[6], Anthony Di Fiore[7,8], Claudia Fontsere[2], Jason A. Hodgson[9,10], Mareike C. Janiak[1,11], Lukas F.K. Kuderna[2], Esther Lizano[2,18], Yoshihito Niimura[12], George H. Perry[9,13], Jia Tang[1], Wesley C. Warren[14], João Pedro de Magalhães[5], Shoji Kawamura[15], Tomàs Marquès-Bonet[2,16,17,18], Roman Krawetz[19], Amanda D. Melin[1,11,20*]

1 Department of Anthropology and Archaeology, University of Calgary, Calgary, Canada
2 Institut de Biologia Evolutiva, Universitat Pompeu Fabra-CSIC, Barcelona, Spain
3 Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19146
4 Doctorado en Ciencias mención Ecología y Evolución, Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
5 Integrative Genomics of Ageing Group, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool L7 8TX, UK
6 Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA
7 Department of Anthropology and Primate Molecular Ecology and Evolution Laboratory, University of Texas at Austin
8 College of Biological and Environmental Sciences, Universidad San Francisco de Quito, Cumbayá, Ecuador
9 Department of Anthropology, The Pennsylvania State University, State College, PA, USA
10 Department of Zoology, University of Cambridge, Cambridge, UK
11 Alberta Children's Hospital Research Institute, Calgary, AB, Canada
12 Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
13 Department of Biology, Huck Institute of Life Sciences, The Pennsylvania State University, State College, PA, USA
14 Division of Animal Sciences, School of Medicine, University of Missouri, Columbia, MO, 65211, USA
15 Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan
16 Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain
17 CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain
18 Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain
19 Department of Cell Biology and Anatomy, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

20 Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

* Corresponding authors

ABSTRACT

Background: Capuchins have the largest relative brain size of any monkey and a remarkable lifespan of 55 years, despite their small body size. Distributed widely across Central and South America, they are inventive and extractive foragers, known for their sensorimotor intelligence, dietary diversity, and ecological flexibility. Despite decades of research into their ecology and life history, little is known about the genomics of this radiation.

Results: We assemble a *de novo* reference genome for *Cebus imitator*, and provide the first genome annotation of a capuchin monkey. We identified 20,740 and 9,556 for protein-coding and non-coding genes, and recovered 23,402 orthologous groups. Through a comparative genomics approach across a diversity of mammals, we identified genes under positive selection associated with longevity and brain development, which are of particular relevance to capuchin and primate comparative biology. Additionally, we compared populations in distinct habitats, facilitated by our novel method for minimally-biased, whole-genome sequencing from fecal DNA using fluorescence activated cell sorting (FACS). By analyzing 23 capuchin genomes from tropical dry forest and rainforest, we identified population divergence in genes involved in water balance, kidney function, and metabolism, consistent with local adaptation to resource seasonality.

Conclusions: Our comparative study of capuchin genomics provides new insights into the molecular basis of brain evolution and longevity. These data also improve our understanding of processes of local adaptation to diverse and physiologically challenging environments. Additionally, we provide a technological advancement in use of non-invasive genomics to study free-ranging mammals through FACS.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

KEYWORDS

Brain size, longevity, intelligence, seasonality, dry forest, non-invasive genomics, flow cytometry, platyrrhine

BACKGROUND

Large brains, long lifespans, extended juvenescence, tool use, and problem solving are hallmark characteristics of the great apes, and are of enduring interest in studies of human evolution [1–6]. Similar suites of traits have arisen in other lineages, including some cetaceans, corvids and, independently, in another radiation of primates, the capuchin monkeys. Like great apes, they have diverse diets, consume and seek out high-energy resources, and engage in complex extractive foraging techniques [7,8] to consume difficult-to-access invertebrates and nuts [8]. Their propensity for tool use and their ecological flexibility may have contributed to their convergence with the great apes [9]offering opportunities for understanding their evolution via comparative methods [10–12].

Capuchins also offer excellent opportunities to study local adaptation to diverse habitats. While little is known about the migration, history, and population dynamics of this species, they presently range from Panama to northern Honduras [13–15], where they occupy a wide diversity of habitats, including rainforests and, in the northern extent of their range, tropical dry forests. Particular challenges of the tropical dry forest are staying hydrated during the seasonally prominent droughts, high temperatures in the absence of foliage, and coping metabolically with periods of fruit dearth (Figure 1). The sensory challenges of food search in dry versus humid biomes are also distinct. For example, odor detection and propagation is affected by temperature and humidity [16,17], and color vision is hypothesized to be adaptive in the search for ripe fruits and young reddish leaves against a background of thick, mature foliage [18], which

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

is absent for long stretches in dry deciduous forests. The behavioral plasticity of capuchins is widely acknowledged as a source of their ability to adapt to these dramatically different habitats [19–22]. However, physiological processes including water balance and metabolic adaptations to low caloric intake, and sensory adaptations to food search, are also anticipated to be targets of natural selection, as seen in other mammals [23–26].

Comparative genomics offers a unique opportunity to examine the molecular underpinnings of traits relevant to the evolution of humans and other great apes. Furthermore, the ecological flexibility of white-faced capuchins allows us to assess the influence of divergent and changing habitats on the process of adaptive evolution in these primates. In order to address the genetic underpinnings of capuchin adaptive evolution, we assembled the first reference genome of *Cebus imitator*. Additionally, to better understand the local adaptation of capuchins, we conducted high-coverage re-sequencing (10X - 47X) of 10 individuals from populations inhabiting distinct environments--one in a lowland evergreen rainforest (n=4), and another from a lowland tropical dry forest (n=6). Importantly, to facilitate the population-wide analyses without the need for potentially harmful invasive sampling of wild primates, we sequenced an additional 13 individuals at low coverage using a novel method for minimally-biased, whole-genome sequencing from fecal DNA using fluorescence-activated cell sorting (fecalFACS) that we developed (Figure 2). In our positive selection analyses, we focus on genes that may underlie sensation, cognition, and lifespan due to their relevance to capuchin-specific biology and adaptation. In our population comparison we predict genes related to water balance, metabolism will differ between dry forest and rainforest populations, reflecting local adaptation to different habitats. With respect to sensory systems, due to higher diversity of flora and fauna, including colorful fruits presented against green foliage, and increased humidity, genes underlying color vision (opsin genes) and chemosensation (olfactory and taste receptor genes) are predicted to be more diverse in rainforest habitats [27–29].

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

RESULTS

1. Comparative Genomics

Genome assembly and gene annotation

Our reference genome assembly for *Cebus imitator* is comprised of 7,742 scaffolds (including single contig scaffolds) with an N50 scaffold length of 5.2 Mb and an N50 contig length of 41 kb. The final ungapped assembly length is 2.6 Gb (GenBank accession: GCA_001604975.1). Our estimate of total interspersed repeats using WindowMasker [30] output is 45.8%. The numbers of annotated genes are 20,740 and 9,556 for protein-coding and non-coding genes, respectively (Table S1). Measures of gene representation using the known human RefSeq set of 56,230 transcripts show an average of >94% coverage with a mean identity of 92.5%. Overall, our draft assembly metrics and gene representation are consistent with other non-human primate (NHP) short-read reference assemblies [31].

Positive selection analyses

We recovered 23,402 Orthologous Groups (OGs). Capuchins share 18,475 OGs with human, 17,589 OGs with rhesus macaque, 15,582 OGs with mouse, and 14,404 OGs with dog. When we included orthologous genes that are present simultaneously in all 15 species, we recovered 7,519 OGs, which we subsequently used in the natural selection analyses ($d_N/d_S=\omega$). We identified 612 genes under positive selection ($p<0.05$ after FDR correction) in the *Cebus* lineage using the branch model. We also performed a branch-site test using codeml in PAML[32] and identified a second set of 748 genes under positive selection in *Cebus* (Table S2 Sheet 2). The results of our enrichment analysis for biological processes using DAVID [33]

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

identified genes that underlie brain development, sensation, and lifespan, which were of particular interest given the derived features of capuchin biology (Figure 3).

Brain development and longevity

Capuchins have the largest brain to body ratio of any monkey, and are known for their sensorimotor intelligence [7] and derived cognitive abilities [8]. Of the 748 genes identified as being under positive selection in the branch-site model, 17 were previously associated with brain development (Table S2, Sheet 6, row 18) and 5 were linked to neurogenesis (Table S2, Sheet 6, row 116). For example, *WDR62*, *BPTF*, *BBS7*, *NUP113*, mutations are directly associated with brain size and related malformations, including microcephaly [34–37]. *MTOR* signaling malfunction is also implicated in developmental brain malformations [38], and *NUP113* is involved in nuclear migration during mammalian brain development [39]. Several genes are linked with brain tumor formation (including *ZNF217*), and others with cognitive ability (e.g. *PHF8 [40]*).

We found 27 aging-related genes, as identified in the GenAge database [41,42], under positive selection in capuchins including *PARP1*, *MTOR*, *SREBF1*, *INSR*, *HTT*, *RB1* and *MDM2* (Table S2, Sheet 7). Of note, poly (ADP-ribose) polymerase 1 (PARP1) putatively serves as a determinant of mammalian aging due to its activity in the recovery of cells from DNA damage. In previous studies, gene expression levels of *PARP1* were inversely correlated with mammalian lifespan [43]. Another large body of research has associated the mechanistic target of rapamycin (*MTOR*) with aging and longevity in various organisms [44], making it a prime candidate for therapeutic interventions in aging [45]. *MTOR* acts as a regulator of cell growth and proliferation while also being generally involved in metabolism and other processes. Additional key genes in aging and metabolism include sterol regulatory element binding transcription factor 1 (*SREBF1*), which acts as a regulator of the metabolic benefits of caloric restriction [46,47], and the insulin

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

receptor (*INSR*), a major player in longevity regulation [48]. As for specific age-related diseases, huntingtin (*HTT*) is under selection in mammals; *HTT* is not only involved in Huntington's disease but has also been associated with longevity in mice [49]. Lastly, various cell cycle regulators (e.g., *RB1*, *MDM2*) are also under positive selection in capuchins, and indeed, cell cycle is an enriched term among positively selected genes (Table S2), though these could be related to other life history traits like developmental schedules that correlate with longevity.

2. Population Genomics and Local Adaptation

Population structure, genetic diversity, and demographic history in Costa Rican white-faced capuchins

Of the 24 capuchin DNA samples that we sequenced and mapped to our reference genome, 15 were fecal-derived (Table S3). When comparing the high coverage tissue-derived genomes from the Santa Rosa site to those generated from our novel application of fluorescence-activated cell sorting to isolate fecal-sourced cells (fecalFACS), we observed no substantial difference in quality, coverage, heterozygosity, or GC content (Figures 2, S1, S2, Supplemental Text). This includes the first (to our knowledge) high-coverage (12.2 X) whole mammalian genome generated from a fecal sample.

The pattern of clustering in our maximum likelihood single nucleotide variant (SNV) tree recapitulates the expected patterns of geographic distance and ecological separation in our samples (Figure 4). Likewise, in the projected PCA all individuals from the seasonal dry forests in the northwest are sharply discriminated from individuals inhabiting the southern rainforests along PC1. These relationships are not perturbed by depth of coverage, or source material (tissue-based vs fecalFACS genomic libraries) (Figure 4). Levels of heterozygosity calculated in overlapping 1 Mb genomic windows (with a step size of 100 kb) were significantly higher in the

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

202 southern population (W = 1,535,400,000, p-value < 2.2e-16; Figure 5A, Figure S3).
203 Furthermore, the median pairwise heterozygosity for each southern individual (range: 0.00065 -
204 0.00071) was higher than any northern monkey (0.00047 - 0.00057) (W = 0, p-value =
205 0.009524; Table S5). In the northern population, we also identified long runs of homozygosity
206 significantly more often (W = 24, p-value = 0.009524), and more of the longest runs (>= 5 Mb)
207 (W = 1315.5, p-value = 0.03053; Figures 5B, S4, S5). Pairwise sequential Markovian coalescent
208 (PSMC) analysis of demographic history (Figure 5C) reveals that white-faced capuchins had a
209 peak effective population size of ~60,000 effective individuals ~1 mya, which declined to fewer
210 than 20,000 during the middle to late Pleistocene transition. After recovering during the middle
211 Pleistocene, they declined precipitously through the late Pleistocene to fewer than 5,000
212 effective individuals.
213
214 ***Local adaptation to seasonal dry forest biome***
215
216 We predicted that genes related to water balance, metabolism, and sensation would
217 differ between dry forest and rainforest populations, reflecting local adaptation to different
218 habitats. To test this, we searched for associations between genes in windows with high F_{ST} and
219 divergent non-synonymous SNVs with high or moderate effect. Of the 299 genes identified in
220 high F_{ST} windows, 39 had highly differentiated non-synonymous SNVs (Figure S6, Tables S6,
221 S7). Our enrichment analysis identified a single significant GO biological process: regulation of
222 protein localization to cilium (GO: 1903564). Remarkably--and in accordance with our
223 hypothesis--disruptions of cilia proteins are predominantly associated with disorders of the
224 kidney and retina (ciliopathies) [50]. Furthermore, enrichment of genes associated with disease
225 states was linked to kidney function, metabolism, and muscular wasting. These genes are good
226 candidates for adaptive resilience to seasonal water and food shortages and warrant further
227 investigation. We highlight several genes of particular promise (Figure 6, Figure S7).

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

228 Evidence of adaptation to food and water scarcity

229

230 Multiple candidate genes indicate that dry-forest capuchins could be adapted to
231 seasonal drought-like conditions and food scarcity. *SERPINC1* encodes antithrombin III, which
232 is involved in anticoagulant and anti-inflammatory responses associated with numerous kidney-
233 related disorders including salt-sensitive hypertension, proteinuria, and nephrotic syndrome [51–
234 56]. *SERPINC1* is overexpressed in the renal cortex of Dahl salt-sensitive rats fed a high salt
235 diet [55], suggesting that variants could be facilitating water/salinity balance in the dry-forest
236 capuchins. Additionally, sequence variants of *AXDND1* (as identified in the GeneCards
237 database) are associated with nephrotic syndrome, a kidney disorder that commonly presents
238 with edema and proteinuria. *BCAS3* is expressed in multiple distal nephron cells types [57], and
239 is associated with four pleiotropic kidney functions (concentrations of serum creatinine, blood
240 urea nitrogen, uric acid, and the estimated glomerular filtration rate based on serum creatinine
241 level) [58]. Although we did not identify any population-specific non-synonymous SNVs in
242 *BCAS3*, the genomic windows encompassing the gene rank among the highest regions of F_{ST} in
243 our dataset, and intronic variation in BCAS3 putatively impacts estimated glomerular filtration
244 rate in humans [57].
245 The association of *BCAS3* and serum creatinine is of particular interest, given field
246 observations of capuchins from SSR, whose urinary creatinine levels are associated with
247 decreased muscle mass during periods of seasonally low fruit availability [59]. Creatinine is a
248 byproduct of the metabolism of creatine phosphate in skeletal muscle, which is normally filtered
249 by the kidneys, and can be used as a clinical biomarker of kidney function, chronic kidney
250 disease [58], and as a monitor of muscle mass [59,60]. Multiple congenital muscular
251 dystrophies--including mild forms that present with muscular wasting--and abnormal circulating
252 creatine kinase concentration (HP:0040081) are associated with candidate genes *ITGA7, ISPD*
253 (*CRPPA*), and SYNE2 [61–64]. Furthermore, transgenic overexpression of *ITGA7* has been

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

shown to reduce muscular pathologies caused by mutations in *LAMA2* [65], which falls in a high F_{ST} window. The relationship among seasonal resource availability, kidney function, and muscle mass is further supported by a potential adaptive role in capuchin sugar metabolism and frugivory. In particular, *GLIS3* is one of two genes known to be associated with type 1 diabetes, type 2 diabetes, and neonatal diabetes [66], and appears to be diverging between populations. Additionally, while insulin receptor substrate (*IRS4*) did not fall in a high F_{ST} window, we observed a non-synonymous SNP of medium effect fixed between populations. Given the appearance of both diabetes and kidney disorders in our gene sets, we conducted an *a posteriori* search of our high F_{ST} gene set for overrepresentation of genes associated with diabetic nephropathy (EFO_0000401) in the GWAS catalogue (Table S6). Seven genes were present in both our gene set of 299 high F_{ST} genes and the diabetic nephropathy set of 117. Given the 16,553 annotated genes with HGNC Gene IDs, seven overlapping genes would occur with $p = 0.00046$ when permuted 100,000 times. We take this as promising evidence that these genes have been under selection in the northwestern population.

Evidence of adaptation in sensory systems

Given the ecological differences between the northern dry and southern wet forests, we predicted that the evergreen, humid environment of the lowland rainforest would favor enhanced diversity of both color vision (opsin genes) and olfaction (olfactory receptor genes) would result in population specific variation in chemosensory and visual genes. We did not find support for this; in both populations, we observed similar polymorphism at each of the three medium/long wave cone opsin tuning sites (180, A/S; 277, F/Y; and 285, T/A) (Table S8). None of these codons is a novel variant [67,68], providing no support for the hypothesis of differences in the perception of photopic (cone-driven) vision between the two biomes. However, we did observe some evidence for population specific variation associated with the photoreceptive layers of the

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

retina. First, creatine kinase plays an important role in providing energy to the retinal pigment epithelium [69]. Secondly, we identified a fixed non-synonymous SNV in *CCDC66*, which falls in a high F_{ST} region and is heavily expressed in photoreceptive layers of the retina. Electroretinography of *CCDC66* -/- mice reveals a significant reduction in scotopic (rod-driven) photoreceptor response [70], indicating a potential effect on vision in low-light conditions. Curiously, *CCDC66* -/- mice, also display neurodegeneration of the olfactory bulb, and have reduced odor discrimination performance of lemon smells [71]. Turning to olfaction, we identified 614 olfactory receptor (OR) genes and pseudogenes in the capuchin reference genome (408 intact, 45 truncated, and 161 pseudogenized (Table S9). To test for population differences in the OR gene repertoire, we assembled each olfactory gene/pseudogene independently in each individual. The proportion of total functional ORs was stable across individuals and populations, with trivial fluctuations (North $\bar{x}$ = 411, s = 1.6; South $\bar{x}$ = 408.5, s = 1.3), possibly driven by a small difference in OR family 5/8/9 (Table S10). We also identified 8 vomeronasal and 28 taste receptor and taste receptor-like genes (Table S11), two of which (*TAS1R* and *TAS2R4*) have non-synonymous SNVs with fixed variants in the north (Table S12). The functional significance of these variants is unknown, but may be revealed via cellular expression systems in future research [72].

DISCUSSION

Comparative genomics of white-faced capuchins

Among primates, capuchin monkeys are known for their relatively large brains, cognitive capacity, and sensorimotor intelligence [7,8,73,74]. Accordingly, it is perhaps unsurprising to see positive selection in the *Cebus* lineage and evidence of shifts in gene function linked to brain function and development relative to other primates. In particular, positive selection in

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

306 *WDR62*, *BPTF*, *BBS7*, *NUP133*, and *MTOR*, and *PHF8* supports our hypothesis that the
307 capuchin lineage has undergone adaptation linked to brain development. The association of
308 several of these genes with size related brain malformations [34–37], such as microcephaly,
309 suggests that they could be influencing the large relative brain size of capuchins. Furthermore,
310 the evidence of selection in PHF8, which is associated with human cognitive capacity, aligns
311 with the link between brain size and intelligence that has been observed in other primates [75].
312 While we highlight here the putative functional roles of these genes, which are based on clinical
313 studies and comparative genomics, we acknowledge that further examination of their function in
314 the context of capuchin biology is warranted.
315 In the context of longevity, it is noteworthy that we observed genes under selection
316 associated with DNA damage response, metabolism, cell cycle and insulin signaling [76]. Of
317 particular interest are: *PARP1*, *MTOR*, *SREBF1*, *INSR1*, and *HTT*. Damage to the DNA is
318 thought to be a major contributor to aging [77]. Previous studies have also shown that genes
319 involved in DNA damage responses exhibit longevity-specific selection patterns in mammals
320 [41]. It is therefore intriguing that *PARP1*, a gene suggested to be a determinant of mammalian
321 aging [43], is under selection in capuchins. Other genomes of long-lived mammals also revealed
322 genes related to DNA repair and DNA damage responses under selection [78,79]. In the context
323 of longevity, it is also noteworthy that we observed genes under selection associated with
324 metabolism, cell cycle and insulin signaling. Other genome sequencing efforts of long-lived
325 mammals also revealed changes in such pathways [79,80]. Intriguingly, short-lived species also
326 exhibit genes under selection related to insulin receptors, raising the possibility that the same
327 pathways associated with aging in model organisms are involved in the evolution of both short-
328 and long lifespans [81], an idea supported by our results. Of course, because aging-related
329 genes often play multiple roles, for example in growth and development, it is impossible to tell
330 for sure whether selection in these genes is related to aging or to other life-history traits, like
331 growth rates and developmental times, that in turn correlate with longevity [82]. Therefore,

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

332 although we should be cautious about the biological significance of our findings, it is tempting to
333 speculate that, like in other species, changes to specific aging-related genes or pathways, could
334 contribute to the longevity of capuchins.
335
336 **Population genomics and local adaptation with fecalFACS**
337
338 Through a novel use of flow cytometry/FACS, we have developed a new method for the
339 isolation of epithelial cells from mammalian feces for population genomics. We generated the
340 first high-coverage, minimally biased mammalian genome solely from feces. Additionally, we
341 have demonstrated that fecalFACS can be used to generate low coverage SNP datasets that
342 are suitable for population assignment and clustering. FecalFACS is cost-effective and
343 minimizes the biases that commonly occur in traditional bait-and-capture approaches to the
344 enrichment of endogenous DNA from feces. Furthermore, fecalFACS does not require costly
345 impractical preservation of biomaterial in liquid nitrogen; rather, we rely on room-temperature
346 stable storage in RNAlater. FecalFACS offers great benefits to the field of mammalian
347 conservation and population genomics.
348 White-faced capuchins are the most northerly distributed member of the Cebinae, having
349 dispersed over the isthmus of Panama in a presumed speciation event with *C. capucinus* in
350 South America ~1.6 mya [13–15]. After expanding during the early late Pleistocene, white-faced
351 capuchins appear to have undergone a dramatic reduction in effective population size. This
352 pattern predates the movement of humans into Central America, and could reflect a series of
353 population collapses and expansions caused by glacial shifts and fluctuating forest cover
354 availability during the Pleistocene. At a finer scale, we observed a clear demarcation between
355 the northern dry- and southern wet-forest populations in Costa Rica. Higher levels of
356 heterozygosity in the south and lower levels in the northwest are in accordance with the
357 hypothesis that capuchins dispersed northwards across Costa Rica. White-faced capuchins in

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

358 SSR are near the northernmost limits of their range, which extends as far north as Honduras,
359 and we predict they may represent some of the least genetically diverse members of their
360 species. Given the limitations of the available sampling sites, it is possible that the appearance
361 of an ecological divide is actually evidence of isolation by distance; however, given that the
362 single individual from Cañas clusters closely with the individuals from SSR, despite a
363 geographic distance of more than 100 km, we suggest that isolation by distance does not
364 completely explain the population differentiation.

365 We found evidence that the recent northern expansion of *Cebus imitator* has undergone
366 local adaptation to the extreme seasonality of rainfall and food availability. The effects of
367 seasonality have been linked to biological consequences for Costa Rican white-faced capuchins
368 in other contexts. Seasonal fluctuations in food abundance and rainfall are associated with
369 compositional changes in the gut microbiome of capuchins at SSR [83], which differs markedly
370 from that observed in capuchins inhabiting nonseasonal forests [84]. Previous capuchin
371 research in the dry forest also demonstrates seasonal negative energy balance and periods of
372 pronounced muscle loss through catabolic processes [59]. These observations fit with the notion
373 that animals living in seasonal environments, or pursuing seasonal migrations, are more likely to
374 have weight fluctuations through binge-subsist cycles that map onto food abundance [85].
375 Accordingly, we observed population-specific variation in genes implicated in water/salt balance,
376 kidney function, muscle wasting, and metabolism (e.g. *SERPINC1*, *BCAS3*, *ITGA7*, *ISPD, and*
377 *GLIS3*). In light of this, we contend that seasonal drought and food shortages would create an
378 environment favoring efficient catabolism when needed and adaptations for maintaining water
379 balance. Given that selection operates on both gene function and regulation, we suspect the
380 observed variation is affecting gene expression or enzymatic efficiency, which offers a
381 promising avenue for future research.

382 Additionally, we found evidence of 408 OR genes, 28 TASR, and 7 VR genes that are
383 putatively functional. These numbers are similar to, or slightly higher than, the number of

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

chemosensory genes identified in other anthropoid primates [86–90]. The VR gene repertoire of capuchins highlights the persistent role of the vomeronasal organ that is used in social communication of other mammals, but that has been nearly lost from all African and Asian monkeys [89]. Like most other primates in the Americas, capuchins possess an intriguing color vision system characterized by extensive intraspecific genotypic and phenotypic variation [32–35]. Contrary to our prediction, dry-forest and rainforest capuchins have similar numbers of color vision, taste and olfactory receptors. This indicates that there are not greater numbers of functional sensory genes in areas of higher vegetative biodiversity or humidity. The tuning of the chemosensory receptors may vary between habitat types and may be elucidated by future work. While we did observe population-specific variation in CCDC66, which is expressed in photoreceptive layers of the retina and may affect odor discrimination, the ecological significance of this result, if any, is unclear at present but may warrant future attention.

CONCLUSION

We provide the first annotated reference assembly for a capuchin monkey. We observed evidence of selection in *C. imitator* on genes involved with brain function and cognitive capacity. These results are in accordance with the remarkably long life span, large brain, and high degree of sensorimotor intelligence that has been observed in capuchins. These genes are good candidates for further investigation of traits which have evolved in parallel in apes and other mammals. Through a novel use of flow cytometry/FACS, we developed a new method for the isolation of epithelial cells from mammalian feces for population genomics. FecalFACS allowed us to generate both the first high-coverage, minimally biased mammalian genome solely from feces, as well as low coverage SNV datasets for population level analyses. In our population level analysis of wet- and dry-forest capuchins, we observed both evidence of population structure between and local adaptation to these different habitats. In particular, we identified

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

410 selection in genes related to food and water scarcity, as well as muscular wasting, all of which
411 have been observed during seasonal extremes in the dry forest population.
412
413 **METHODS**
414
415 *Study populations and sample collection*
416 Central American white-faced capuchins *(Cebus imitator)*, a member of the gracile
417 radiation of capuchins (genus *Cebus*) [91], were recently recognized as a species, distinct from
418 *C. capucinus* in South America [14]. *Cebus imitator* occupies a wide diversity of habitats,
419 spanning lowland rainforests and cloud forests in Panama and throughout southern, eastern,
420 and central Costa Rica, and tropical dry forests in northwestern Costa Rica and Nicaragua. The
421 annual precipitation and elevation of rainforest versus dry forest biomes in their current range
422 vary dramatically, leading to considerable variation in the resident flora and fauna [92,93]. We
423 sampled individual Costa Rican capuchin monkeys from populations inhabiting two distinct
424 habitats. 1) lowland rainforest around Quepos, Puntarenas Province; and 2) tropical dry forest at
425 two sites in Guanacaste Province. In total, we collected samples from 23 capuchins, a list of
426 which is provided in table S3.
427 We sampled capuchins inhabiting a lowland tropical rainforest biome by collaborating
428 with *Kids Saving the Rainforest* (KSTR) in Quepos, Costa Rica. We acquired blood samples
429 from 4 wild capuchins from nearby populations who were undergoing treatment at the facility
430 (although we were unable to collect paired fecal samples). For one of these individuals, an adult
431 male white-faced capuchin that was mortally wounded by a vehicle in Costa Rica, we
432 additionally sampled tissues from several organs. DNA derived from the kidney was used for the
433 reference genome assembly.
434 We collected 21 samples from 19 individuals in the northern tropical dry forest. 16 fecal
435 samples and 4 tissue samples were from free-ranging white-faced capuchin monkeys (*Cebus*

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

436 *imitator*) at in the Sector Santa Rosa (SSR), part of the Área de Conservación Guanacaste in
437 northwestern Costa Rica, which is a 163,000 hectare tropical dry forest nature reserve (Figure
438 1). Behavioral research of free-ranging white-faced capuchins has been ongoing at SSR since
439 the 1980's which allows for the reliable identification of known individuals from facial features
440 and bodily scars [94]. The 16 fresh fecal samples were collected from 14 white-faced capuchin
441 monkeys immediately following defecation (Table S3). We placed 1 mL of feces into conical 15
442 mL tubes pre-filled with 5 mL of RNAlater. RNAlater preserved fecal samples were sent to the
443 University of Calgary, where they were stored at room temperature for up to three years. To
444 evaluate other preservation methods, we also collected two additional capuchin monkey fecal
445 samples (SSR-FL and a section of SSR-ML), which we stored in 1X PBS buffer and then frozen
446 in liquid nitrogen with a betaine cryopreservative [95]. Given the logistical challenges of carrying
447 liquid nitrogen to remote field sites, we prioritized evaluation of samples stored in RNAlater. We
448 also collected tissue and blood samples opportunistically. During the course of our study, 4
449 individual capuchin monkeys died of natural causes at SSR, from whom we were able to collect
450 tissue samples, which were stored in RNAlater. Additionally, we collected a blood sample from 1
451 northern dry-forest individual housed at KSTR that originated near the town of Cañas, which is
452 ~100 km southwest of SSR.
453
454 *Genome-wide sequencing, genome assembly and gene annotation*
455 We assembled a reference genome for *Cebus imitator* from DNA extracted from the
456 kidney of a male Costa Rican individual (KSTR64) using a short read approach (Illumina HiSeq
457 2500). Based on a genome estimate of 3 Gb, the total sequencing depth generated was 81X,
458 including 50X of overlapping read-pairs (200 bp insert), 26X and 5X of 3 and 8 kbs insert read
459 pairs, respectively. The combined sequence reads were filtered and assembled using default
460 parameter settings with ALLPATHS-LG [96]. To improve the quality of gene annotation, we
461 isolated total RNA from the whole blood of an adult male white-faced capuchin (ID: CNS-HE)

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

462 permanently residing at the KSTR wildlife rehabilitation center. The blood was immediately
463 stored in a PAXgene blood RNA tube (Qiagen), and frozen at ultralow temperatures for
464 subseqeunt use. To extract total RNA, we used the PAXgene Blood RNA kit following the
465 manufacturer's recommended protocols. A RiboZero library construction protocol was followed
466 according to the manufacturer's specifications and sequenced on an Illumina HiSeq 2000
467 instrument creating 150 bp paired-end reads. We assembled the FASTQ sequence files into
468 transcripts with Trinity, and submitted the assembled transcriptome to the National Center for
469 Biotechnology Information (NCBI) to assist in gene annotation. The capuchin genome assembly
470 was annotated with the NCBI pipeline previously described here:
471 (http://www.ncbi.nlm.nih.gov/books/NBK169439/).
472
473 *Phylogenetic arrangement and data treatment*
474 The phylogenetic arrangement in this study included 14 species as outgroups to *C.*
475 *imitator*: three Platyrrhini (*Callithrix jacchus*, *Aotus nancymaae*, *Saimiri* boliviensis), six
476 Catarrhini (*Macaca mulatta*, *Rhinopithecus roxellana*, *Nomascus leucogenys*, *Pan troglodytes*,
477 *Homo sapiens*, *Pongo abelii*), one Strepsirrhini (*Microcebus murinus*), one rodent (*Mus*
478 *musculus*), and three Laurasiatheria (*Canis lupus familiaris*, *Bos taurus,* and *Sus scrofa*).
479 Genomic cds were downloaded from Ensembl and NCBI (Table S13). The sequences per
480 genome were clustered using CD-HITest version 4.6 [97] with a sequence identity threshold of
481 90% and an alignment coverage control of 80%. To remove low quality sequences and keep the
482 longest transcript per gene, we used TransDecoder.LongOrfs and TransDecoder.Predict
483 (https://transdecoder.github.io) with default criteria.
484
485 *Orthology identification*
486 The orthology assessment was performed with OMA stand-alone v. 2.3.1 [98]. The OMA
487 algorithm makes strict pairwise “all-against-all" sequence comparisons and identifies the

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

orthologous pairs (genes related by speciation events) based on evolutionary distances. These orthologous genes were clustered into Orthologous Groups (OGs). All OGs included one ortholog sequence from capuchin and at least one outgroup. The tree topology was obtained from TimeTree (http://www.timetree.org/). We identified 7,519 OGs present among the 15 species. Each orthogroup shared by all species was translated into amino acids using the function pxtlate -s in phyx [99]. Amino acid sequences were aligned using the L-INS-i algorithm from MAFFT v.7 [100]. We generated codon alignments using pxaa2cdn in phyx. To avoid false positives in low quality regions, the codon alignments were cleaned with the codon.clean.msa algorithm in rphast [101], using human as a reference sequence. We used conservative methodologies of homology and data cleaning to obtain a smaller number of orthologous genes that avoided false positives with high confidence.

Positive natural selection analysis through codon-based models of evolution and enrichment tests

To evaluate signals consistent with positive selection in the *C. imitator* genome, we explored variation in the ratio of non-synonymous and synonymous substitutions ($d_N/d_S=\omega$) in the ancestor of *Cebus*. We used branch and branch-site substitution models with a maximum likelihood approach in PAML v4.9 [32], which we implemented through the python framework ETE-toolkit with the ete-evol function [102]. We compared the null model where the omega (ω) value in the branch marked as foreground was set with 1, with the model where the ω value was estimated from the data [103]. Likelihood ratio tests (LRT) were used to test for significance between the models and probability values were adjusted with a false discovery rate correction for multiple testing with a q-value < 0.05 for the two positive selection models (branch and branch-site).

We performed functional annotation analysis using DAVID bioinformatics resources - DAVID 6.7 [33], to ascertain which ontology processes the genes with signals of positive

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

514 selection were involved. We focussed the enrichment analysis on two functional categories:
515 Biological Processes (BP) and Genetic Association Database (GAD). GAD is a database from
516 complex diseases and disorders in humans. Finally, the genes with positive section signal were
517 intersected with the GenAge database (build 19,307 genes) [42].
518
519 *FACS*
520 Before isolating cells by fluorescence-activated cell sorting, fecal samples were prepared
521 using a series of washes and filtration steps. Fecal samples were vortexed for 30 s and
522 centrifuged for 30 s at 2,500 g. Then the supernatant was passed through a 70 µm filter into a
523 50 mL tube and washed with DPBS. After transferring the resultant filtrate to a 15 mL tube, it
524 was centrifuged at 1,500 RPM for 5 minutes to pellet the cells. Then we twice washed the cells
525 with 13 mL of DPBS. We added 500 µL of DPBS to the pellet and re-filtered through a 35 µm
526 filter into a 5 mL FACS tube. We prepared a negative control (to control for auto-fluorescence)
527 with 500 µL of DPBS and one drop of the cell solution. To the remaining solution, we added 1
528 µL of AE1/AE3 Anti-Pan Cytokeratin Alexa Fluor® 488 antibody or TOTO-3 DNA stain, which
529 we allowed to incubate at 4°C for at least 30 minutes.
530 We isolated cells using a BD FACSAria™ Fusion (BD Biosciences) flow cytometer at the
531 University of Calgary Flow Cytometry Core. To sterilize the cytometer's fluidics before
532 processing each sample, we ran a 3% bleach solution through the system for four minutes at
533 maximum pressure. We assessed background fluorescence and cellular integrity by processing
534 the negative control sample prior to all prepared fecal samples. For each sample we first gated
535 our target population by forward and side scatter characteristics that were likely to minimize
536 bacteria and cellular debris (Figure S8). Secondary and tertiary gates were implemented to
537 remove cellular agglomerations. Finally, we selected cells with antibody or DNA fluorescence
538 greater than background levels. In cases when staining was not effective, we sorted solely on
539 the first three gates. Cells were pelleted and frozen at -20°C.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

DNA Extraction and Shotgun Sequencing

We extracted fecal DNA (fDNA) with the QIAGEN DNA Micro kit, following the "small volumes of blood" protocol. To improve DNA yield, we increased the lysis time to three hours, and incubated 50 µL of 56°C elution buffer on the spin column membrane for 10 minutes. DNA concentration was measured with a Qubit fluorometer. Additionally, to calculate endogenous DNA enrichment, we extracted DNA directly from five fecal samples prior to their having undergone FACS. We extracted DNA from the nine tissue and blood samples using the QIAGEN Gentra Puregene Tissue kit and DNeasy blood and tissue kit, respectively.

For the fecal samples, DNA was fragmented to 350 bp with a Covaris sonicator. We built whole genome sequencing libraries with the NEB Next Ultra 2 kit using 10-11 PCR cycles. Fecal genomic libraries were sequenced on an Illumina NextSeq (2x150 PE) at the University of Calgary genome sequencing core and an Illumina HighSeq 4000 at the McDonnell Genome Institute at Washington University in St. Louis (MGI). Using ½ of one HiSeq 4000 lane, we achieved an average coverage of 12.2X across the *Cebus imitator* 1.0 genome (sample SSR-ML). Other fecal samples were sequenced to average depths of 0.1-4.4X (Table S4). High-coverage (10.3-47.6X), whole genome shotgun libraries were prepared for the blood and tissue DNA samples and sequenced on an Illumina X Ten system at MGI. For population analyses within capuchins, we mapped genomic data from all 23 individuals sequenced (Table S3) to the reference genome.

Mapping and SNV Generation

Reads were trimmed of sequencing adaptors with Trimmomatic [104]. Subsequently, we mapped the *Cebus* reads to the *Cebus imitator* 1.0 reference genome (GCF_001604975.1) with BWA mem [105] and removed duplicates with Picard Tools (http://broadinstitute.github.io/picard/) and SAMtools [106]. We called SNVs for each sample

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

566 independently using the *Cebus* genome and the GATK UnifiedGenotyper pipeline (*-out_mode*
567 *EMIT_ALL_SITES*) [107]. Genomic VCFs were then combined using GATK's CombineVariants
568 restricting to positions with a depth of coverage between 3 and 100, mapping quality above 30,
569 no reads with mapping quality zero, and variant PHRED scores above 30. Sequencing reads
570 from one of the high coverage fecal samples (SSR-FL) bore a strong signature of human
571 contamination (16%), and were thus excluded from SNV generation. We included reads from
572 nine tissue/blood samples and one frozen fecal sample with high coverage (SSR-ML). In total,
573 we identified 4,184,363 SNVs for downstream analyses.

574 To remove potential human contamination from sequenced libraries, we mapped
575 trimmed reads to the *Cebus imitator* 1.0 and human (hg38) genomes simultaneously with
576 BBsplit [108]. Using default BBsplit parameters, we binned separately reads that mapped
577 unambiguously to either genome. Ambiguously mapping reads (i.e. those mapping equally well
578 to both genomes) were assigned to both genomic bins, and unmapped reads were assigned to
579 a third bin. We calculated the amount of human genomic contamination as the percentage of
580 total reads unambiguously mapping to the human genome (Table S4). After removing
581 contaminant reads, all libraries with at least 0.5X genomic coverage were used for population
582 structure analysis.

583 In order to test the effect of fecalFACS on mapping rates, we selected five samples at
584 random (SSR-CH, SSR-NM, SSR-LE, SSR-PR, SSR-SN) to compare pre- and post-FACS
585 mapping rates. To test for an increase in mapping percentage, we ran a one-sample paired
586 Wilcoxon signed-rank test on the percentages of reads that mapped exclusively to the *Cebus*
587 genome before and after FACS. Additionally, we ran Pearson's product moment correlations to
588 test for an effect of the number of cells (log10 transformed) on rates of mapping, read
589 duplication, and nanograms of input DNA. The above tests were all performed in R.
590
591 *High coverage fecal genome comparison*

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

We made several comparisons between our high-coverage feces-derived genome and the blood/tissue-derived genomes using window-based approaches. For each test, the feces-derived genome should fall within the range of variation for members of its population of origin (SSR). Deviations from this, for example all fecal genomes clustering together, would indicate biases in our DNA isolation methods. To assess this, we constructed 10 KB windows with a 4KB slide along the largest scaffold (21,314,911 bp) in the *C. imitator* reference genome. From these windows, we constructed plots of coverage density and the distribution of window coverage along the scaffold. Secondly, we assessed the level of heterozygosity in 1 MB / 200 KB sliding windows throughout the ten largest scaffolds. For each high-coverage genome, we plotted the density distribution of window heterozygosity. We measured genome-wide GC content with the Picard Tools CollectGcBiasMetrics function. The percentage of GC content was assessed against the distribution of normalized coverage and the number of reads in 100 bp windows per the number reads aligned to the windows.

Population Structure

Given the large degree of difference in coverage among our samples, (less than 1X to greater than 50X), we performed pseudodiploid allele calling on all samples. For each library, at each position in the SNV set, we selected a single, random read from the sequenced library. From that read, we called the variant information at the respective SNV site for the given library. In so doing, we generated a VCF with a representative degree of variation and error for all samples.

To assess population structure and infer splits between northern and southern groups of Costa Rican white-faced capuchins, we constructed principal components plots with EIGENSTRAT [109] and built population trees with TreeMix [110]. Because we ascertained variants predominantly with libraries that were of tissue/blood origin, we built principal components solely with SNVs from these libraries and projected the remaining fecal libraries

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

618 onto the principal components. For our maximum likelihood trees, we used two outgroups
619 (*Saimiri sciureus*, and *Cebus albifrons*), with *S. sciureus* serving as the root of the tree. Given
620 the geographic distance and anthropogenic deforestation between northern and southern
621 populations, we assumed no migration. To account for linkage disequilibrium, we grouped SNVs
622 into windows of 1,000 SNVs. Population size history was inferred from the highest coverage
623 non-reference individual, SSR-RM08 with PSMC [111], using default parameters
624 (https://github.com/lh3/psmc).
625
626 *Local adaptation, F_{ST}, heterozygosity, relatedness*
627 For all analyses of local adaptation and heterozygosity between populations, we
628 excluded individuals from our low coverage dataset. We tested for the degree of relatedness
629 among all high and low coverage individuals using READ [112] and identified two of the high-
630 coverage individuals from SSR, SSR-ML and SSR-CR, as potential first degree relatives (Figure
631 S9, Table S14). For all statistical analyses of high-coverage samples, we removed SSR-ML,
632 because SSR-CR was sequenced to higher average depth.
633 For each individual, we calculated heterozygosity in 1 Mb / 200 Kb sliding windows
634 across the genome for all scaffolds at least 1 Mb in length. Windows were generated with
635 BedTools windowMaker [113] and heterozygosity was calculated as the per-site average within
636 each window. Based upon a visual inspection of the average heterozygosity values across the
637 genome (Figure 5), we classified a window as part of a run of homozygosity if the window's
638 average heterozygosity fell below 0.0002. Descriptive statistics and two-sided Wilcoxon tests
639 were calculated in R.
640 For each high-coverage sample, we calculated the Hudson's F_{ST} ratio of averages [114]
641 in 20 kb windows with a slide of 4 Kb across the genome. Among the genes present in each
642 window in the top 0.5% and top 0.1% of F_{ST} values, we searched for SNPs with high or
643 moderate effects using SnpEff and identified those SNPs with high F_{ST} values (> 0.75) using

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

644 VCFtools. We searched for functional enrichment of our population gene set using ToppFun in
645 the ToppGene Suite [115]. In an effort to identify candidate genes for further investigation, we
646 set low thresholds for significance (FDR p-value < 0.1 and the minimum number of genes per
647 category to 1).
648
649 Chemosensory genes
650 The chemosensory behaviors of capuchins have been well-studied [31], and taste and
651 olfaction are suspected to play an important role in their foraging ecology. 614 orthologous olfactory
652 receptor genes were identified in the *Cebus* reference genome using previously described
653 methods [116]. Briefly, putative OR sequences were identified by conducting TBLASTN
654 searches against the capuchin reference assembly using functional human OR protein
655 sequences as queries with an e-value threshold of 1e-20. For each reference-derived OR gene,
656 we added 500 bp of flanking sequence to both the 5' and 3' ends in a bed file. For each
657 individual, we extracted the OR gene region from the gVCF and generated a consensus
658 sequence defaulting to the reference allele at variable site using bcftools [117]. The number of
659 intact, truncated, and pseudogenized OR genes in each individual were identified using the
660 ORA pipeline [118]. We considered an OR gene to be putatively functional if its amino acid
661 sequence was at least 300 amino acids in length. ORA was further used to classify each OR
662 into the appropriate class and subfamily with an e-value cutoff of 1e-10 to identify the functional
663 OR subgenome for each individual [118,119]. Taste and vomeronasal receptor genes were
664 identified in the NCBI genome annotation, and variable positions were located by scanning the
665 VCF with VCFtools [120] and bash. The positions of *Cebus* opsin tuning sites have been
666 identified previously [68]. With the high-coverage dataset, we identified the allele(s) present at
667 each locus in the VCF. For each low-coverage fecal-derived genome, we located the position of
668 the tuning site in the bam file using SAMtools [117] tview and manually called the variant when
669 possible.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

DECLARATIONS

Ethics approval and consent to participate

This research adhered to the laws of Costa Rica, the United States, and Canada and complied with protocols approved by the Área de Conservación Guanacaste and by the Canada Research Council for Animal Care through the University of Calgary's Life and Environmental Care Committee (ACC protocol AC15-0161). Samples were collected with permission from the Área de Conservación Guanacaste (ACG-PI-033-2016) and CONAGEBIO (R-025-2014-OT-CONAGEBIO; R-002-2020-OT-CONAGEBIO). Samples were exported from Costa Rica under permits from CITES and Area de Conservacion Guanacaste (2016-CR2392/SJ #S 2477, 2016-CR2393/SJ #S 2477, DGVS-030-2016-ACG-PI-002-2016; 012706) and imported with permission from the Canadian Food and Inspection Agency (A-2016-03992-4).

Consent for Publication

Not applicable

Availability of data and materials

The reference genome is available at NCBI through BioProjects PRJNA298580 and PRJNA328123. RNAseq reads used in genome annotation can be accessed through PRJNA319062. The sequencing reads used in the local adaptation will be released by NCBI upon publication (PRJNA610850), and are available immediately to reviewers and editors upon request from the corresponding author.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

696

697 **Competing Interests**

698

699 The authors declare that they have no competing interests.

700

701 **Funding**

702

703 Funding was provided by Washington University in St. Louis, the Canada Research
704 Chairs Program, and a National Sciences and Engineering Council of Canada Discovery grant
705 (to A.D.M.), the Alberta Children's Hospital Research Institute (to A.D.M. and J.D.O), the Beatriu
706 de Pinós postdoctoral programme of the Government of Catalonia's Secretariat for Universities
707 and Research of the Ministry of Economy and Knowledge (to J.D.O), and the Japan Society for
708 the Promotion of Science 15H02421 and 18H04005 (S.K.). This work was partly funded by a
709 Methuselah Foundation grant to J.P.M. To the Comisión Nacional de Investigación Científica y
710 Tecnológica (CONICYT) - Chile through the doctoral studentship number N°21170433 and the
711 scholarship from MECESUP AUS 2003 to D.T.M. GenAge is funded by a Biotechnology and
712 Biological Sciences Research Council (BB/R014949/1) grant to J.P.M. TMB is supported by
713 BFU2017-86471-P (MINECO/FEDER, UE), Howard Hughes International Early Career, Obra
714 Social "La Caixa" and Secretaria d'Universitats i Recerca and CERCA Programme del
715 Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880).
716 C.F is supported by "La Caixa" doctoral fellowship program. EL is supported by CGL2017-
717 82654-P (MINECO/FEDER, UE). The funding bodies played no role in the design of the study,
718 in the collection, analysis, and interpretation of data, or in writing the manuscript.

719

720

721

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Author contributions

ADM and JDO conceived of the project

JDO, ADM, YN, WCW, RK, JdC, JPM, and TMB contributed to the study design

JDO, DTM, MdM, MJM, LFK, and MCJ performed computational analyses

JDO, RK, and JT conducted flow cytometry

JDO, ADM, CF, JH, and EL conducted molecular lab work

ADM, GHP, ADF, JAH, and SK contributed samples

ADM, RK, and TMB provided computational and laboratory resources

JDO, ADM, MJM, JPM, and DTM wrote the manuscript with commentary from all authors

Acknowledgements

We thank R. Blanco Segura and M. M. Chavarria and staff from the Área de Conservación Guanacaste and Ministerio de Ambiente y Energía. From Kids Saving the Rainforest we thank the volunteers and veterinarians, especially Pia Martin, Carmen Soto, Jennfier Rice and Chip Braman. For assistance in the lab, we acknowledge Kelly Kries, Gwen Duytschaever, Jene Weatherhead, Laurie Kennedy, and Yiping Liu. We thank Aoife Doherty and Mengjia Li for exploratory analyses of selection in aging-related genes and acknowledge Patrick Minx and Kim Kyung for assistance with bioinformatics. We also thank Jessica Lynch and Hazel Byrne for comparative data. Thanks to R. Gregory in the Centre for Genomic Research and I.C. Smith in the Advanced Research Computing at the University of Liverpool for the access to the computing resources. To Dr. J.C Opazo for the insightful discussion.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

FIGURES

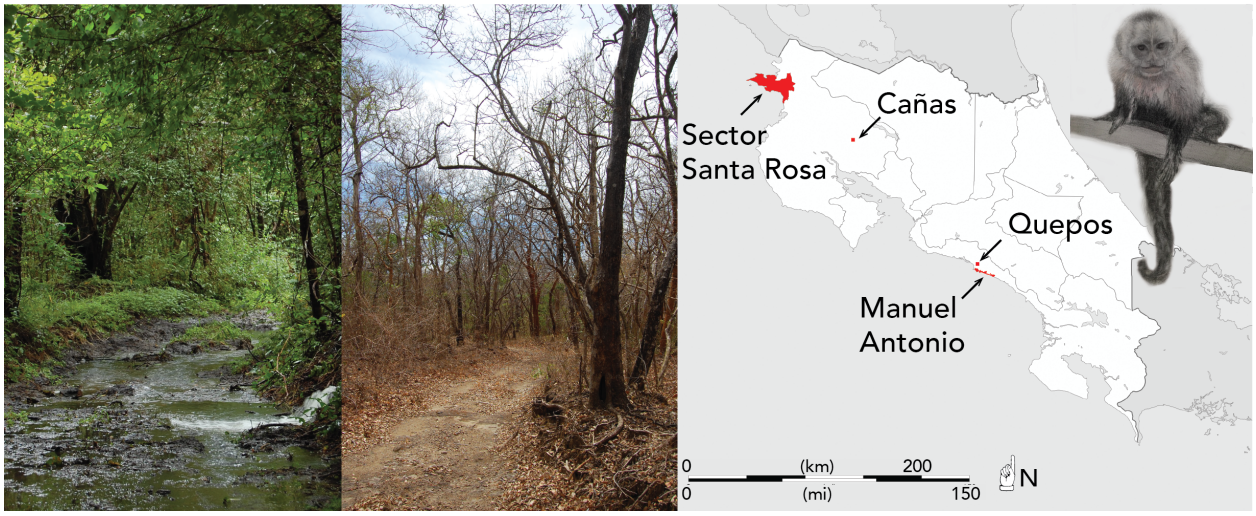

Figure 1: Sector Santa Rosa (SSR) during wet (left) and dry (middle) seasons. Right: Map of sampling locations in Costa Rica. The two northern sites, SSR and Cañas, have tropical dry forest biomes, whereas the two southern sites, Quepos and Manuel Antonio are tropical wet forests. Photos - Amanda Melin; Drawing of white faced capuchin monkey - Alejandra Tejada-Martinez; Map: Eric Gaba – Wikimedia Commons user: Sting

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

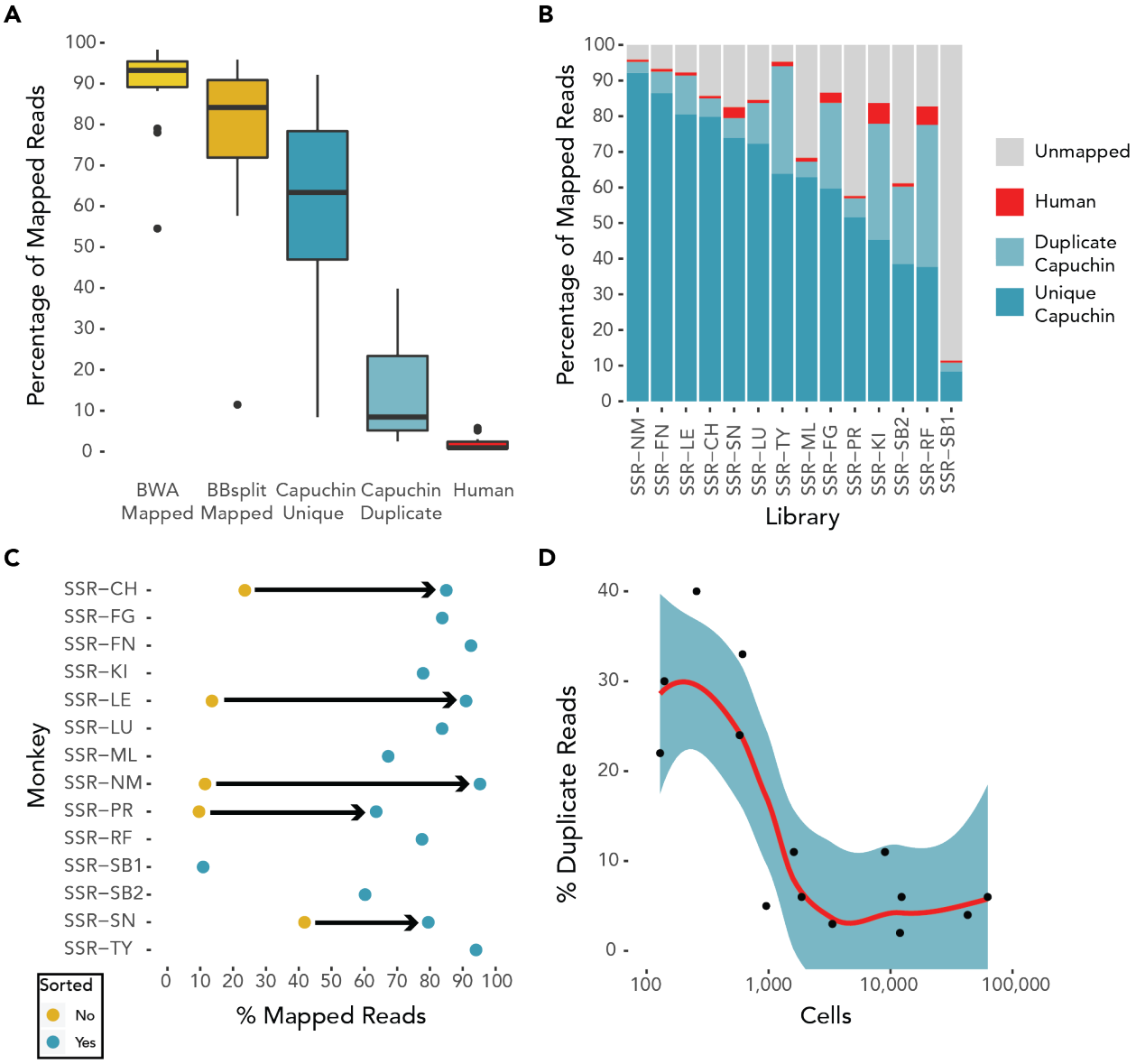

Figure 2: Mapping percentages of sequencing reads from RNAlater preserved fDNA libraries prepared with FACS for A) all samples [Box-plot elements: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers], and B) individual libraries. C) Increase in mapping rate for RNAlater preserved samples. D) Relationship between mapped read duplication and number of cells with LOESS smoothing. The duplicate rate decreases sharply once a threshold of about 1,000 cells is reached.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

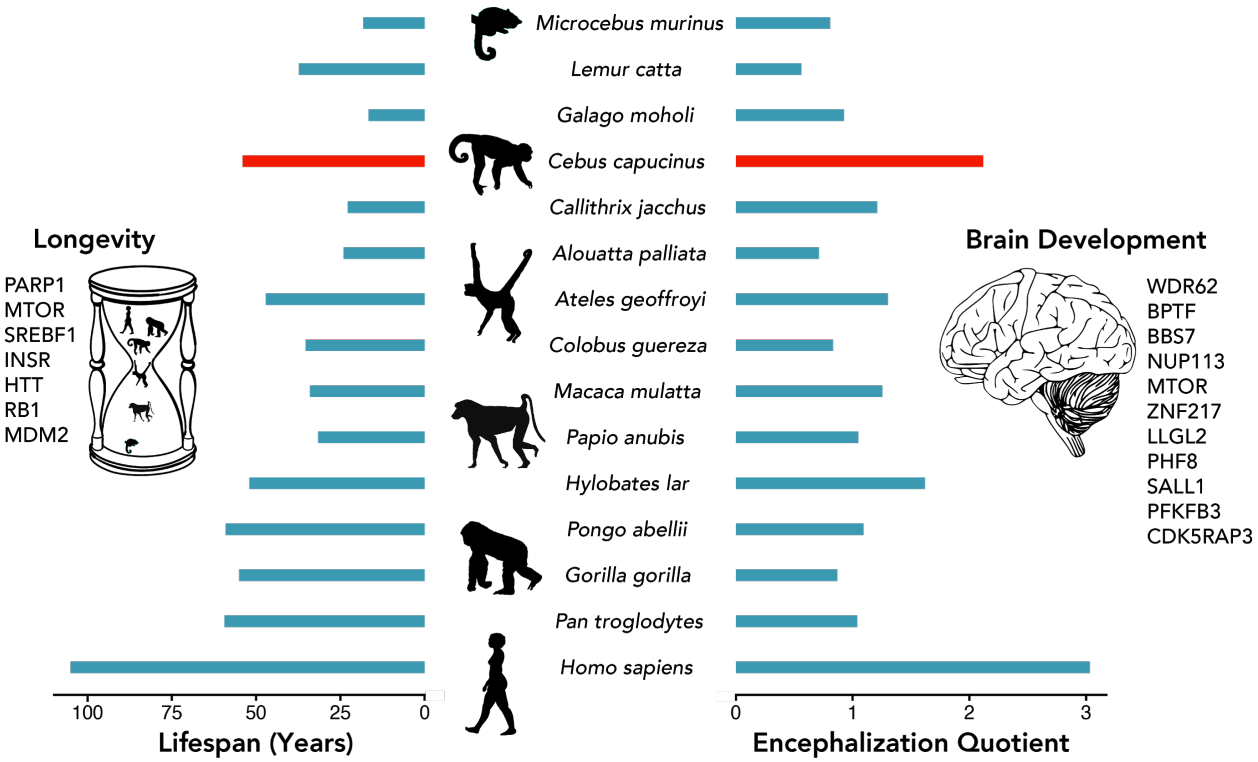

Figure 3: Genes under positive selection in white-faced capuchin monkeys that are associated with longevity and brain development. Values from *Cebus capucinus* are used in place of *Cebus imitator*, given the recent taxonomic split.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

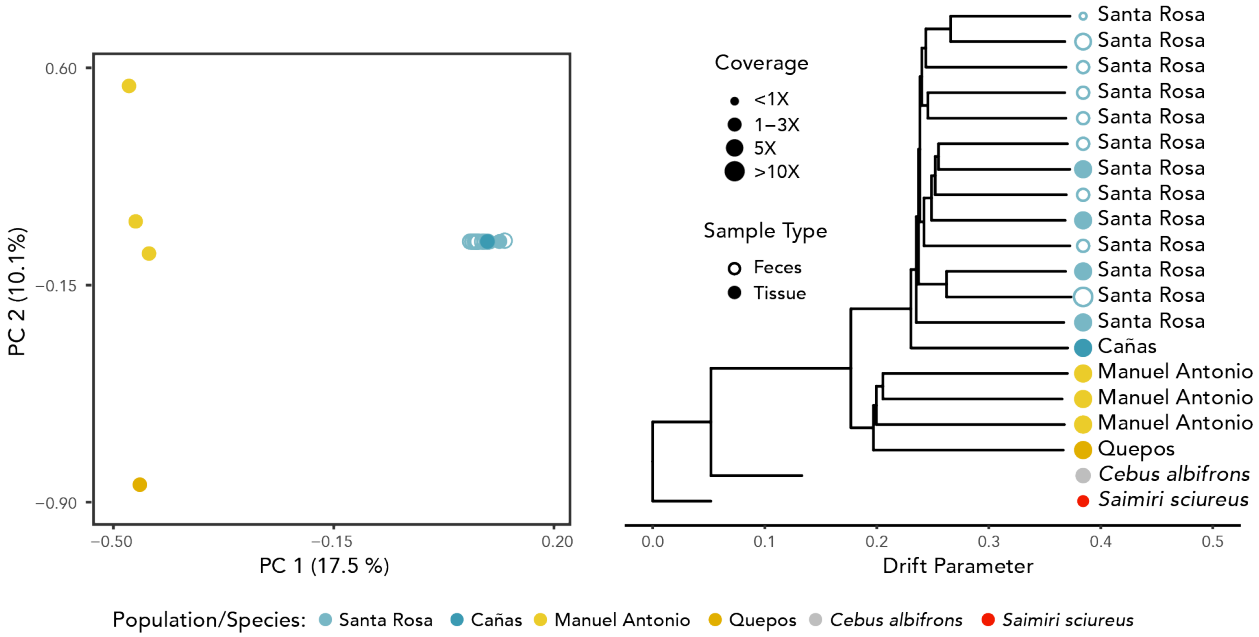

Figure 4: Population subdivision in *Cebus imitator*. Left: Principal components of 13 fecal and 10 blood/tissue libraries from white faced capuchins. Individuals from northern and southern sites separate on PC 1. Low- and high-coverage *C. imitator* samples from Santa Rosa plot in the same cluster. Right: Maximum likelihood tree of 9 fecal and 10 blood/tissue libraries from *C. imitator* (samples with less than 0.5X coverage were excluded). Among the white-faced capuchin samples, individuals from northern (dry forest) and southern (wet forest) regions form the primary split; secondary splits reflect the individuals from different sites within regions. The short branch lengths of the outgroups are a result of only polymorphic positions within *C. imitator* being used to construct the tree.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

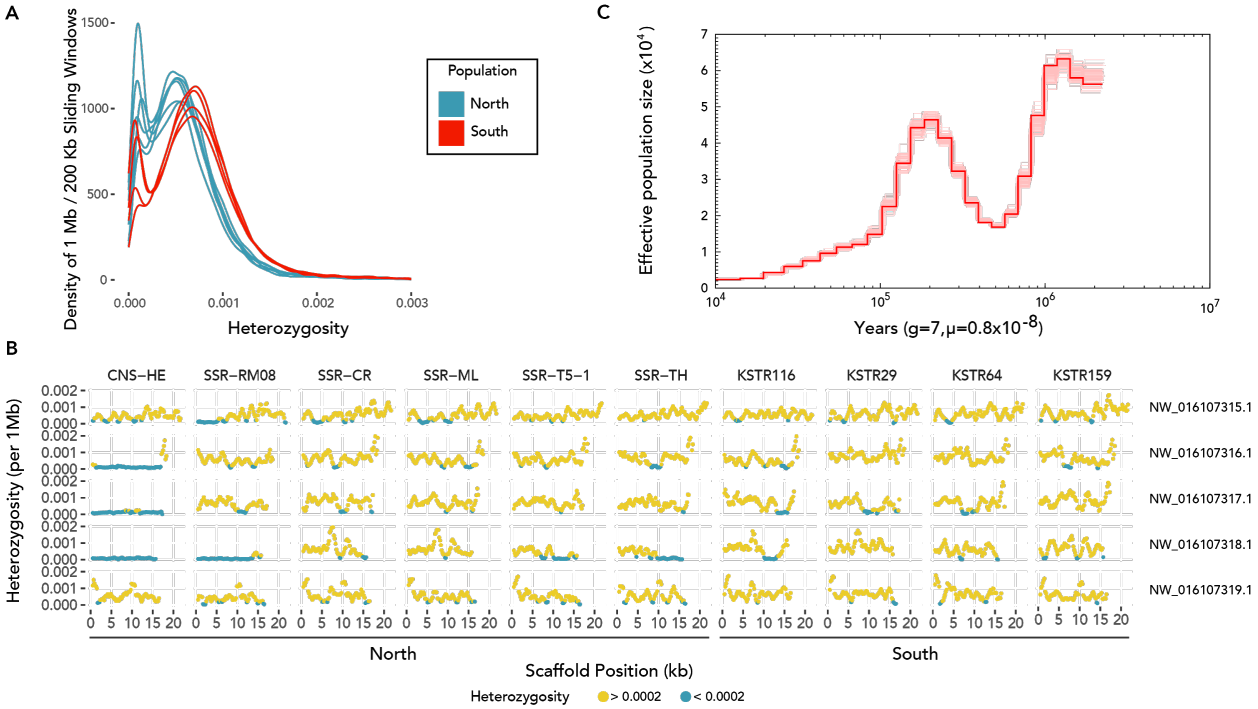

Figure 5: A: Density plot of 1 Mb windows with a slide of 200 Kb in northern and southern populations. The distribution of windows from the northern population indicates lower heterozygosity than the southern distribution.The individuals from the southern population show consistently higher values. B: Long runs of homozygosity in the 5 largest scaffolds. Blue dots represent windows with depleted heterozygosity. The individuals with the longest runs of homozygosity come from the northern population. C: PSMC plot of effective population size over time.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

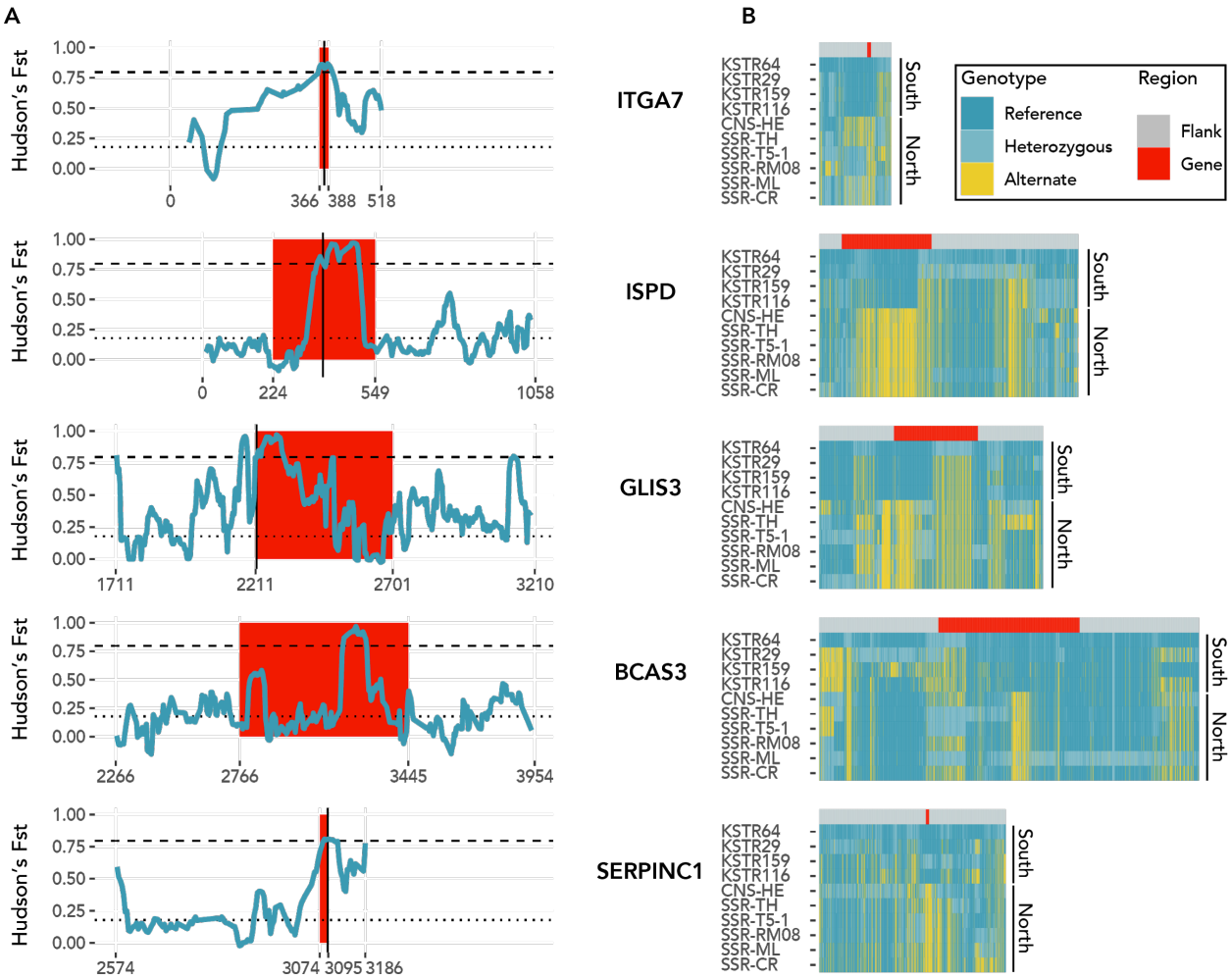

Figure 6: Highly differentiated genes between wet and dry forest populations involved in diabetes, kidney function, and creatinine levels. A: Hudson's F_{ST} within windows of 20kb with a 4 kb slide. Gene regions are in red, flanked by 500 kb (or length to beginning or end of scaffold) of sequence. X-axis values correspond to position along the scaffold. The dotted line indicates average F_{ST} value across all windows (F_{ST} = 0.178), and the dashed line represents the top 0.5% of values (F_{ST} = 0.797). Vertical black lines indicate a non-synonymous SNP with an Fst >= 0.750, excluding BCAS3 (see Results). B: Heatmaps indicating the pattern of SNP variation within and surrounding highly divergent genes. SNVs within the genes are located under the red band and those within 200 kb of flanking region under the gray bands.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

REFERENCES

1. DeCasien AR, Williams SA, Higham JP. Primate brain size is predicted by diet but not sociality. Nature Ecology &Amp; Evolution. Macmillan Publishers Limited, part of Springer Nature.; 2017;1:0112.

2. Street SE, Navarrete AF, Reader SM, Laland KN. Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. Proc Natl Acad Sci U S A. 2017;114:7908–14.

3. Aiello LC, Wheeler PE. The expensive-tissue hypothesis. Curr Anthropol. 1995;36:199–221.

4. Washburn SL. Tools and human evolution. Sci Am. 1960;203:63–75.

5. Kaplan H, Hill K, Lancaster J, Magdalena Hurtado A. A theory of human life history evolution: Diet, intelligence, and longevity [Internet]. Evolutionary Anthropology: Issues, News, and Reviews. 2000. p. 156–85.

6. Russon AE, Begun DR. Evolutionary origins of great ape intelligence: an integrated view [Internet]. The Evolution of Thought. 2004. p. 353–68. Available from: http://dx.doi.org/10.1017/cbo9780511542299.023

7. Melin AD, Young HC, Mosdossy KN, Fedigan LM. Seasonality, extractive foraging and the evolution of primate sensorimotor intelligence. J Hum Evol. 2014;71:77–86.

8. Fragaszy DM, Visalberghi E, Fedigan LM. The Complete Capuchin: The Biology of the Genus Cebus. Cambridge University Press; 2004.

9. De Petrillo F, Rosati AG. Ecological rationality: Convergent decision-making in apes and capuchins. Behav Processes. 2019;164:201–13.

10. Barrett BJ, Monteza-Moreno CM, Dogandžić T, Zwyns N, Ibáñez A, Crofoot MC. Habitual stone-tool-aided extractive foraging in white-faced capuchins, Cebus capucinus. R Soc open sci. 2018;5:181002.

11. Emery NJ. Are Corvids “Feathered Apes”? Cognitive evolution in crows, jays, rooks and jackdaws. In: Watanabe S, editor. Comparative Analysis of Minds. Keio University Press: Tokyo.; 2004. p. 181–213.

12. Marino L. Convergence of complex cognitive abilities in cetaceans and primates. Brain Behav Evol. 2002;59:21–32.

13. Lynch Alfaro JW, Boubli JP, Olson LE, Di Fiore A, Wilson B, Gutiérrez-Espeleta GA, et al. Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys: Biogeography of Neotropical capuchin monkeys. J Biogeogr. 2012;39:272–88.

14. Boubli JP, Rylands AB, Farias IP, Alfaro ME, Alfaro JL. Cebus phylogenetic relationships: a preliminary reassessment of the diversity of the untufted capuchin monkeys. Am J Primatol. 2012;74:381–93.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

847 15. Johnson SE, Brown KA. The Specialist Capuchin? Using Ecological Niche Models to
848 Compare Niche Breadth in Mesoamerican Primates. In: Kalbitzer U, Jack KM, editors. Primate
849 Life Histories, Sex Roles, and Adaptability: Essays in Honour of Linda M Fedigan. Cham:
850 Springer International Publishing; 2018. p. 311–29.

851 16. Kuehn M, Welsch H, Zahnert T, Hummel T. Changes of pressure and humidity affect
852 olfactory function. Eur Arch Otorhinolaryngol. 2008;265:299–302.

853 17. Benignus VA, Prah JD. Flow thresholds of nonodorous air through the human naris as a
854 function of temperature and humidity. Percept Psychophys. 1980;27:569–73.

855 18. Mollon JD. “Tho’she kneel'd in that place where they grew…” The uses and origins of
856 primate colour vision. J Exp Biol [Internet]. jeb.biologists.org; 1989; Available from:
857 https://jeb.biologists.org/content/146/1/21.short

858 19. Campos FA, Fedigan LM. Behavioral adaptations to heat stress and water scarcity in white-
859 faced capuchins (Cebus capucinus) in Santa Rosa National Park, Costa Rica. Am J Phys
860 Anthropol. Wiley Online Library; 2009;138:101–11.

861 20. Mosdossy KN, Melin AD, Fedigan LM. Quantifying seasonal fallback on invertebrates, pith,
862 and bromeliad leaves by white-faced capuchin monkeys (Cebus capucinus) in a tropical dry
863 forest. Am J Phys Anthropol. 2015;158:67–77.

864 21. Proffitt T, Luncz LV, Falótico T, Ottoni EB, de la Torre I, Haslam M. Wild monkeys flake
865 stone tools. Nature. 2016;539:85–8.

866 22. Izar P, Verderane MP, Peternelli-Dos-Santos L, Mendonça-Furtado O, Presotto A, Tokuda
867 M, et al. Flexible and conservative features of social systems in tufted capuchin monkeys:
868 comparing the socioecology of Sapajus libidinosus and Sapajus nigritus. Am J Primatol.
869 2012;74:315–31.

870 23. Merkt JR, Taylor CR. “Metabolic switch” for desert survival [Internet]. Proceedings of the
871 National Academy of Sciences. 1994. p. 12313–6. Available from:
872 http://dx.doi.org/10.1073/pnas.91.25.12313

873 24. Pontzer H, Raichlen DA, Shumaker RW, Ocobock C, Wich SA. Metabolic adaptation for low
874 energy throughput in orangutans. Proc Natl Acad Sci U S A. 2010;107:14048–52.

875 25. Schmidt-Nielsen K, Schmidt-Nielsen B. Water metabolism of desert mammals 1. Physiol
876 Rev. 1952;32:135–66.

877 26. Cain JW III, Krausman PR, Rosenstock SS, Turner JC. Mechanisms of Thermoregulation
878 and Water Balance in Desert Ungulates. Wildl Soc Bull. Wiley Online Library; 2006;34:570–81.

879 27. Olender T, Waszak SM, Viavant M, Khen M, Ben-Asher E, Reyes A, et al. Personal receptor
880 repertoires: olfaction as a model. BMC Genomics. 2012;13:414.

881 28. Jacobs RL, MacFie TS, Spriggs AN, Baden AL, Morelli TL, Irwin MT, et al. Novel opsin gene
882 variation in large-bodied, diurnal lemurs. Biol Lett [Internet]. 2017;13. Available from:
883 http://dx.doi.org/10.1098/rsbl.2017.0050

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

884 29. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo J-M, et al. Evolutionary
885 history and adaptation from high-coverage whole-genome sequences of diverse African hunter-
886 gatherers. Cell. 2012;150:457–69.

887 30. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. WindowMasker: window-based masker for
888 sequenced genomes. Bioinformatics. 2006;22:134–41.

889 31. Rice ES, Green RE. New Approaches for Genome Assembly and Scaffolding. Annu Rev
890 Anim Biosci. 2019;7:17–40.

891 32. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol.
892 2007;24:1586–91.

893 33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
894 lists using DAVID bioinformatics resources [Internet]. Nature Protocols. 2009. p. 44–57.
895 Available from: http://dx.doi.org/10.1038/nprot.2008.211

896 34. Memon MM, Raza SI, Basit S, Kousar R, Ahmad W, Ansar M. A novel WDR62 mutation
897 causes primary microcephaly in a Pakistani family. Mol Biol Rep. 2013;40:591–5.

898 35. Stankiewicz P, Khan TN, Szafranski P, Slattery L, Streff H, Vetrini F, et al.
899 Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and
900 Speech Delay, Postnatal Microcephaly, and Dysmorphic Features. Am J Hum Genet.
901 2017;101:503–15.

902 36. Guo J, Higginbotham H, Li J, Nichols J, Hirt J, Ghukasyan V, et al. Developmental
903 disruptions underlying brain abnormalities in ciliopathies. Nat Commun. 2015;6:7857.

904 37. Fujita A, Tsukaguchi H, Koshimizu E, Nakazato H, Itoh K, Kuraoka S, et al. Homozygous
905 splicing mutation in NUP13 3 causes Galloway-Mowat syndrome [Internet]. Annals of
906 Neurology. 2018. p. 814–28. Available from: http://dx.doi.org/10.1002/ana.25370

907 38. Crino PB. mTOR signaling in epilepsy: insights from malformations of cortical development.
908 Cold Spring Harb Perspect Med [Internet]. 2015;5. Available from:
909 http://dx.doi.org/10.1101/cshperspect.a022442

910 39. Bertipaglia C, Gonçalves JC, Vallee RB. Nuclear migration in mammalian brain
911 development. Semin Cell Dev Biol. 2018;82:57–66.

912 40. Laumonnier F, Holbert S, Ronce N, Faravelli F, Lenzner S, Schwartz CE, et al. Mutations in
913 PHF8 are associated with X linked mental retardation and cleft lip/cleft palate. J Med Genet.
914 2005;42:780–6.

915 41. Li Y, de Magalhães JP. Accelerated protein evolution analysis reveals genes and pathways
916 associated with the evolution of mammalian longevity. Age . 2013;35:301–14.

917 42. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. Human Ageing
918 Genomic Resources: new and updated databases. Nucleic Acids Res. 2018;46:D1083–90.

919 43. Grube K, Bürkle A. Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13
920 mammalian species correlates with species-specific life span. Proc Natl Acad Sci U S A.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

921 1992;89:11759–63.

922 44. Cornu M, Albert V, Hall MN. mTOR in aging, metabolism, and cancer. Curr Opin Genet Dev.
923 2013;23:53–62.

924 45. de Magalhães JP, Stevens M, Thornton D. The Business of Anti-Aging Science. Trends
925 Biotechnol. 2017;35:1062–73.

926 46. Fujii N, Narita T, Okita N, Kobayashi M, Furuta Y, Chujo Y, et al. Sterol regulatory element-
927 binding protein-1c orchestrates metabolic remodeling of white adipose tissue by caloric
928 restriction. Aging Cell. Wiley Online Library; 2017;16:508–17.

929 47. Plank M, Wuttke D, van Dam S, Clarke SA, de Magalhães JP. A meta-analysis of caloric
930 restriction gene expression profiles to infer common signatures and regulatory mechanisms. Mol
931 Biosyst. 2012;8:1339–49.

932 48. Kenyon CJ. The genetics of ageing. Nature. 2010;464:504–12.

933 49. Zheng S, Clabough EBD, Sarkar S, Futter M, Rubinsztein DC, Zeitlin SO. Deletion of the
934 huntingtin polyglutamine stretch enhances neuronal autophagy and longevity in mice. PLoS
935 Genet. 2010;6:e1000838.

936 50. Long H, Huang K. Transport of Ciliary Membrane Proteins. Front Cell Dev Biol. 2020;7:381.

937 51. Lu Z, Wang F, Liang M. SerpinC1/Antithrombin III in kidney-related diseases. Clin Sci .
938 2017;131:823–31.

939 52. Kauffmann RH, Veltkamp JJ, Van Tilburg NH, Van Es LA. Acquired antithrombin III
940 deficiency and thrombosis in the nephrotic syndrome. Am J Med. 1978;65:607–13.

941 53. Lau SO, Tkachuck JY, Hasegawa DK, Edson JR. Plasminogen and antithrombin III
942 deficiencies in the childhood nephrotic syndrome associated with plasminogenuria and
943 antithrombinuria. J Pediatr. 1980;96:390–2.

944 54. Citak A, Emre S, Sirin A, Bilge I, Nayır A. Hemostatic problems and thromboembolic
945 complications in nephrotic children. Pediatr Nephrol. 2000;14:138–42.

946 55. Liang M, Lee NH, Wang H, Greene AS, Kwitek AE, Kaldunski ML, et al. Molecular networks
947 in Dahl salt-sensitive hypertension based on transcriptome analysis of a panel of consomic rats.
948 Physiol Genomics. 2008;34:54–64.

949 56. Tomczykowska M, Bielak J, Bodys A. Evaluation of platelet activation, plasma antithrombin
950 III and alpha2-antiplasmin activities in hypertensive patients. Ann Univ Mariae Curie
951 Sklodowska Med. 2003;58:15—20.

952 57. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, et al. Trans-ethnic kidney
953 function association study reveals putative causal genes and effects on kidney-specific disease
954 aetiologies. Nat Commun. 2019;10:29.

955 58. Okada Y, Sim X, Go MJ, Wu J-Y, Gu D, Takeuchi F, et al. Meta-analysis identifies multiple
956 loci associated with kidney function-related traits in east Asian populations. Nat Genet.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

957 2012;44:904–9.

958 59. Bergstrom ML, Emery Thompson M, Melin AD, Fedigan LM. Using urinary parameters to
959 estimate seasonal variation in the physical condition of female white-faced capuchin monkeys
960 (Cebus capucinus imitator). Am J Phys Anthropol [Internet]. 2017; Available from:
961 http://dx.doi.org/10.1002/ajpa.23239

962 60. Myers VC, Fine MS. The creatine content of muscle under normal conditions. Its relation to
963 the urinary creatinine. Proc Soc Exp Biol Med. SAGE Publications; 1912;10:10–1.

964 61. Cirak S, Foley AR, Herrmann R, Willer T, Yau S, Stevens E, et al. ISPD gene mutations are
965 a common cause of congenital and limb-girdle muscular dystrophies. Brain. 2013;136:269–81.

966 62. Roscioli T, Kamsteeg E-J, Buysse K, Maystadt I, van Reeuwijk J, van den Elzen C, et al.
967 Mutations in ISPD cause Walker-Warburg syndrome and defective glycosylation of α-
968 dystroglycan. Nat Genet. Nature Publishing Group; 2012;44:581–5.

969 63. Mayer U, Saher G, Fässler R, Bornemann A, Echtermeyer F, von der Mark H, et al.
970 Absence of integrin alpha 7 causes a novel form of muscular dystrophy. Nat Genet.
971 1997;17:318–23.

972 64. Zhang Q, Bethmann C, Worth NF, Davies JD, Wasner C, Feuer A, et al. Nesprin-1 and -2
973 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for
974 nuclear envelope integrity. Hum Mol Genet. 2007;16:2816–33.

975 65. Doe JA, Wuebbles RD, Allred ET, Rooney JE, Elorza M, Burkin DJ. Transgenic
976 overexpression of the α7 integrin reduces muscle pathology and improves viability in the dy(W)
977 mouse model of merosin-deficient congenital muscular dystrophy type 1A. J Cell Sci.
978 2011;124:2287–97.

979 66. Amin S, Cook B, Zhou T, Ghazizadeh Z, Lis R, Zhang T, et al. Discovery of a drug
980 candidate for GLIS3-associated diabetes. Nat Commun. 2018;9:2681.

981 67. Melin AD, Fedigan LM, Hiramatsu C, Kawamura S. Polymorphic color vision in white-faced
982 capuchins (Cebus capucinus): Is there foraging niche divergence among phenotypes? Behav
983 Ecol Sociobiol. 2007;62:659–70.

984 68. Hiramatsu C, Tsutsui T, Matsumoto Y, Aureli F, Fedigan LM, Kawamura S. Color-vision
985 polymorphism in wild capuchins (Cebus capucinus) and spider monkeys (Ateles geoffroyi) in
986 Costa Rica. Am J Primatol. 2005;67:447–61.

987 69. Kennedy BG, Haley BE, Mangini NJ. Creatine kinase in human retinal pigment epithelium.
988 Exp Eye Res. 2000;70:183–90.

989 70. Gerding WM, Schreiber S, Schulte-Middelmann T, de Castro Marques A, Atorf J, Akkad DA,
990 et al. Ccdc66 null mutation causes retinal degeneration and dysfunction. Hum Mol Genet.
991 2011;20:3620–31.

992 71. Schreiber S, Petrasch-Parwez E, Porrmann-Kelterbaum E, Förster E, Epplen JT, Gerding
993 WM. Neurodegeneration in the olfactory bulb and olfactory deficits in the Ccdc66 -/- mouse
994 model for retinal degeneration. IBRO Reports. Elsevier; 2018;5:43–53.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

995 72. Tsutsui K, Otoh M, Sakurai K, Suzuki-Hashido N, Hayakawa T, Misaka T, et al. Variation in
996 ligand responses of the bitter taste receptors TAS2R1 and TAS2R4 among New World
997 monkeys. BMC Evol Biol. 2016;16:208.

998 73. Roth G, Dicke U. Evolution of the brain and intelligence. Trends Cogn Sci. 2005;9:250–7.

999 74. Gibson KR. Evolution of human intelligence: the roles of brain size and mental construction.
1000 Brain Behav Evol. 2002;59:10–20.

1001 75. Reader SM, Laland KN. Social intelligence, innovation, and enhanced brain size in primates.
1002 Proc Natl Acad Sci U S A. 2002;99:4436–41.

1003 76. Muntané G, Farré X, Rodríguez JA, Pegueroles C, Hughes DA, de Magalhães JP, et al.
1004 Biological Processes Modulating Longevity across Primates: A Phylogenetic Genome-Phenome
1005 Analysis. Mol Biol Evol. 2018;35:1990–2004.

1006 77. Freitas AA, de Magalhães JP. A review and appraisal of the DNA damage theory of ageing.
1007 Mutat Res. 2011;728:12–22.

1008 78. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, et al. Comparative analysis of
1009 bat genomes provides insight into the evolution of flight and immunity. Science. 2013;339:456–
1010 60.

1011 79. Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, et al. Insights into the evolution
1012 of longevity from the bowhead whale genome. Cell Rep. 2015;10:112–22.

1013 80. Seim I, Fang X, Xiong Z, Lobanov AV, Huang Z, Ma S, et al. Genome analysis reveals
1014 insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat Commun.
1015 2013;4:2212.

1016 81. Valenzano DR, Benayoun BA, Singh PP, Zhang E, Etter PD, Hu C-K, et al. The African
1017 Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of
1018 Lifespan. Cell. 2015;163:1539–54.

1019 82. de Magalhães JP, Costa J, Church GM. An analysis of the relationship between
1020 metabolism, developmental schedules, and longevity using phylogenetic independent contrasts.
1021 J Gerontol A Biol Sci Med Sci. 2007;62:149–60.

1022 83. Orkin JD, Campos FA, Myers MS, Cheves Hernandez SE, Guadamuz A, Melin AD.
1023 Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest.
1024 ISME J. 2019;13:183–96.

1025 84. Mallott EK, Amato KR. The microbial reproductive ecology of white-faced capuchins (Cebus
1026 capucinus). Am J Primatol. 2018;e22896.

1027 85. Young RA. Fat, Energy and Mammalian Survival. Integr Comp Biol. Narnia; 1976;16:699–
1028 710.

1029 86. Niimura Y, Matsui A, Touhara K. Acceleration of Olfactory Receptor Gene Loss in Primate
1030 Evolution: Possible Link to Anatomical Change in Sensory Systems and Dietary Transition. Mol
1031 Biol Evol. 2018;35:1437–50.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

87. Go Y. Lineage-specific expansions and contractions of the bitter taste receptor gene repertoire in vertebrates. Mol Biol Evol. 2006;23:964–72.

88. Li D, Zhang J. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. Mol Biol Evol. 2014;31:303–9.

89. Yoder AD, Larsen PA. The molecular evolutionary dynamics of the vomeronasal receptor (class 1) genes in primates: a gene family on the verge of a functional breakdown. Front Neuroanat. 2014;8:153.

90. Moriya-Ito K, Hayakawa T, Suzuki H, Hagino-Yamagishi K, Nikaido M. Evolution of vomeronasal receptor 1 (V1R) genes in the common marmoset (Callithrix jacchus). Gene. 2018;642:343–53.

91. Alfaro JWL, Lynch Alfaro JW, de Sousa E Silva J, Rylands AB. How Different Are Robust and Gracile Capuchin Monkeys? An Argument for the Use of Sapajus and Cebus. Am J Primatol. 2012;74:273–86.

92. Janzen D. Tropical dry forest: area de Conservación Guanacaste, northwestern Costa Rica. In: Perrow M DA, editor. Handbook of Ecological Restoration: Restoration in Practice. Cambridge University Press, Cambridge; 2002. p. 559–83.

93. Campos FA. A synthesis of long-term environmental change in Santa Rosa. In: Kalbitzer U, Jack KM, editors. Primate Life Histories, Sex Roles, and Adaptability - Essays in Honour of Linda M Fedigan. Springer, New York; 2018.

94. Fedigan L, Rose-Wiles L. See how they grow: Tracking capuchin monkey populations in a regenerating Costa Rican dry forest. In: Norconk MA, Rosenberger AL, Garber PA, editors. Adaptive radiations of Neotropical primates. Springer; 1996. p. 289–307.

95. Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. Nat Protoc. 2014;9:1038–48.

96. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8.

97. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

98. Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA, et al. OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. 2019;29:1152–63.

99. Brown JW, Walker JF, Smith SA. Phyx: phylogenetic tools for unix. Bioinformatics. 2017;33:1886–8.

100. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

1069 101. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with
1070 space/time models. Brief Bioinform. 2011;12:41–51.

1071 102. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of
1072 Phylogenomic Data. Mol Biol Evol. 2016;33:1635–8.

1073 103. Zhang J. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive
1074 Selection at the Molecular Level [Internet]. Molecular Biology and Evolution. 2005. p. 2472–9.
1075 Available from: http://dx.doi.org/10.1093/molbev/msi237

1076 104. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
1077 data. Bioinformatics. 2014;30:2114–20.

1078 105. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
1079 Bioinformatics. 2009;25:1754–60.

1080 106. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1081 Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

1082 107. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
1083 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
1084 sequencing data. Genome Res. 2010;20:1297–303.

1085 108. Bushnell B. BBMap short read aligner. University of California, Berkeley, California URL
1086 http://sourceforge.net/projects/bbmap. 2016;

1087 109. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal
1088 components analysis corrects for stratification in genome-wide association studies. Nat Genet.
1089 2006;38:904–9.

1090 110. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide
1091 allele frequency data. PLoS Genet. 2012;8:e1002967.

1092 111. Li H, Durbin R. Inference of human population history from individual whole-genome
1093 sequences. Nature. 2011;475:493–6.

1094 112. Monroy Kuhn JM, Jakobsson M, Günther T. Estimating genetic kin relationships in
1095 prehistoric populations. PLoS One. 2018;13:e0195491.

1096 113. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
1097 Bioinformatics. 2010;26:841–2.

1098 114. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the
1099 impact of rare variants. Genome Res. 2013;23:1514–21.

1100 115. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment
1101 analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37:W305–11.

1102 116. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian
1103 evolution. PLoS One. journals.plos.org; 2007;2:e708.

1104 117. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and

bioRxiv preprint doi: https://doi.org/10.1101/366112. this version posted July 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

1105 population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–
1106 93.

1107 118. Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. Ecological adaptation
1108 determines functional mammalian olfactory subgenomes. Genome Res. 2010;20:1–9.

1109 119. Hayden S, Bekaert M, Goodbla A, Murphy WJ, Dávalos LM, Teeling EC. A cluster of
1110 olfactory receptor genes linked to frugivory in bats. Mol Biol Evol. 2014;31:917–27.

1111 120. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, De Pristo MA, et al. The variant call
1112 format and VCFtools. BIOINFORMATICS APPLICATIONS NOTE [Internet]. 2011;27. Available
1113 from: http://dx.doi.org/10.1093/bioinformatics/btr330