# Convergent and lineage-specific genomic changes contribute to adaptations in sugar-consuming birds


## Author list

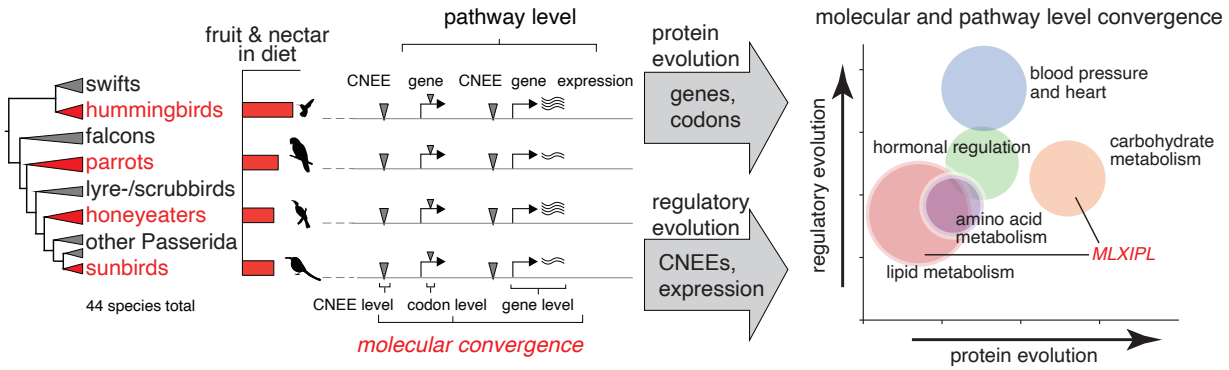
Ekaterina Osipova[*,1], Meng-Ching Ko[*,2], Konstantin M. Petricek[3,4], Simon Yung Wa Sin[5], Thomas Brown[6], Sylke Winkler[7], Martin Pippel[7], Julia Jarrells[7], Susanne Weiche[8], Mai-Britt Mosbech[9], Fanny Taborsak-Lines[10], Chuan Wang[11], Orlando Contreras-Lopez[10], Remi-André Olsen[12], Phil Ewels[12], Daniel Mendez-Aranda[2], Andrea Gaede[13], Keren Sadanandan[2], Gabriel Weijie Low[2], Amanda Monte[14], Ninon Ballerstädt, Nicolas M. Adreani[15], Lucia Mentesana[16], Auguste von Bayern[17], Alejandro Rico-Guevara[18], Scott Edwards[19], Carolina Frankl-Vilches[14], Heiner Kuhl[14], Antje Bakker[14], Manfred Gahr[14], Douglas Altshuler[20], William Buttemer[21], Michael Schupp[3], Maude Baldwin[†,2], Michael Hiller[†,22,23,24], Tim Sackton[†,1]

* co-first author
† co-last

1. Informatics Group, Harvard University
2. Evolution of Sensory and Physiological Systems, Max Planck Institute for Biological Intelligence
3. Institute of Pharmacology, Max Rubner Center (MRC) for Cardiovascular Metabolic Renal Research, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin
4. Institute of Pharmacology, Max Rubner Center (MRC) for Cardiovascular Metabolic Renal Research, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin
5. School of Biological Sciences, The University of Hong Kong
6. Leibniz Institute of Zoo and Wildlife Research
7. Max Planck Institute of Molecular Cell Biology and Genetics
8. Center for Molecular and Cellular Bioengineering, Technische Universität Dresden
9. Science for Life Laboratories, Department of Immunology, Genetics and Pathology, Uppsala University
10. Science for Life Laboratories, Division of Gene Technology, KTH Royal Institute of Technology
11. Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institutet
12. Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University
13. Royal Veterinary College
14. Department of Behavioural Neurobiology, Max Planck Institute for Biological Intelligence
15. Biology Department, Facultad de Ciencias, Universidad de la República
16. Faculty of Sciences, Republic University
17. Evolution of intelligent minds research group, Max Planck Institute for Biological Intelligence
18. Department of Biology, University of Washington
19. Department of Organismal and Evolutionary Biology, Harvard University
20. Department of Zoology, University of British Columbia
21. School of Earth, Atmospheric, and Life Sciences, University of Wollongong
22. LOEWE Centre for Translational Biodiversity Genomics
23. Faculty of Bioscience, Goethe-University
24. Senckenberg Research Institute, Frankfurt

# Abstract

Although high-sugar diets are associated with metabolic diseases in humans, several bird lineages have independently evolved to primarily subsist on simple sugars from flower nectar or fruits. In this study, we address a key question of the repeatability of molecular evolution by investigating the convergent and lineage-specific molecular mechanisms underlying dietary adaptations in four major sugar-consuming bird lineages: hummingbirds, parrots, honeyeaters, and sunbirds. We assembled nine new genomes for sugar-consuming species and their closely related non-sugar feeding outgroup species and generated 90 tissue-specific transcriptomes for six key species. We identified signatures of positive selection in both protein-coding and non-coding regulatory sequences, and found positive selection targets the same genes more frequently in sugar-feeders compared to non-sugar feeding controls, suggesting that adapting to a high-sugar diet requires changes in a limited number of genetic elements. At the functional level, pathways associated with energy homeostasis, carbohydrate metabolism, heart function, and hormonal regulation showed convergent selection signals in both protein-coding and regulatory evolution, while lipid and amino acid metabolism demonstrated mostly regulatory evolution. Notably, we observed striking evidence for convergent adaptation in MLXIPL, a transcription factor regulating sugar and lipid homeostasis, manifesting in both sequence and regulatory changes across all sugar-feeders. With functional assays, we demonstrated that hummingbird MLXIPL enhances sugar-induced transcriptional activity in HEK239 cells, suggesting its central role in the evolution of adaptations to high-sugar diets. Our findings elucidate the main genomic targets in the evolution of sugar-feeding at both molecular and pathway levels.


# Introduction

Animals vary tremendously in the foods they consume, and this imposes diverse requirements on their morphology, physiology, and behavior. Among birds, several groups including hummingbirds, sunbirds, and many honeyeaters and parrots independently adapted to feeding primarily on sugar-rich nectars and fruit [1,2]. Nectar provides mostly sugars [3] and contains little protein and fat, a challenge that birds overcome by supplementing nectar consumption with insects or pollen [4]. In humans, excessive sugar consumption is a major risk factor for diabetes

mellitus type 2 (T2D), a disease characterized by the failure of organs to regulate glucose uptake and production in response to insulin (insulin resistance). In contrast, nectarivorous birds exhibit morphological and metabolic adaptations to their sugar-rich diet, and revealing the genomic basis of these adaptations may provide insights that could be ultimately harnessed for human metabolic disease therapy.

Similar diets can select for convergent morphological and functional adaptations. In nectarivorous birds, these include elongated bills and elaborated (often brush-tipped) tongues to extract nectar efficiently [5,6]; short digestive tracts, and modified kidneys that aid in processing large amounts of liquid [1]; an enhanced ability to extract and metabolize sugars [7]; high mass-specific metabolic rates [8,9]; elevated paracellular transport rates to increase intestinal glucose transport; high sucrase activities to hydrolyze sucrose into glucose and fructose in the small intestine, and low daily protein requirements [1,2]. Hummingbirds have the striking ability to metabolize ingested sugar directly instead of converting a substantial portion into fatty acids, and almost their entire energy metabolism is fueled by recently ingested sugars [10].

Previous studies, largely focused on hummingbirds, have identified some genomic correlates of physiological and metabolic adaptations to high-sugar diets. For example, hummingbirds lost the gluconeogenic muscle enzyme FBP2, which enhanced glycolysis and mitochondrial respiration in muscle cells and likely facilitated the evolution of their energy-demanding hovering flight [11]. Transcriptomic and genomic analyses of hummingbirds have suggested fast evolution of lipogenic pathways [12] and revealed changes in sugar transporters [13]. Studies of taste receptors have revealed that hummingbirds and songbirds have convergently repurposed the amino-acid taste receptor (T1R1-T1R3) to sense sugars in nectar [14,15]. Despite these advances, many relevant genomic adaptations remain to be discovered, and the extent to which convergent adaptations to sugar-rich diets arise from convergent molecular changes is not known. Molecular convergence - defined here as changes affecting the sequence or regulation of the same gene in at least two independent lineages - can indicate relatively constrained possibilities for adapting to a high-sugar diet, in contrast to changes affecting distinct (but potentially functionally-similar) genes or pathways. The repeated evolution of bird lineages consuming sugar-rich diets thus provides a powerful system to investigate the role of convergent molecular changes in adaptation, a central question in evolutionary biology.

To address these questions and uncover adaptation to high-sugar diets, we combined genome sequencing with comprehensive comparative genomic and transcriptomic analyses and experimental validations to uncover both convergent and lineage-specific genomic adaptations to a sugar-rich diet in four avian clades containing nectar-specialists: hummingbirds, sunbirds, honeyeaters, and parrots. We include species with high levels of fruit or nectar consumption in these four focal ‘sugar-consuming’ lineages. Whereas in hummingbirds, elevated sugar consumption evolved once and was retained throughout the clade, in songbirds (as well as in parrots), sugar may have been an early part of the diet [15,16]. Nevertheless, we expect that high levels of sugar intake in honeyeaters, sunbirds, and many parrots likely led to elaborated metabolic and physiological adaptations, detectable as genomic signatures. We performed an integrated analysis, considering several types of evolutionary signatures in protein-coding genes

and in gene expression, as well as changes in non-coding genomic regions to address the prevalence of convergent or lineage-specific molecular changes at multiple levels.

## 1. Genome assembly, annotation, and comparative analyses

To investigate the genomic basis of adaptations to sugar-rich diets across hummingbirds, parrots, honeyeaters, and sunbirds, we used long and short read sequencing to assemble new genomes for five nectar- or fruit-taking birds and four outgroup species (Fig. 1A-C, table S1). These genomes are highly complete: we recovered 95.5% to 97.3% of 8,338 universal single-copy avian orthologs (BUSCO odb10 gene set) [17] (fig. S1C, table S1). We also assessed gene completeness with TOGA, a method that integrates comparative gene annotation, inferring and classifying orthologous genes [18], which recovered ≥10,092 intact coding genes of 15,133 ancestral bird genes across our nine new assemblies (Fig. 1D). Together with 35 previously generated assemblies [19,20], these genomes covered all four sugar-consuming clades, each with several representative species, and provided the basis for our analysis.

Comprehensive protein-coding gene annotations are key to studying molecular changes underlying phenotypic differences. To produce highly complete protein-coding gene annotations, we adapted a gene annotation pipeline [21] that combined TOGA's ortholog predictions, transcriptomic data, protein alignments, and *ab initio* gene predictions (fig. S1D). Applying this pipeline to our nine new and eight previous assemblies that lacked an annotation generated by NCBI [22] produced annotations with a high BUSCO completeness of 93.2% - 99.6% for all 17 genomes, comparable or better than existing NCBI annotations (Fig. 1E).

Adaptations to sugar consumption may be the result of the convergent and lineage-specific evolution of protein-coding genes as well as regulatory evolution, which manifests as variation in gene regulatory elements or gene expression patterns [11,13]. Our analyses are designed to capture all of these aspects. To investigate protein-coding evolution, we screened 11,688 one-to-one orthologs identified with TOGA [23] for signatures of episodic positive selection in each branch in the phylogeny using the branch-site method implemented in aBSREL [24]. We also identified convergent amino acid substitutions in sugar-consuming clades with CSUBST [25], a method that measures the ratio of the non-synonymous to synonymous convergence rate. To investigate regulatory evolution, we screened 363,476 avian conserved non-exonic elements (CNEEs) [26–29] for rate shifts in one or more sugar-consuming lineages with the Bayesian method PhyloAcc [30]. Furthermore, we generated 90 transcriptomic (RNA-seq) datasets for five tissues relevant to metabolism and digestion for three pairs of sugar-consuming species with a non-sugar outgroup, each in three biological replicates (Fig. 1F, table S11). We investigated gene expression changes with (i) pairwise differential expression analyses, considering comprehensive sets of >11,000 one-to-one orthologs and ranking expression changes by a metric that jointly considers significance and effect size (fig. S2A); and (ii) a multifactor analysis to detect smaller sugar-associated tissue-specific expression shifts between the key metabolic tissues - pectoralis and liver - that are likely explained by selection rather than drift (see fig. S2B and Methods). Together, these methodological approaches provide a comprehensive analysis of genomic and transcriptomic changes in sugar-consuming species.

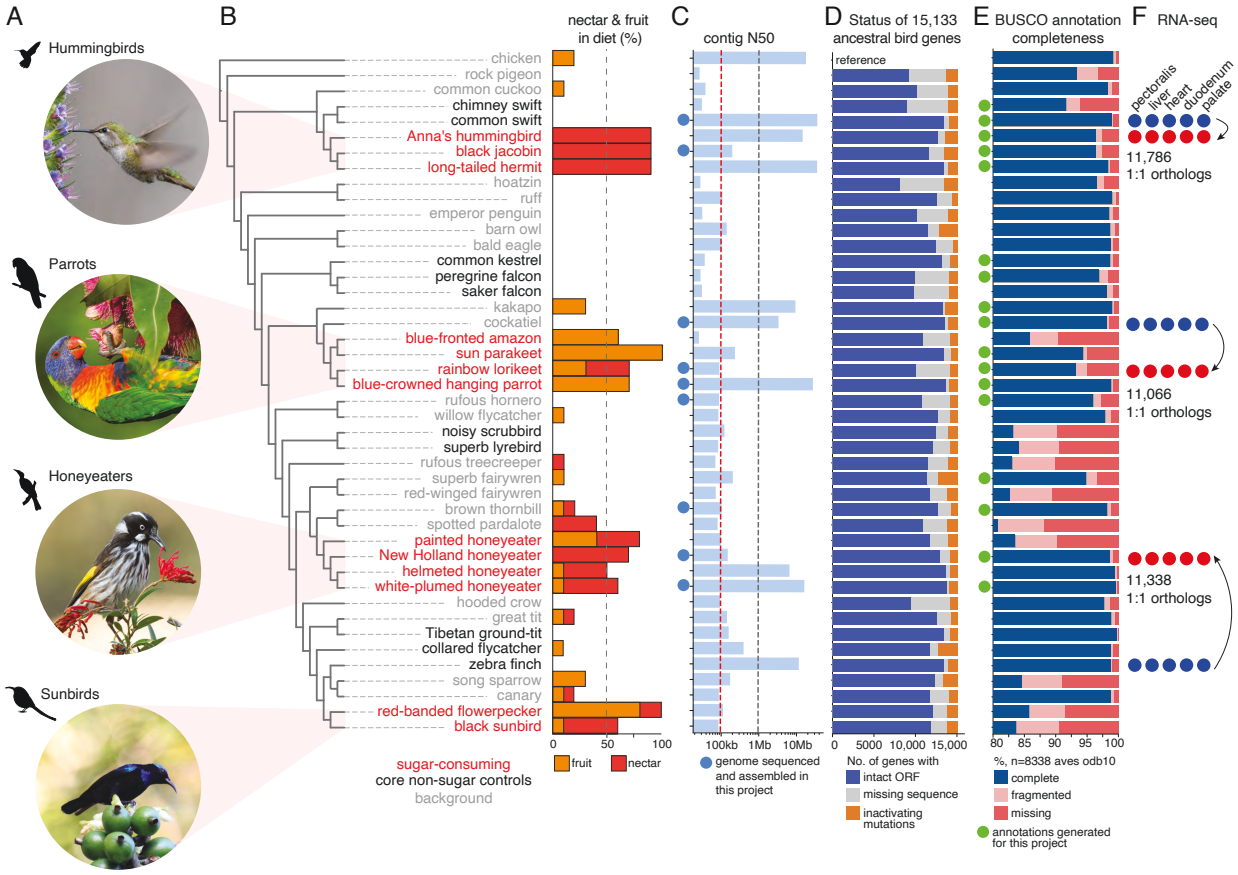



Figure 1. Genome assemblies and annotations

(A) Species representing the four sugar-consuming groups. Bird images from the Macaulay Library at the Cornell Lab of Ornithology: ML426078561, ML431374521, ML285318411, ML205719751.

(B) Phylogeny and diet of the birds included in the study. Based on the percentage of dietary nectar and fruit intake, we classified species as core sugar-consuming birds (red font), core non-sugar controls (black), or background (gray). Core non-sugar control sets were defined as groups phylogenetically close to sugar-consuming groups and comprising a similar number of species, but not taking any nectar in the diet. These were used as direct controls in the downstream analyses.

(C) Comparison of genome assembly contiguity. The contig N50 value indicates that 50% of the assembly consists of contigs of at least that size. The nine newly-sequenced birds are indicated with a blue circle.

(D) Status of 15,133 ancestral bird genes, classified by TOGA as those with an intact reading frame (blue), with gene inactivating mutations (premature stop codons, frameshifts, splice site disruptions, and deletions of exons or entire genes; orange), or missing or incomplete coding sequences due to assembly gaps or fragmentation (gray).

(E) Status of 8,338 near-universally conserved aves genes (BUSCO, odb10) in annotations generated in this study (green circle) or NCBI RefSeq annotations.

(F) Tissue-specific transcriptomes generated for pairs of sugar - non-sugar control species. Black arrow specifies the direction of DESeq2 analysis (for details, see Methods and table S11).

## 2. Adaptation to sugar-feeding involves convergent evolution in both protein-coding and regulatory sequences.

Our positive selection screen revealed 4,911 protein-coding genes with a selection signal on at least one branch in the 44-bird phylogeny, of which 1,030 (21%) are under selection predominately in sugar-consuming branches (with a more frequent signal of selection in target branches compared to background and core control). The 1,030 genes under selection in sugar-consuming groups include 501, 373, 331, and 323 that are under selection in hummingbirds, honeyeaters, parrots, and sunbirds respectively (Fig. 2A, fig. S3A, table S2). These genes are much more likely to show convergent patterns of evolution than genes under selection predominantly in non-sugar species. Of the 1,030 sugar-selected genes, 340 (33.0%) are under selection in multiple convergent sugar-consuming lineages, with 216 (21.0%), 90 (8.7%) and 34 (3.3%) genes being selected in two, three and all four lineages (Fig. 2A). In matched non-sugar controls - swifts, falcons, lyre- and scrubbirds, and control Passerida lineages (fig. S3B,C) - we find fewer genes (787) selected predominantly in these lineages (with a non-sugar relevance ratio >0.5), and lower proportions under selection in two (132 genes, 16.8%), three (40 genes, 5.1%) or all four (10 genes, 1.3%) lineages (fig. S2B-D). Using a two-sided Fisher's exact test, we found that in comparison to the controls, sugar-consuming clades show significantly more convergent selection signals for five of the six species pairs (the exception is parrots-sunbirds), and all three- and four-way comparisons (Fig. 2B, fig. S3D).

Convergent signals of positive selection can be due to changes in the same or different amino acids. Of the 11,688 codon alignments used in the selection screen, CSUBST [25] identified pairwise codon-level convergence among two sugar-consuming clades in only 55 genes (503 genes among non-sugar clades). Although CSUBST also detects convergence involving more than two branches, no gene exhibited convergent codon substitutions among more than two sugar branches. To quantify the extent of overlap between convergently selected genes and CSUBST signals, we assessed the fraction of genes showing CSUBST signals in the overlapping clades. In sugar-consuming lineages, only 3 out of 340 (0.88%) convergently selected genes exhibited overlapping CSUBST signals, compared to 11 out of 182 (6.04%) genes in non-sugar control lineages. This discrepancy suggests that although convergent positive selection is more prevalent in sugar-consuming lineages, it most often occurs through changes at different amino acid positions rather than identical substitutions. These findings indicate that the observed positive selection signals in three or four lineages are often due to changes at different codons.

We observe similar patterns of convergent evolution in CNEEs. Using PhyloAcc, we identified 11,607 (3.2% of 363,476) CNEEs accelerated in sugar-feeders: 6,930 in hummingbirds, 7,269 in parrots, 3,141 in honeyeaters, and 2,181 in sunbirds (fig. S5A). Among these, 3,768 CNEEs (32%) were accelerated in two sugar-consuming clades, 1,398 (12%) in three, and 450 (3.9%) across all four (Fig. 2D). Among the same core non-sugar control lineages described above, we find fewer accelerated CNEEs and a lower proportion of convergent acceleration: 3,642 in total, with 1,771 in swifts; 1,072 in falcons; 892 in lyre- and scrubbirds; and 1,233 in control Passerida lineages (fig. S5B), with 773 (21%) CNEEs convergently accelerated in two clades, 218 (6.0%) in three, and 39 (1.1%) in all four (Fig. 2D). Two-sided Fisher's exact tests revealed significantly

more convergent CNEE acceleration in sugar-consuming birds than in controls for five out of six pairs (the exception being the hummingbird-sunbird pair) and in all comparisons involving three or four groups (Fig. 2E, fig. S5C).

Convergent regulatory changes can also occur at the gene level, where distinct accelerated CNEEs regulate the same target gene (Fig. 2D). To investigate this, we first assigned CNEEs to potential target genes, using the procedure of GREAT [31], which considers both a promoter-proximal and an extended domain that reaches until the nearest upstream and downstream genes within a 1 Mb limit. The 11,607 CNEEs accelerated in sugar-feeders are associated with 4,992 genes (3,958 genes for hummingbirds; 3,802 for parrots, 2,424 for honeyeaters; and 2,048 for sunbirds; fig. S5D), whereas the 3,642 CNEEs accelerated in core non-sugar groups are associated with 2,905 genes (1,805 genes in swifts; 1,375 in falcons; 974 in lyre- and scrubbirds; and 1,489 in control Passerida species). We found more gene level convergence in sugar-consuming species compared to non-sugar controls, with 1,404 (28%) vs. 772 (27%) genes being associated with accelerated CNEEs in two clades, 995 (20%) vs. 532 (18%) in three clades, and 1,282 (26%) vs. 301 (10%) in all four clades (Fig. 2D, fig. S5E). Similar to the CNEE-level analysis, a two-sided Fisher's exact test revealed a significantly higher gene-level convergence for five out of six pairs (all pairs except the hummingbird-sunbird pair) and in all scenarios involving three or four groups (Fig. 2E).

Together, these results show higher convergence in sugar-feeders affecting both protein-coding and regulatory regions. In the following sections, we present functional enrichments of genes with sequence, regulatory and expression changes in sugar-feeders, relating to carbohydrate, lipid, amino acid metabolism, heart function, and overlaps with human T2D loci.

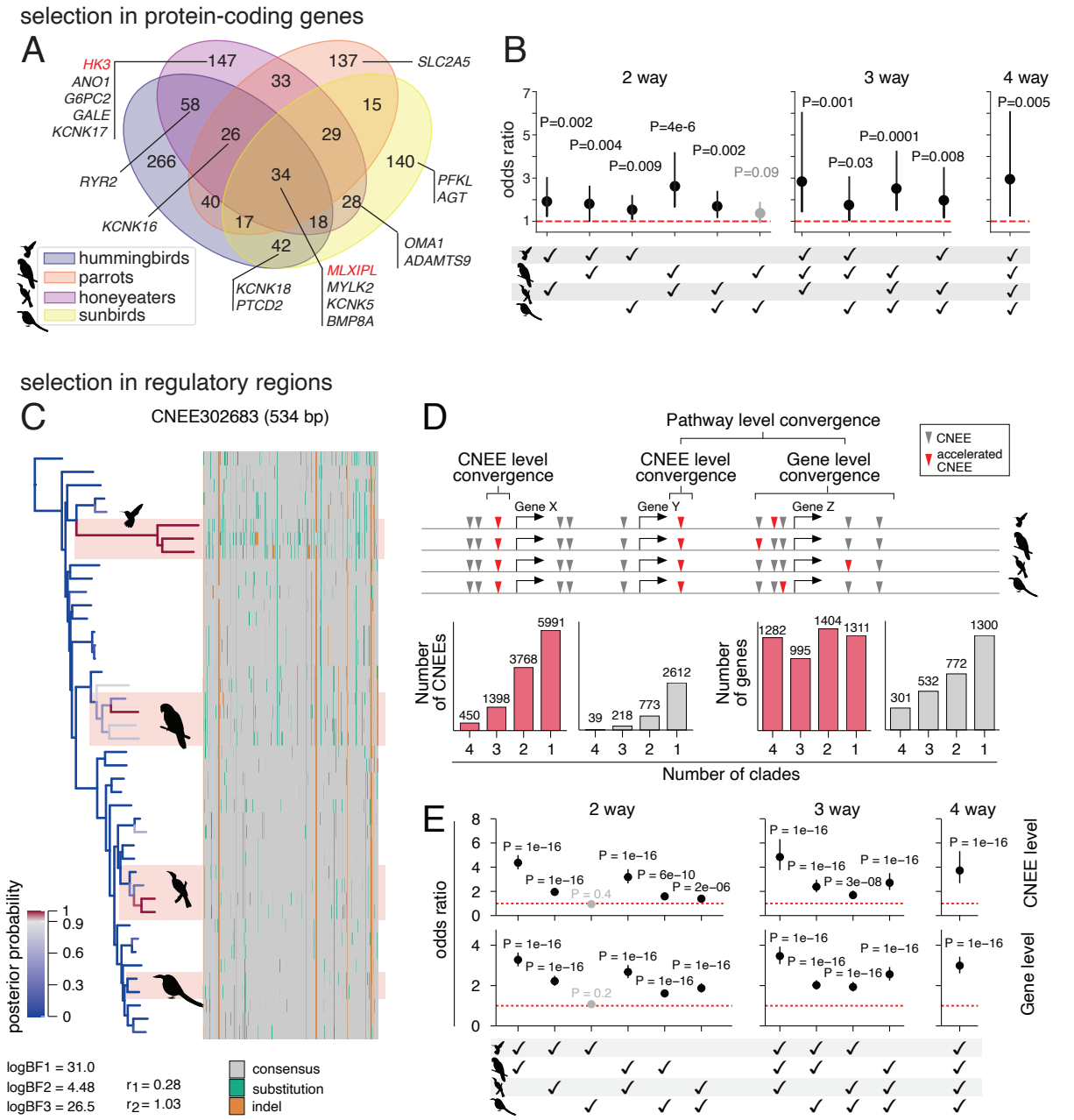



**Figure 2.** Genome-wide screens reveal higher convergence in sugar-consuming lineages in protein-coding and regulatory regions.

(A) Number and overlap of genes under positive selection (p-value<0.05% and relevance ratio>0.5) in four sugar-consuming clades.

(B) Excess of convergent positive selection in sugar-feeders. Counts of genes showing convergent vs. lineage-specific positive selection were compared between sugar-consuming groups and matching core non-sugar controls, considering all pairs, triplets, and the quadruple. The odds ratio is >1 for all comparisons. Significant differences are in black font.

(C) Example of a CNEE (associated with *MAT2B*) with accelerated substitution rates in hummingbird, parrot, and honeyeater lineages. Branches in the tree (left) are colored according to the posterior probability of being in an accelerated state; those with a posterior probability >0.9 are depicted in red. Bayes factors, conserved (r1), and accelerated rates (r2) are below the tree; branch lengths correspond to the posterior mean of the substitution rate. The sequence alignment representation uses gray color for positions that are identical to the predominant nucleotide; turquoise and orange indicate deviations from the consensus (substitutions or indels).
(D) Convergent accelerations can be at the CNEE, gene, or pathway level. The number of accelerated CNEEs (bottom left) and their associated genes (bottom right) illustrate the extent of overlap across sugar-consuming (red-filled) or core non-sugar (gray-filled) clades. The x-axis indicates the number of overlapping clades, representing the degree of convergence.
(E) Excess of convergent acceleration among sugar-feeders at CNEE (top) and gene levels (middle), showing more frequent acceleration in two-, three-, and four-way comparisons than in paired non-sugar controls. Significant differences are shown in black.
A two-sided Fisher's exact test was used in (B) and (E).

## 3. Both sequence and regulatory evolution affect carbohydrate metabolism in sugar-feeders.

### 3.1. Convergent and lineage-specific changes in glucose-utilization pathways in sugar-feeders.

Genes showing convergent positive selection in at least two sugar-consuming clades are enriched in Gene Ontology (GO) terms associated with energy homeostasis and sugar metabolism ("multicellular organismal-level homeostasis" and "carbohydrate biosynthesis") (Fig. 3A, table S3). Many of the genes contributing to these enrichments are important regulators of metabolic stress response, including three notable genes with evidence for convergent selection in three or four sugar-consuming clades: *OMA1*, *BMP8A*, and *MLXIPL* (Fig. 2A, table S3). *OMA1* and *BMP8A* are both involved in regulating energy homeostasis in mice [32,33]. *MLXIPL* encodes the transcription factor ChREBP (carbohydrate-responsive element-binding protein) that regulates cellular responses to high concentrations of simple carbohydrates such as glucose and fructose [34,35]. ChREBP is activated by glucose metabolites and promotes expression of glycolytic and lipogenic pathways, thus enhancing the capacity to metabolize ample amounts of sugar [36,37]. Similarly, genes associated with convergently accelerated CNEEs are enriched in metabolic processes and electron transport chain (Fig. 3C). In contrast, a similar analysis did not identify any enriched pathways for convergently accelerated CNEEs in at least two core non-sugar clades.

In addition to molecular convergence, higher-level convergence can occur at the pathway level, which may involve selection or regulatory changes in different genes with similar functions. We identified sugar-specific pathway enrichments by first identifying GO enrichments for each of the four sets of genes under selection in the four sugar-consuming clades, and then excluded identical or semantically similar GO terms enriched in genes selected in the respective control lineages. The only pathway convergently enriched for genes under selection in all four sugar-consuming clades is related to the stabilization of membrane potential (Fig. 3B), and is driven by

four potassium transporters having a broad range of functions from the regulation of insulin secretion in the pancreas (*KCNK16*) to muscle contraction, cardiac function, and neuronal signal transduction (*KCNK5*, *KCNK17*, and *KCNK18*). Interestingly, fruit-eating bats also demonstrate expression shifts in several potassium channels [38]. Overall, pathways convergently enriched for selected genes in at least three sugar-consuming clades are similar to pathways enriched for convergently selected genes, and include terms associated with a shift in nutrient source, such as energy homeostasis, response to dietary excess and diet-induced thermogenesis (Fig. 3B, table S6).

Overall, we detected an overwhelming signal of accelerated protein-coding and regulatory evolution in a number of genes involved in carbohydrate metabolism (Fig. 3D). First, sugar-feeders exhibit a higher capacity for sucrose hydrolysis compared to generalist birds [39,40]. We found that two genes central for sugar breakdown, *AMY2A* and *SI*, are associated with CNEEs accelerated in all four sugar-feeders (fig. S6A**,** table S7). *AMY2A* encodes α-amylase, which hydrolyzes polysaccharides into simpler sugars [41], while sucrase-isomaltase (*SI*) breaks these down further, accounting for approximately 80% of intestinal maltase and almost all sucrase activity [42,43]. *SI* is part of a dual enzyme system crucial for disaccharide and starch digestion in vertebrates, with duplications and functional convergence observed across different lineages [44]. Enhanced mono-saccharide import is likely facilitated by several glucose transporters GLUT1 (encoded by *SLC2A1)*, GLUT2 (*SLC2A2*), and SGLT1 (*SLC5A1*), all of which show differential expression and acceleration in associated CNEEs across at least two sugar-consuming groups. Furthermore, the fructose transporter GLUT5 (*SLC2A5*) shows higher expression in all three tested sugar-consuming groups and signals of positive selection in sugar-consuming parrots (Fig. 3D). Key glycolytic enzymes such as *HK3, PGD, PFKL*, and *ALDOB* exhibit convergent signals across sugar-consuming groups in multiple analyses. *PFKP*, encoding an isoform of phosphofructokinase, which catalyzes a rate-limiting step in glycolysis, shows signatures of regulatory evolution (either differential expression or PhyloAcc) across all four sugar-consuming groups. Enolase genes *ENO1* and *ENO4* also exhibit signals of convergent regulatory evolution. The final step of gluconeogenesis, catalyzed by glucose 6-phosphatase (encoded by *G6PC* and *G6PC2*), shows sequence and regulatory evolution in at least two sugar-consuming groups (Fig. 3D).

Several genes regulating glucose and fructose metabolism also have convergently accelerated CNEEs and show convergent evidence of differential expression between sugar-consuming and non-sugar-consuming species. *AKR1B10L2* encodes an aldose reductase, an enzyme that converts glucose to sorbitol in an NADPH-dependent manner [45]. As part of the polyol pathway, *AKR1B10L2* plays a crucial role in glucose regulation by facilitating the conversion of excess glucose to fructose, which is then metabolized to pyruvate at a faster rate than glucose [46]. Glycolysis and gluconeogenesis are closely linked to glycogen production, where we also detected several genes with signals of sequence (*GYG1*) and regulatory evolution (*GYS1* and *GYS2*, and genes encoding protein phosphatase 1 subunits: *PPP1R**).

Finally, glycolysis feeds into the TCA cycle, the central hub of cellular metabolism, where many genes also show convergent regulatory evolution signals (Fig. 3D). Additionally, *MDH1*, encoding malate dehydrogenase 1 - a key component of the malate-aspartate shuttle connected to the TCA

cycle - shows protein and regulatory evolution across all four sugar-feeders. Since glycolysis generates NADH, which inhibits regulatory enzymes in the TCA cycle [47], cytosolic *MDH1* is critical. It oxidizes NADH to NAD+ by reducing oxaloacetate to malate, maintaining redox balance and regulating ATP production.

## 3.2. Functional shift in honeyeater hexokinase 3.

Our positive selection screen detected a strong and specific signal in *HK3* (Hexokinase 3) at the ancestor of honeyeaters (fig. S7). As one of five hexokinase enzymes, the HK3 homodimer catalyzes the phosphorylation of glucose in the first and one of the rate-limiting steps of glycolysis (Fig. 3D,E). HK3 contains a catalytic C-domain and a regulatory N-domain, which face each other in the asymmetric homodimer and form two substrate-binding sites [48,49]. To investigate the structural location of amino acid residues affected in the honeyeater protein, we modeled the structure of the honeyeater enzyme based on the available structure of a human hexokinase 2 as a template. Notably, most sites under selection are located on the interface between the catalytic and regulatory domains (Fig. 3E). Such proximity suggests that selection in the honeyeater protein targeted allosteric regulation and did not affect the catalytic site directly. In addition to positive selection on the branch leading to all honeyeaters, CSUBST also identified codon convergence shared between this branch and the ancestral hummingbird branch, with three convergent codons being located in close proximity to the sites identified by PAML (Fig. 3E). Additionally, our expression data showed the dramatic upregulation of *HK3* across all tested tissues (except the heart) in the honeyeater (Fig. 3F), further emphasizing the importance of the evolution of protein sequence and regulation of *HK3* in honeyeaters.

To assess functional changes in honeyeater HK3, we purified the HK3 proteins from the New Holland honeyeater and its closely related outgroup species (superb lyrebird and rufous treecreeper) and conducted a hexokinase activity assay. Consistent with previous studies on HK3 from several species [50–54], we observed substrate inhibition, according to which enzymatic activity decreases with excess substrate after reaching a maximum turnover rate at a glucose concentration of 0.5 mM [55]. Therefore, we compared enzymatic kinetic properties by determining Vmax (maximum velocity), Km (Michaelis constant, indicating the affinity for the substrate), and Ksi (substrate inhibition constant, indicating the concentration at which substrate begins to inhibit the enzyme). The honeyeater HK3 exhibited significantly higher Vmax (21.1 mM/min/mL) compared to control species (treecreeper: 8.0; lyrebird: 10.6 mM/min/mL; non-overlapping 95% CIs). The honeyeater HK3 also demonstrated the highest Km (65.9 μM), significantly different from the treecreeper (26.2 μM) and higher than the lyrebird (30.6 μM) (Fig. 3G, fig. S8). Honeyeater hexokinase 3 has a low affinity for glucose (high Km) and a high maximum rate, similar to mammalian glucokinase making the enzyme ideal for processing glucose at high blood glucose concentrations [56]. This indicates a shift in the kinetic properties of the enzyme, reflecting adaptations to high-glucose diet in honeyeaters.

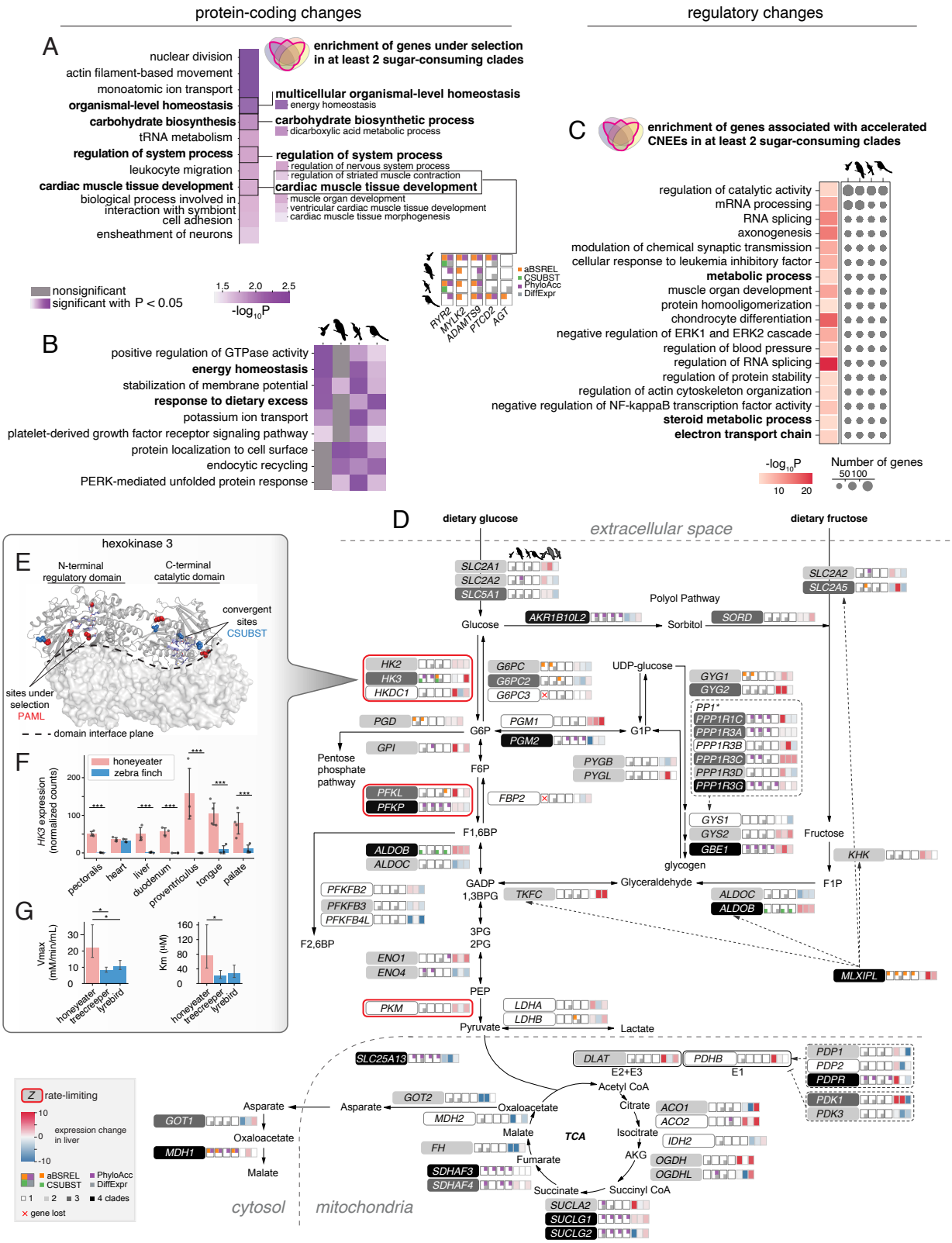

protein-coding changes
regulatory changes
A
enrichment of genes under selection in at least 2 sugar-consuming clades
nuclear division
actin filament-based movement
monoatomic ion transport
organismal-level homeostasis
carbohydrate biosynthesis
tRNA metabolism
regulation of system process
leukocyte migration
cardiac muscle tissue development
biological process involved in interaction with symbiont
cell adhesion
ensheathment of neurons
multicellular organismal-level homeostasis
energy homeostasis
carbohydrate biosynthetic process
dicarboxylic acid metabolic process
regulation of system process
regulation of nervous system process
regulation of striated muscle contraction
cardiac muscle tissue development
muscle organ development
ventricular cardiac muscle tissue development
cardiac muscle tissue morphogenesis
nonsignificant
significant with P < 0.05
1.5
2.0
2.5
-log10P
RYR2
MYLK2
ADAMTS9
PTCD2
AGT
aBSREL
CSUBST
PhyloAcc
DiffExpr
B
positive regulation of GTPase activity
energy homeostasis
stabilization of membrane potential
response to dietary excess
potassium ion transport
platelet-derived growth factor receptor signaling pathway
protein localization to cell surface
endocytic recycling
PERK-mediated unfolded protein response
C
enrichment of genes associated with accelerated CNEEs in at least 2 sugar-consuming clades
regulation of catalytic activity
mRNA processing
RNA splicing
axonogenesis
modulation of chemical synaptic transmission
cellular response to leukemia inhibitory factor
metabolic process
muscle organ development
protein homooligomerization
chondrocyte differentiation
negative regulation of ERK1 and ERK2 cascade
regulation of blood pressure
regulation of RNA splicing
regulation of protein stability
regulation of actin cytoskeleton organization
negative regulation of NF-kappaB transcription factor activity
steroid metabolic process
electron transport chain
-log10P
10
20
Number of genes
50
100
D
dietary glucose
extracellular space
dietary fructose
SLC2A1
SLC2A2
SLC5A1
SLC2A2
SLC2A5
Polyol Pathway
Glucose
AKR1B10L2
Sorbitol
SORD
HK2
HK3
HKDC1
G6PC
G6PC2
G6PC3
UDP-glucose
GYG1
GYG2
PP1*
PPP1R1C
PPP1R3A
PPP1R3B
PPP1R3C
PPP1R3D
PPP1R3G
PGD
PGM1
PGM2
G6P
G1P
GPI
Pentose phosphate pathway
F6P
PYGB
PYGL
PFKL
PFKP
FBP2
GYS1
GYS2
GBE1
Fructose
F1,6BP
ALDOB
ALDOC
glycogen
KHK
PFKFB2
PFKFB3
PFKFB4L
GADP
1,3BPG
TKFC
Glyceraldehyde
F1P
ALDOC
ALDOB
F2,6BP
3PG
2PG
ENO1
ENO4
MLXIPL
PEP
PKM
LDHA
LDHB
Pyruvate
Lactate
SLC25A13
DLAT
PDHB
E2+E3
E1
PDP1
PDP2
PDPR
PDK1
PDK3
Acetyl CoA
Asparate
GOT2
Oxaloacetate
Citrate
ACO1
ACO2
GOT1
Asparate
MDH2
Oxaloacetate
Malate
TCA
Isocitrate
IDH2
MDH1
Malate
FH
Fumarate
AKG
OGDH
OGDHL
SDHAF3
SDHAF4
Succinate
Succinyl CoA
SUCLA2
SUCLG1
SUCLG2
cytosol
mitochondria
Z rate-limiting
10
0
-10
expression change in liver
aBSREL
CSUBST
PhyloAcc
DiffExpr
1
2
3
4 clades
gene lost
E
hexokinase 3
N-terminal regulatory domain
C-terminal catalytic domain
convergent sites
CSUBST
sites under selection
PAML
domain interface plane
F
honeyeater
zebra finch
HK3 expression (normalized counts)
200
100
0
pectoralis
heart
liver
duodenum
proventriculus
tongue
palate
G
Vmax (mM/min/mL)
30
20
10
0
Km (μM)
160
120
80
40
0
honeyeater
treecreeper
lyrebird

Figure 3. Genes involved in carbohydrate metabolism and cardiac function exhibit convergent changes across sugar-consuming clades.

(A) GO Biological Processes significantly enriched (p-value<0.05) for genes showing signal of selection in at least two sugar-consuming groups. Purple color indicates significance. The insets show child terms of the selected pathways. The genes driving the enrichment in the cardio-associated pathways are shown in the gene inset summarizing all analyses results.
(B) GO Biological Processes significantly enriched (p-value<0.05) for positively selected genes in at least three sugar-consuming clades. The terms relevant to metabolic adaptations are in bold.
(C) Enrichment analysis of sugar-specific accelerated CNEEs. Heatmap shows pathways with significant enrichment in both binomial and hypergeometric tests (FDR-adjusted P < 0.05). Color intensity indicates strength of binomial test enrichment (-log10 P). Circle sizes represent gene counts near accelerated CNEEs for each sugar-consuming clade.
(D) Integrating results from aBSREL, PhyloAcc, CSUBST, and differential expression analysis (DiffExpr, both between-species pairwise comparisons across all analyzed tissues and multifactor cross-tissue analyses), highlights changes in sugar-feeders in glucose and fructose metabolism (glycolysis, gluconeogenesis, glycogen synthesis, TCA cycle). Signals from the four analyses are represented in the four quadrants of a square. Each square represents a sugar-consuming clade, arranged in the order of hummingbirds, parrots, honeyeaters, and sunbirds. The heatmap panels to the right of the quadrants illustrate the differential expression in the liver of hummingbirds, parrots, and honeyeaters compared to their respective outgroups. Genes with signals from the combined analyses detected in all four sugar-consuming clades are shaded black, those in three clades are dark gray, and those in two clades are light gray. The rate-limiting steps of glycolysis are highlighted with red boxes. The dotted arrows indicate some of the downstream targets of MLXIPL.
(E) Protein model of the honeyeater hexokinase 3 based on the human *HK3* ortholog.
(F) Expression of *HK3* across tissues in the New Holland honeyeater compared to its non-sugar outgroup (zebra finch).
(G) Kinetic characteristics of the honeyeater recombinant HK3 compared to its control outgroups (treecreeper and lyrebird). Error bars indicate the 2.5% and 97.5% confidence intervals.

## 4. Pathways regulating heart function show convergent evolution in protein-coding genes.

Apart from carbohydrate metabolism, genes with convergent positive selection are also enriched in terms relating to heart development and performance ("regulation of system process" and "cardiac muscle tissue development") (Fig. 3A, table S3). Additionally, enrichment analysis of genes associated with accelerated CNEEs across all four sugar-consuming clades revealed significant enrichment for heart development (table S9). Both glucose and fructose - the main monosaccharides in nectar - are linked to elevated blood pressure, likely through the insulin pathway (glucose) and by altering nitric oxide levels (fructose) [57,58], suggesting a mechanism of selection on genes regulating blood vessel and heart muscle development in species naturally exposed to high blood sugar levels. All four sugar-consuming clades show signatures of positive selection in *MYLK2* (Fig. 2A, Fig. 3A (inset)), a gene encoding a kinase crucial for myosin structure development in cardiac muscle. Hummingbirds, honeyeaters, and sunbirds show signatures of selection in *ADAMTS9*, a gene encoding a secreted enzyme that contributes to extracellular matrix organization and is required for the differentiation of vascular smooth muscle cells that

regulate systemic blood pressure [59]. Hummingbirds and honeyeaters show signatures of selection and codon convergence in *RYR2* (Fig. 3A (inset) and fig. S4), encoding the major Ca2+ release channel in the sarcoplasmic reticulum of cardiomyocytes, which, in addition to mediating cardiac muscle contraction, stimulates oxidative phosphorylation in cardiomyocyte mitochondria by the Ca2+-dependent activation of PDH, the enzyme that converts pyruvate to acetyl-CoA and links glycolysis with glucose oxidation [60,61]. Apart from its well-known role in the heart, *RYR2* is also expressed in pancreatic beta-cells, where it regulates Ca2+ release from the endoplasmic reticulum, thus affecting insulin secretion and glucose homeostasis [62,63]. Hummingbirds and sunbirds show signatures of selection in *PTCD2*, which encodes an RNA-binding protein that is required for mitochondria function in heart and skeletal muscle by facilitating the efficient translation of the respiratory complex IV catalytic subunit COX3 [64,65]. *AGT*, which is under selection in sunbirds, encodes angiotensinogen, the precursor of the main vasoconstrictor controlling blood pressure. Together, our comparative screens detect molecular convergence in heart-related genes as a novel signature of adaptations to sugar-rich diets.

## 5. Convergent evolution of regulatory pathways in insulin signaling, renin-angiotensin system, and type 2 diabetes genes.

Given the shared metabolic pathways and physiological traits involved in heart function and glucose metabolism in sugar-consuming birds, we investigated an overlap with loci implicated in T2D, a human disease intricately linked to these processes. We used data from a recent large-scale meta-analysis that identified 611 loci associated with T2D; these were grouped into eight clusters based on their genetic contributions [66]. Two-sided Fisher's exact tests revealed a significant enrichment of sugar-accelerated CNEEs in T2D loci clusters related to body fat and dysfunctional beta cells, while non-sugar-accelerated CNEEs showed no such enrichment (Fig. 4A). Interestingly, we find little evidence for convergent positive selection among proteins encoded by T2D loci: only a single gene (*CTNNAL1*) is positively selected in all four sugar-consuming clades. In contrast, we find convergently accelerated CNEEs near 74 T2D loci, including a number of genes that are crucial mediators of insulin signaling, beta cell function, and glucose homeostasis such as *IRS2*, *IGF1*, *IGF2*, and *KCNQ1* (table S10) [67–70], consistent with T2D having a predominantly regulatory basis.

Sugar-consuming birds face unique osmoregulatory challenges when processing electrolyte-poor nectar of varying sugar concentrations [71,72]. These challenges likely drove adaptations in the interplay between insulin signaling and the renin-angiotensin-aldosterone system (RAAS). These adaptations could enable rapid shifts between water conservation and excretion, while maintaining stable blood pressure and osmotic balance. Consistent with this hypothesis, we found genes associated with the regulation of blood pressure to be enriched in the PhyloAcc analysis across three sugar-consuming clades (Fig. 4B-D). In these species, we observed convergent acceleration of CNEEs in several components of this pathway. For instance, IGF1 can bind to both its primary receptor, IGF1R, and the insulin receptor (INSR), triggering a cascade involving PI3-kinase and insulin receptor substrates (IRS), ultimately activating serum and glucocorticoid-regulated kinase 1 (SGK1) [68,73]. Notably, we found accelerated CNEEs associated with *IGF1, IGF2*, *IRS1*, and *SGK1* in all four sugar-consuming lineages. SGK1 plays a crucial role in modulating

the activity of the epithelial sodium channel (ENaC), which is essential for sodium reabsorption and fluid balance—processes particularly important for birds consuming large volumes of dilute nectar [74]. The RAAS system is further regulated by aldosterone, which stimulates ENaC activity through receptors such as *NR3C2* (mineralocorticoid receptor) and, to a lesser extent, *NR3C1* (glucocorticoid receptor). Both of these receptors exhibit accelerated evolution in their associated CNEEs in sugar-consuming species. This stimulation increases the expression of targets including *SGK1*, which then inhibits *NEDD4L*, an ENaC inhibitor.

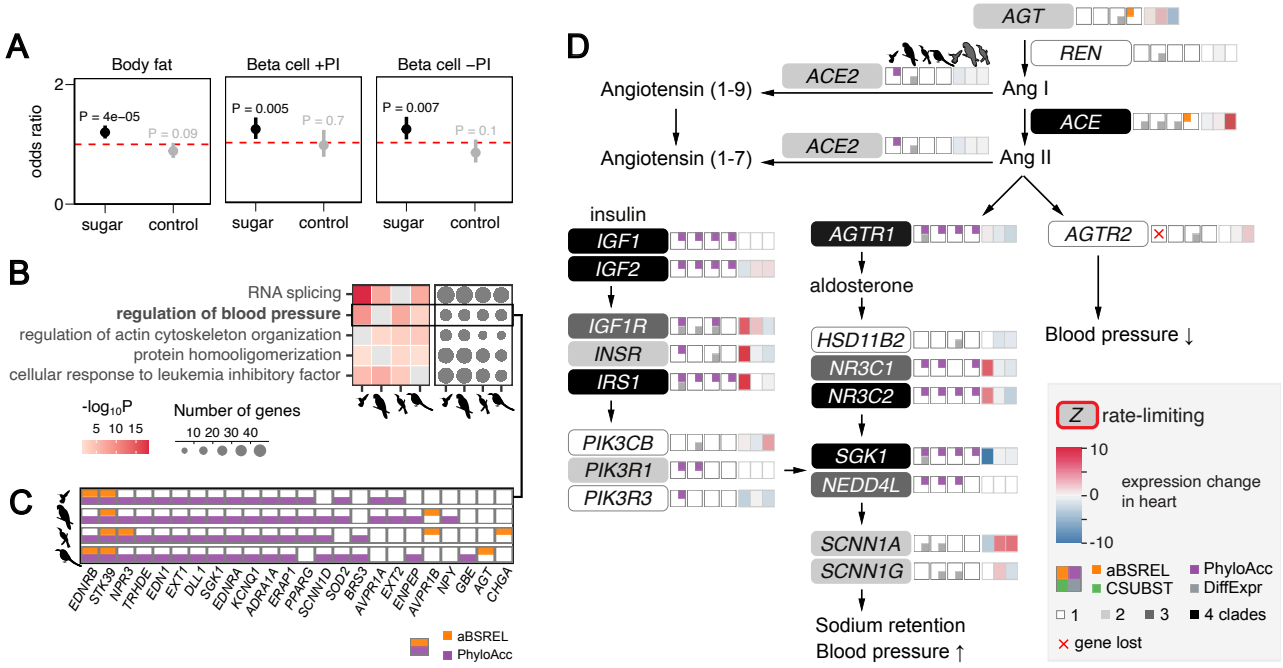



Figure 4. Summary of integrated results from different methods in the renin-angiotensin-aldosterone system (RAAS), insulin signaling, and type 2 diabetes genes.

(A) Two-sided Fisher's exact tests show enrichments of sugar-accelerated CNEEs in T2D loci clusters associated with body fat and beta cell deficiency with high (+) or low (-) proinsulin. Significant differences are indicated in black.
(B) GO term enrichment analysis for genes associated with sugar-accelerated CNEEs. The heatmap displays FDR-adjusted binomial p-values (-log10P) for terms enriched in at least three sugar-consuming clades. Only GO terms with significant results in both binomial and hypergeometric tests (FDR-adjusted p-values < 0.05) are shown. Enriched terms include the regulation of blood pressure. The full list of enrichments is in table S8.
(C) The panel displays sugar-accelerated CNEEs (purple) and minimal positive selection signals (orange) for the regulation of blood pressure, demonstrating strong convergent signals in regulatory elements rather than in protein-coding genes.
(D) Integrating results from aBSREL, PhyloAcc, CSUBST, and differential expression analyses (DiffExpr) highlight changes in genes involved in renin-angiotensin-aldosterone system. Schematics as in Fig. 3D.

## 6. Large effects of regulatory evolution in lipid and amino acid metabolism.

Genes with upregulated expression in sugar-feeders show enrichments in a wide range of processes associated with lipid metabolism across all tested tissues (Fig. 5A, fig. S2E, table S13), although only fatty acid metabolism in the liver shows convergent enrichment in all three sugar-consuming clades (Fig. 5A, table S13). In addition, genes upregulated in the liver of sugar-consuming species are enriched for processes associated with sugar metabolism (pyruvate, monosaccharide, and carbohydrate metabolic processes) and the regulation of hormone secretion (hummingbirds and honeyeaters) (fig. S2E, table S13). Among downregulated genes, we observed a consistent enrichment of genes involved in lipid metabolic processes for all tested tissues in all sugar-feeders (Fig. 5B, fig. S2F, table S14). This expression shift is substantially driven by the downregulation of genes with roles in fatty acid metabolism in multiple sugar-consuming species, including *ASPG* (hummingbird, lorikeet), *ACSL4* (honeyeater, lorikeet), *DECR2* (hummingbird, lorikeet, honeyeater), *ECHDC1* (hummingbird, lorikeet), *FABP1* (hummingbird, lorikeet, honeyeater), and *HACL1* (hummingbird).

Examining genes involved in lipid metabolism in more detail, we found that most genes involved in fatty acid (FA) synthesis and FA beta-oxidation showed signatures of protein sequence and regulatory evolution in at least one sugar-consuming lineage (Fig. 5C). Notably, we saw a striking upregulation of genes of the FA synthesis in all sugar-consuming clades, while the genes responsible for FA beta-oxidation are downregulated. *ACACA* encodes acetyl-CoA carboxylase, which catalyzes the rate-limiting step in FA synthesis, and it shows convergent upregulation in all sugar-consuming clades (Fig. 5C, fig. S2G) together with CNEE rate shifts in two clades. This illustrates an example of convergent regulatory evolution in the key flux-determining step of the pathway.

Additionally, genes downregulated in pectoralis, liver, and duodenum are enriched in amino acid metabolism, with duodenum showing a convergent 3-way enrichment (Fig. 5B, table S14). Many genes involved in amino acid metabolism belong to the branched chain amino acid (BCAA) pathway; these include: *BCKDHB* (hummingbird, honeyeater), *ASRGL1* (hummingbird, honeyeater, lorikeet), *MCCC2* (hummingbird, lorikeet), *GOT1* (hummingbird, honeyeater), *GOT2* (hummingbird, lorikeet), *ALDH4A1* (hummingbird, lorikeet) (fig. S9).

Highlighting the expression shifts in these two key metabolic pathways, we also found significant downregulation in one or more sugar-consuming species for a number of genes with roles in both BCAA and lipid metabolism, including *ACADSB* (hummingbird, lorikeet, honeyeaters), *ACAD6L* (honeyeater, hummingbird), *ACAD8* (hummingbird), *ACAT1* (hummingbird, lorikeet), *PCCA* (hummingbird, lorikeet), and *PCCB* (hummingbird, lorikeet) (fig. S9). Because flower nectar is poor in protein and fat, sugar-feeders may have to conserve these nutrients. These results suggest that sugar-feeders impose breaks on amino acid and lipid catabolism across all tissues.

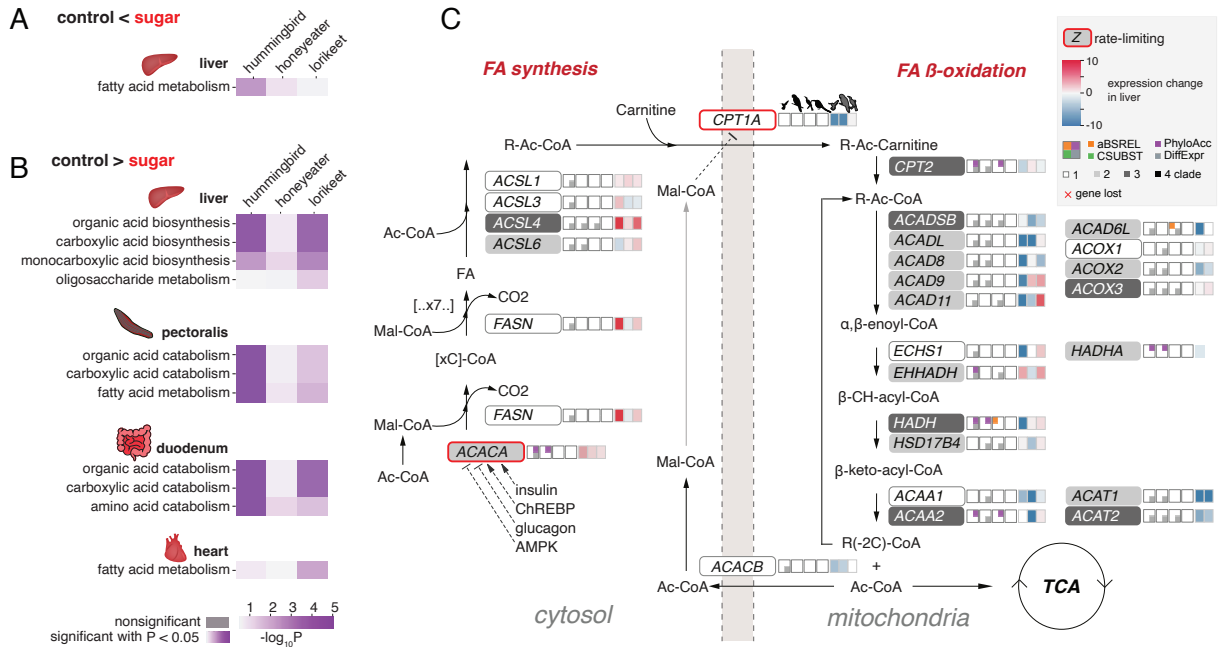



Figure 5. Sugar-feeders show large changes in the regulation of lipid and amino acid metabolism.

(A, B) GO biological processes significantly enriched (p-value<0.05) for (A) up- and (B) downregulated genes in all three sugar-feeders.
(C) Large effects of regulatory evolution in lipid metabolic pathways and contrasting expression signals in fatty acid synthesis and degradation. Schematics as in Fig. 3D.

## 7. Functional changes in sugar-feeders' *MLXIPL* enhance sugar response.

Among the genes that are positively selected in all four sugar-consuming clades, *MLXIPL* is of particular interest: *MLXIPL* encoded ChREBP dimerizes with its interaction partner MLX and binds to DNA, initiating the expression of target genes that promote glucose metabolism and lipogenesis (Fig. 6C). Many substitutions specific to sugar-consuming clades cluster in the putative glucose-response activation conserved element (GRACE domain) [75], which is thought to mediate transcriptional activity in response to glucose [76] (Fig. 6A). Furthermore, *MLXIPL* as well as downstream target genes associated with *de novo* lipogenesis, glycogen synthesis, and glucose and fructose metabolism (Fig. 6E and fig. S10B) are upregulated in all three sugar-consuming clades relative to non-sugar outgroups. In hummingbirds, *MLXIPL* expression in the duodenum was the highest among all tissues analyzed (Fig. 6B), possibly indicating the greater importance of intestinal compared to hepatic *MLXIPL* expression for fructose tolerance [77–80]. Hummingbirds also show exceptionally high hepatic expression of lipogenic target genes of ChREBP, which are associated with a dramatic accumulation of lipids in their livers (Fig. 6D).

To test the functional consequences of the sequence changes we detected in sugar-feeders, we synthesized hummingbird and swift *MLXIPL* together with its corresponding interaction partner

*MLX*. We expressed the vectors ectopically in human embryonic kidney cells (HEK293) (Fig. 6F) and used carbohydrate response element (ChoRE)-driven luciferase reporter assay [81] to test two predominantly expressed *MLXIPL* isoforms (fig. S10C,D). We found that for both birds, the long isoform drives a much higher ChoRE response compared to the short one, suggesting the long isoform is the main functional isoform in birds (Fig. 6G, fig. S10E,I). Notably, we found that hummingbird ChREBP shows a higher ChoRE response compared to the swift protein, at both low and high concentrations of glucose (Fig. 6G, fig. S10I), a difference that was also reflected in the higher capacity of the hummingbirds' protein to drive expression of the ChREBP target genes *TXNIP* [82] and *RGS16* [83] (Fig. 6H,I, fig. S10J,K). Together, these results show that hummingbird MLXIPL drives a higher response to glucose compared to the swift protein.

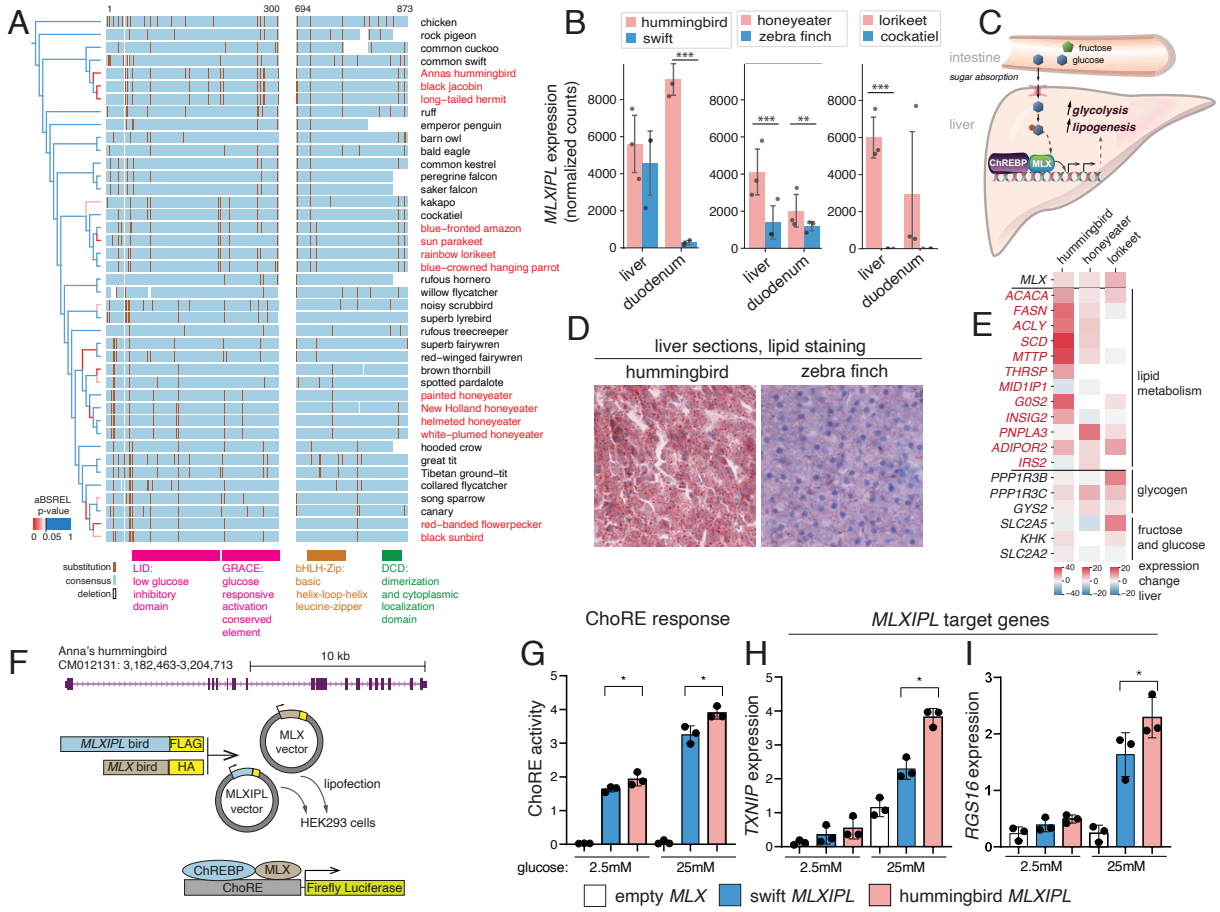



Figure 6. Positive selection and functional analysis of *MLXIPL*

(A) Parts of MLXIPL protein alignment overlapping known functional domains. Substitutions are marked with brown lines. Branches in the phylogeny are colored according to the aBSREL p-value.

(B) *MLXIPL* expression in liver and duodenum of sugar-consuming species compared to their corresponding outgroups.

(C) the role of MLXIPL-encoded carbohydrate response binding protein (ChREBP) in intestines and liver.

(D) Oil-red-O lipid staining of liver sections.

(E) Expression changes in MLXIPL target genes involved in lipid and sugar metabolism.

(F) Vector design and reporter assay to assess transcriptional activity of the recombinant MLXIPL proteins.
(G) Transcriptional activity of the long MLXIPL isoforms in the luciferase reporter assay with the hummingbird and swift MLXIPL-MLX recombinant protein pairs.
(H, I) Expression changes in two target genes of MLXIPL.

# Discussion

Using the repeated evolution of sugar-rich diets as a system, we addressed two fundamental questions: (i) what is the contribution of convergent vs. lineage-specific molecular changes to convergent adaptations and (ii) what are genomic adaptations to sugar-rich diets in four avian lineages. Our comprehensive analyses integrating genomic screens for multiple evolutionary signatures revealed that adaptations to these specialized diets involve convergent evolution in both regulation (detected with PhyloAcc and DESeq2) and proteins (detected with aBSREL and CSUBST). Specifically, in comparison to non-sugar-consuming control clades, we found significantly more convergent positive gene selection and convergent accelerated evolution in regulatory regions (CNEEs) in multiple sugar-consuming clades.

Our work demonstrates the value of studying multiple independent trait origins, as we are able to identify convergent evidence for selection across multiple clades for candidate genes previously studied for relevance in a single clade. For example, our study revealed that the main fructose transporter *SLC2A5* and key genes of the fatty acid biosynthesis (*ACACA* and *FASN*) not only have protein sequence and expression changes in hummingbirds [12,13], but also in honeyeaters, sunbirds, and sugar-feeding parrots. Furthermore, molecular convergence can occur over even larger evolutionary distances, as sugar-consuming bats also show sequence and regulatory changes in a key glycolytic gene *ALDOB* [38,84] as well as regulatory changes in gluconeogenic genes *G6PC* and *PCK1* [38].

These convergent genomic changes show considerable functional coherence, and are associated primarily with energy homeostasis, metabolism, cardiac function and blood pressure regulation, and insulin signaling, pointing to a common molecular toolkit required to adapt to high-sugar diets. However, there are specific differences among pathways in the extent of changes in protein sequences or regulation. For example, whereas carbohydrate metabolism, heart function and blood pressure regulation and hormonal regulation demonstrate evidence for both protein and regulatory evolution, lipid and protein metabolism predominantly showed evidence for regulatory evolution (Fig. 7). This distinction indicates that for some biological processes, selection may favor protein-coding changes that likely affect gene function whereas for other pathways, regulatory changes that likely affect gene expression are preferred.

In addition to molecular convergence, we found that lineage-specific changes contribute to sugar-diet adaptations in all four sugar-consuming clades. Interestingly, these are largely gene-level elaborations of the same pathways we detected with convergent signals of selection, highlighting the common toolkit needed for a specialized, high-sugar diet. A striking example of a lineage-

specific change is hexokinase 3, a rate-limiting glycolytic enzyme, that was targeted by both protein and regulatory evolution in honeyeaters. Our functional assays demonstrated a shift in catalytic function in the honeyeater enzyme. Analogous to “many roads lead to Rome”, the observed lineage-specific changes in the same pathways suggest that adaptations in physiological and metabolic processes can often be realized by different genes within broad functional networks.

However, many of the most compelling adaptations we observe are convergently shared among all four clades. A particularly important example is convergent positive selection in *MLXIPL,* encoding ChREBP, the master transcription factor involved in energy homeostasis regulation. Through functional assays, we demonstrated how sequence and expression changes in *MLXIPL* could contribute to dramatic shifts in energy homeostasis in sugar-feeders. We hypothesize that *MLXIPL* may be the critical trigger in the upregulation of lipid synthesis, as it activates the rate-limiting enzyme of the pathway and multiple other genes involved in lipogenesis. The observed convergent upregulation of lipogenesis and downregulation of fatty acid catabolism likely reflects the energetic and substrate preferences in avian sugar-feeders: during feeding, birds convert glucose into fat rather than heavier glycogen, and during fasting, they use stored fat, upregulating lipid degradation pathways [85,86].

Overall, our results highlight the power of comparative studies, and suggest that adaptations to high-sugar diets in birds require modifications to a constrained set of organismal functions, including pathways directly involved in metabolism and hormonal regulation, as well as pathways such as heart function or the blood pressure regulation that are critical to adapt to the physiological consequences of high sugar consumption.

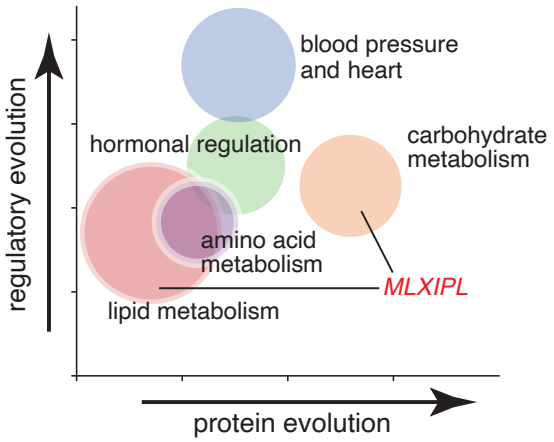



Figure 7. Main detected biological processes in the protein vs regulatory evolution space.

For each of the main biological processes we identified as affected in sugar-feeders, we calculated regulatory (signal in CNEEs and gene expression) and protein (signal in protein-coding genes) evolution scores (see Methods for details). The center of each circle represents the coordinate in the space and the size represents the size of the biological process (total number of genes annotated for this biological process).

# Methods

## Sample collection

**Bird specimens for genome and transcriptome sequencing**

Genomes were sequenced from tissue or blood samples from 9 species (black jacobin, common swift, rainbow lorikeet, blue-crowned hanging parrot, cockatiel, New Holland honeyeater, white-plumed honeyeater, brown thornbill, and rufous hornero) representing three of the four focal sugar-consuming groups as well as outgroup lineages. Transcriptomes from species in three sugar-consuming groups (Anna's hummingbird, rainbow lorikeet, New Holland honeyeater) and three outgroup species (common swift, cockatiel, zebra finch) were sequenced from a replicate panel of tissue samples. Details of specimens used for each target clade are described below.

*Hummingbirds and swifts*

For hummingbirds and their closest relatives, the swifts, we sequenced the genome of a black jacobin (*Florisuga fusca*) using a liver tissue sample collected in Santa Teresa, Espirito Santo, Brazil with the permission of the Brazilian Institute of Environment and Renewable Natural (IBAMA) under the Biodiversity Information and Authorization System (SISBio) permit numbers 30319-1, 41794-2 and 49097-1 (Genetic Heritage #A8E6064). A series of tissues for transcriptomic analysis was collected from three male Anna's hummingbirds (*Calypte anna*) collected in Vancouver, British Columbia following permits obtained from Environment and Climate Change Canada under Scientific Permit SC-BC-2021-0001-01. Tissue samples from four common swifts (one individual was used for genome sequencing, three for transcriptomics) were obtained from terminally-injured individuals in the care of a rehabilitator in Germany that were euthanized at a veterinary clinic (following the regulations of the European Union and the German Animal Welfare Regulation Governing Experimental Animals (TierSchVersV) and the guidelines of the Government of Upper Bavaria).

*Parrots*

For the parrot clade, DNA was obtained from a blood sample of a captive rainbow lorikeet (*Trichoglossus moluccanus*) from the Loro Parque zoo in Tenerife and used for genome sequencing. Tissue samples for RNA-seq collected from four captive-bred rainbow lorikeet individuals from Sydney, Australia, as well as a tissue sample from a captive-bred cockatiel (*Nymphicus hollandicus*) used for genome sequencing were collected following procedures approved by the Animal Welfare Committee (AWC) at Deakin University Australia and following permits from the Department of Sustainability and Environment of Victoria, Australia; tissues from each individual are accessioned into the Museum of Comparative Zoology at Harvard University (MCZ Ornithology # 364103, 364104, 364108, 364107). An additional lorikeet genome (not part of the main analysis, but generated using the same procedure as the thornbill and New Holland honeyeater genomes and uploaded on NCBI) was generated from the individual accessioned under MCZ 364103; a tissue sample from this individual was also used for Hi-C scaffolding for the primary lorikeet genome Cockatiel tissues for transcriptomic analysis were obtained from three individuals bred in captivity in Germany (following the regulations of the European Union and the German Animal Welfare Regulation Governing Experimental Animals (TierSchVersV)). DNA extracted from a blood sample from a blue-crowned hanging parrot (*Loriculus galgulus*)

obtained from the Schoenbrunn Zoo in Vienna and kept in aviaries in Seewiesen, Germany was used for genome sequencing.

*Honeyeaters and other passerines*

For the honeyeater clade, tissue samples from two honeyeater species and a thornbill were collected following approval by the Animal Welfare Committee (AWC) at Deakin University Australia, and according to permits from the Departments of Sustainability and Environment of Victoria and New South Wales, Australia. A tissue sample of a white-plumed honeyeater, *Ptilotula penicillata* was collected at Fowlers, Gap, Australia (accessioned under MCZ 364075). A brown thornbill (*Acanthiza pusilla*; tissue accessioned under MCZ 364070) as well as three New Holland honeyeaters (*Phylidonyris novaehollandia*e) were collected in Geelong, Victoria: an individual with material accessioned under MCZ 364095 was used for genome and transcriptomes; two other individuals (accessioned under MCZ 364096 and MCZ 364097) were used for transcriptomes. Tissue samples from three zebra finches were collected from animals bred in captivity in Seewiesen, Germany, according to the regulations of the European Union and the German Animal Welfare Regulation Governing Experimental Animals (TierSchVersV) and used for transcriptomics. A blood sample from a wild-caught rufous hornero (*Furnarius rufus*) was collected following permits from the ethical committee of animal use (CUEA) of the Faculty of Sciences in the Universidad de la República (Nr. 186) and the Ministry of Environment of Uruguay (Nr. 138/6) and used for genome sequencing.

**Defining analysis groups and species selection for whole genome analysis**

Four focal sugar-consuming groups were examined, called here 'core sugar clades', which are composed of species with a diet consisting of at least 50 percent nectar and fruit combined, using information from the EltonTraits database [16]. The first clade, hummingbirds, represents a clear instance of a single origin of nectarivory, including here the early-branching lineages of jacobins and hermits (represented by our recently-sequenced genome of a long-tailed hermit [11] as well as the well-studied and distantly related Anna's hummingbird. The exact origin of nectarivory in the other clades is less clear-cut, motivating us to include species with high levels of fruit as well as nectar consumption. In parrots, multiple origins of nectarivory likely exist (including in lorikeets, *Loriculus* hanging parrots, New World parakeets, swift parrots, as well as kākās); in addition, many species also consume large amounts of sugar-rich fruits. We focused on rainbow lorikeets, an extreme nectarivore (like many of the closely-related lories), and included a hanging parrot representative (*Loriculus galgulus*) as well as the highly frugivorous sun parakeet and blue-crowned amazon in our focal clade. For honeyeaters, we sequenced two species of honeyeater (New Holland and white-plumed), and included the published genomes of the painted honeyeater and the helmeted honeyeater [87]. A genome of the brown thornbill, a member of the sister group to honeyeaters (including thornbills and pardalotes) was also sequenced; as many species in this group are largely insectivorous, but also take nectar and lerp, this lineage was classified as 'ambiguous'. This category also used for other species close to but outside of the primary in-group sets whose diet contained fruit or nectar (but less than 50%), and for which it was unclear if some adaptations for sugar-feeding could have already have evolved, given either their phylogenetic position and diet or other information (such as data on taste receptor function). Lastly, the

published genomes of the red-banded flowerpecker and black sunbird comprised the fourth focal clade, the 'sunbird' clade.

Outgroup sets ('core non-sugar') were selected from closely related clades where nectar and fruit are not a part of the diet: two swifts (for the hummingbird clade), three falcons (for the parrot clade), a lyrebird and scrubbird (Menuroidea; here called 'lyre- and scrubbirds') (the outgroup clade for honeyeaters), and zebra finch, Tibetan ground tit, and collared flycatcher (here called 'control Passerida') for the sunbird clade. Lineages not classified as core sugar, non-sugar-consuming, or ambiguous, were considered as background species for the genomic analyses. Background lineages were selected to ensure balanced representation across the bird phylogeny and were chosen on the basis of having a BUSCO gene annotation completeness of at least 80%. Outgroups for transcriptomic analyses differed slightly from genomic analyses because of tissue availability: as an outgroup for rainbow lorikeets, the cockatiel, an entirely granivorous member of the non-sugar-consuming Cacatuidae family of parrots (and part of the 'ambiguous' set for the genomic analyses) was selected. As a note, as sugar-consuming behavior has been observed in some *Nestor* parrots (e.g., kākā, *N. meridionalis*) and cacatuids as well as kakapos, kākās and keas take fruit [16,88], nectar and fruit-taking behavior may have existed in the ancestor of all parrots prior to the divergence of Cacatuidae and Psittacidae. For honeyeaters, the distantly-related granivorous zebra finch (part of the control non-sugar-consuming Passerida set for genomic analyses) was used as an outgroup.

## Genome sequencing

***Acanthiza pusilla, Phylidonyris novaehollandiae***

Genomic DNA was isolated from tissue samples collected as described above and accessioned in the Museum of Comparative Zoology Ornithology collection from a brown thornbill *Acanthiza pusilla* (catalog number: 364070), and a New Holland honeyeater, *Phylidonyris novaehollandiae* (364095) DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. We measured DNA concentration with a Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, USA). We performed whole-genome library preparation, sequencing, and assembly (following [89,90]). In brief, a DNA fragment library of 220-bp insert size was prepared using the PrepX ILM 32i DNA Library Kit (Takara), and mate-pair libraries of 3-kb insert size were prepared using the Nextera Mate Pair Sample Preparation Kit (cat. No. FC-132-1001, Illumina). We performed DNA shearing for the fragment and mate-pair library preparations using Covaris S220. We used the 0.75% agarose cassette in the Pippin Prep (Sage Science) for size selection of the mate-pair library (target size 3 kb, "tight" mode). We then assessed fragment and mate-pair library qualities using the High Sensitivity D1000 ScreenTape for the Tapestation (Agilent) and High Sensitivity DNA Kit for the Bioanalyzer (Agilent), respectively, and quantified the libraries with qPCR (KAPA Library Quantification Kit) prior to sequencing. We sequenced the libraries on an Illumina HiSeq 2500 instrument (High Output 250 kit, PE 125-bp reads) at the Bauer Core facility at Harvard University.

***Florisuga fusca***

Genomic DNA (gDNA) was isolated from a female black Jacobin's liver using the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany), as per manufacturer instructions. The specimen was captured and euthanized according to Monte et al., 2023 [91], under the capture permits 41794-1 and genetic patrimony register number AF0CC37, licensed by Ministério do Meio Ambiente, Brazil. Whole-genome shotgun libraries were prepared using Illumina's TrueSeq 2 kit, creating paired-end libraries with 500 and 800 bp fragment sizes, as well as a 3500 bp Nextera mate pair library. Sequencing was executed on an Illumina HiSeq2500 system, and mate pair reads were processed using NEXTCLIP [92].

***Furnarius rufus* and *Trichoglossus moluccanus***

High-molecular-weight genomic DNA was extracted using the MagAttract HMW DNA Kit (Qiagen) following the manufacturer's protocol. For *Furnarius rufus*, genomic DNA was isolated from blood preserved in Queen's buffer, while for *Trichoglossus moluccanus*, a blood sample tissue preserved in ethanol was used. For both species, 10X Genomics libraries were prepared using the Chromium Genome Kit (v2 chemistry) and sequenced on an Illumina NovaSeq 6000 platform (151 bp paired-end reads with 8 bp dual index). Additionally, for *T. moluccanus*, a Hi-C library was constructed (using a tissue sample from another lorikeet individual, accessioned into the MCZ (364103)) using the Dovetail Hi-C Kit (cat# 21004, Manual v.1.01 2-01-18). For this, 33mg of flash-frozen powdered muscle tissue were used. The resulting barcoded library was sequenced using the same Illumina platform and configuration to ca. 50 M reads.

***Ptilotula penicillata, Nymphicus hollandicus, Loriculus galgulus, Apus apus***

**Extraction of ultra-long genomic DNA**

High molecular weight (HMW) genomic DNA (gDNA) from snap-frozen muscle tissues of one common swift (*Apus apus*) and one white-plumed honeyeater (*Ptilotula penicillata*) individual been extracted following the Bionano Prep Animal Tissue DNA Isolation Fibrous Tissue Protocol (Document Number 30071, document revision C). Briefly, 30-40 mg of muscle tissue was cut into 3 mm pieces and fixed with 2% formaldehyde and Homogenization Buffer provided by the kit. The mildly fixed tissue was homogenized with a Qiagen Tissueruptor and the homogenized tissue was embedded in 2% agarose plugs. The plugs were treated further with Proteinase K and RNase A, and then washed with 1X Wash Buffer from the kit and home-made 1X TE Buffer (pH 8.0). DNA was recovered from plugs and purified finally by drop dialysis against TE Buffer (pH 8.0).

Using Bionano SP Blood and Cell DNA Isolation Kit (PN 80030, Bionano Genomics Inc, *USA*), HMW gDNA from blood samples preserved in ethanol of one blue-crowned hanging parrot (*Loriculus galgulus*) was isolated applying the modified Bionano Prep SP Frozen Human Blood DNA Isolation Protocol (Document Number 30246, document revision C). In brief, the ethanol supernatant was firstly removed and the collected blood pellet was washed 3 times in 1 ml of 1 x PBS buffer followed by resuspending the pellet in cell buffer. After cell lysis with Proteinase K, the gDNA was precipitated by isopropanol and attached to the Nanobind disk. Several times of washing steps were performed and the purified gDNA was finally eluted in the Elution Buffer.

gDNA of one cockatiel (*Nymphicus hollandicus*) was extracted from muscle tissue that was preserved in ethanol using the Circulomics Nanobind Tissue Big DNA kit (Circulomics Inc.,

PacBio, USA) according to the manufacturer's instructions. For initial ethanol removal, the tissue was washed three times for 30 min each in 45 ml of EtOH removal buffer (40nM NaCl, 20nM Tris (pH7.5) and 30mM EDTA).

gDNA concentration was determined using the Qubit dsDNA BR (Broad Range) assay kit (ThermoFisher, USA). The purity of the sample was measured with the NanoDrop, assessing curve shape and the 260/280 nm and 260/230 nm values. The integrity of the HMW gDNA was determined by pulse field gel electrophoresis using the Pippin PulseTM device (SAGE Science). The majority of the long and ultralong gDNA was between 50 and 500 kb in length. All pipetting steps of ultra-long and long gDNA have been done very carefully with wide-bore pipette tips.

**PacBio continuous long read (CLR) library preparation and sequencing**

For species *Apus apus, Ptilotula penicillata*, and *Loriculus galgulus,* **l**ong insert libraries have been prepared as recommended by Pacific Biosciences (PacBio) according to the 'Guidelines for preparing size-selected >30 kb SMRTbell templates making use of the SMRTbell express Template kit 2.0. Briefly, ultra-long gDNA was sheared to 75 kb fragments with the MegaRuptor device (Diagenode) and 4-8 µg sheared gDNA was used for library preparation. The PacBio SMRTbell library was size-selected for fragments larger than 26 kb, 26kb, 20kb and 30kb with the BluePippinTM device, respectively, according to the manufacturer's instructions. All size selected libraries were sequenced for 15 hours on the Pacbio Sequel® II on 8M SMRT™ cells using the Sequel® II Binding 2.0 and the Sequel® II Sequencing 2.0 chemistry.

**PacBio High Fidelity (HiFi/CCS) library preparation and sequencing**

For species *Nymphicus hollandicus*, the gDNA entered library preparation using a PacBio HiFi library preparation protocol "Preparing HiFi Libraries from Low DNA Input Using SMRTbell Express Template Prep Kit 2.0". Briefly, gDNA was sheared to 8 kb with the MegaRuptor device (Diagenode) and 2.6 µg sheared gDNA was used for library preparation. With Ampure beads size-selection, fragments shorter than 3 kb were removed. The size-selected library was prepared for loading following the instructions generated by the SMRT Link software (PacBio, version 10) and the 'HiFi Reads' application. The HiFi library was sequenced on 8M SMRT™ Cells for 30 hours with the Sequel® II Binding 2.2 and the Sequel® II Sequencing 2.0 chemistry on the Sequel® II.

**10x genomic linked read sequencing**

Ultra-long gDNA from species *Apus apus, Ptilotula penicillata*, and *Loriculus galgulus* was used for 10x genomic linked read sequencing following the manufacturer's instructions (10X genomics Chromium Reagent Kit v2, revision B). Briefly, 1 ng of ultralong gDNA was loaded into 10X genome GEM droplets (Gel Bead-In-EMulsions) making use of the Chromium device. gDNA molecules were amplified in these individual GEMS in an isothermal incubation using primers that contain a specific 16 bp 10x barcode and the Illumina R1 sequence. After breaking the emulsions, pooled amplified barcoded fragments were purified, enriched and went into Illumina sequencing library preparation, as described in the protocol. Sequencing has been done on a NovaSeq 6000 S1 flow cell using the 2x150 cycles paired-end regime plus 8 cycles of i7 index.

**Bionano optical mapping**

For species *Apus apus, Ptilotula penicillata*, and *Loriculus galgulus,* optical mapping was done with the Bionano Prep Direct Label and Stain DLS DNA Kit (catalog #8005, Bionano Genomics, San Diego), according to the manufacturer's Protocol (Document Number 30206, Document Revision F). Briefly, 750 ng of ultra-long gDNA was fluorescently labeled at defined sequences, making use of the nicking-free Bionano Direct Label Enzyme (DLE-1). For further visualization, the DLE-1 labeled gDNA backbone was stained with DL-Green. Labeled molecules were imaged using the Bionano Saphyr system. Data were generated from one Bionano flow cell with a total yield of 341, 997, 215 and 320 Gb, respectively, in molecules larger than 150 kb.

**Chromatin conformation capturing (HiC)**

Chromatin conformation capturing was done using the Arima-HiC kit (Article Nr. A410109) for *Ptilotula penicillata* and ARIMA-HiC High Coverage Kit (Article Nr. A101030-ARI) for *Apus apus,* and *Loriculus galgulus*, following the Arima documents (User Guide for Animal Tissues, Part Number A160132 V01 and Part Number A160162 v00; Guide to Preparing Crosslinked Nucleated Blood). Briefly, 66 mg flash-frozen powdered muscle tissue from *Ptilotula penicillata*, 46 mg muscle tissue in ethanol from *Nymphicus hollandicus*, 50 µl pure blood sample from *Apus apus* and 50 µl blood preserved in ethanol from *Loriculus galgulus* were crosslinked chemically. All the tissue and blood samples that are stored in ethanol were firstly treated with ethanol removal steps (3 times washing in 1 ml of 1 x PBS buffer) before crosslinking procedure. The crosslinked genomic DNA was digested with a cocktail of two (*Ptilotula penicillata)* or four (*Apus apus, Loriculus galgulus)* restriction enzymes. The 5'-overhangs were filled in and labelled with biotin. Spatially proximal digested DNA ends were ligated, and the ligated biotin containing fragments were enriched and went for Illumina library preparation, following the ARIMA user guide for Library preparation using the Kapa Hyper Prep kit (ARIMA Document Part Number A160139 v00). The barcoded HiC libraries ran on an S4 flow cell of an Illumina NovaSeq 6000 with 2x150 cycles.

## Genome assembly

***Apus apus*, *Ptilotula penicillata, Nymphicus hollandicus,* and *Loriculus galgulus.***

The genomes for *Apus apus, Ptilotula penicillata* and *Loriculus galgulus* were assembled using PacBio Continuous Long Reads (CLR), 10X Genomics linked-reads, Bionano DLS optical maps and Arima Chromatin Conformation Capture (HiC) libraries as follows. Initial contigs were assembled from PacBio subreads using Falcon (v1.8.1) and Falcon Unzip (v1.3.7) [93]. Remaining haplotypic duplications in the primary contig set were then removed using purge-dups (v.0.0.3) [94]. The genome of *Nymphicus hollandicus* was assembled using PacBio HiFidelity (HiFi) Circular Consensus Sequences (CCS) and Arima HiC libraries.

Scaffolding was performed using 10x genomics linked Illumina reads, Bionano optical maps, and HiC chromosome conformation capture Illumina read pairs. To this end, 10x reads were first mapped to the contigs using Long Ranger (v2.2.2) and scaffolded using Scaff10X (v4.2). The 10x linked-reads were then re-mapped to the scaffolded assembly produced by Scaff10X and any false joins were removed using the Break10X command from Scaff10X. Next, we used optical maps from Bionano DLE1-labelled DNA molecules. The Bionano assembly was produced from optical-mapped reads using Bionano Solve Assembly (v3.5.1) with parameters "non-haplotype" and "no-Extend-and-Split". The initial scaffolds from Scaff10X were then further scaffolded using

Bionano Solve Hybrid Scaffold (v3.5.1). HiC reads were then mapped to the resulting scaffolds using bwa-mem (v0.7.17-r1188) [95,96]. We then followed the VGP scaffolding pipeline (https://github.com/VGP/vgp-assembly/tree/master/pipeline/salsa) [20], which uses salsa2 (v2.2) [97]. Finally, we manually curated the scaffolds to join those contigs missed by salsa2 and break those joins which were spuriously created. In order to visually evaluate the scaffolding, reads were again mapped to the scaffolded assembly using bwa-mem, filtered and deduplicated using pairtools (v0.3.0) and the multi-cooler file for higlass (v0.4.17) was produced using cooler (v0.8.11). We then curated the assembly through an iterative process of joining together scaffolds which were missed by the automated scaffolding software and breaking any false joins between two chromosomes until no further corrections could be made to the assembly.

After scaffolding, we closed assembly gaps using the PacBio CLR read data. To this end, we mapped the original subreads.bam files to the scaffolded assembly using pbmm2 (version 1.3.0) with arguments "--preset SUBREAD -N 1". For the assembly of *Apus apus*, we then used Dentist (v1.0.0) [98]. For the assemblies of *Ptilotula penicillata* and *Loriculus galgulus* we used gccp (v2.0.2) to close gaps. Based on the read-piles created by reads spanning across gap regions, we can create a consensus sequence to replace the N sequences in our genome. We used gcpp to polish the gap regions and their 2 kb upstream and downstream flanks. We then replaced the assembly gap and its flanking region with those regions that were polished with high-confidence by arrow (no N's remaining in the polished sequence and no lower-case a/c/g/t).

To increase base level accuracy, we polished the resulting assembly using the 10X linked-reads as done in the VGP pipeline (https://github.com/VGP/vgp-assembly/tree/master/pipeline/freebayes-polish) [20]. Briefly, 10X reads were mapped to the genome using Long Ranger (v2.2.2) and variants were called using freebayes (v1.3.2) with argument "-g 600" to ignore regions with coverage over 600X. The detected variants were filtered using bcftools (v1.12-21-ga865a16) for variants with quality score greater than 1 and genotype of homozygous alt (AA) or heterozygous (Aa) using the command bcftools view -i 'QUAL>1 && (GT="AA" || GT="Aa")'. The consensus was then called using the command bcftools consensus -i'QUAL>1 && (GT="AA" || GT="Aa")' -Hla, which takes the longest allele in heterozygous cases. This procedure was performed twice. Finally, we used Merqury (v1.0) [99] with the 10x data. For the assembly of *Loriculus galgulus* we performed further polishing into two haplogroups. Chromosomes were phased by applying an adapted version of the DipAsm pipeline [100], calling heterozygous sites using 10x reads and creating phase blocks by linking heterozygous sites in 10x, Hi-C and PacBio reads using Hapcut2 (git commit 1ee1d58) and whatshap (v1.6). The whatshap-tagged PacBio and 10x reads, binned into H0 and H1, and H0 and H2, were then used to haplotype-polish the chromosomes. Arrow was run using the combined H0 and H1 haplotagged PacBio CLR reads, followed by freebayes polishing using the combined H0 and H1 haplotagged 10x reads as above. Correspondingly the H2 and H0 tagged reads were used for the second haplotype.

***Florisuga fusca***

The genome of *Florisuga fusca* was assembled from illumina short reads (see above) using a hybrid approach similar to [101]. Initial assembly was conducted with IDBA-UD [102], followed by further

assembly using CELERA ASSEMBLER v8.1 [103], supported by subsets of the short-read data. Scaffolds were ordered using SSPACE2 [104] and aligned to the galGal4 assembly via RAGOUT v1 [105]. Gap filling was performed with GAPCLOSER [106] to enhance contig N50. The data have been deposited with links to BioProject accession number PRJNA622897 in the NCBI BioProject database, genome (https://www.ncbi.nlm.nih.gov/bioproject/), the Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAWXA000000000.

***Acanthiza pusilla* and *Phylidonyris novaehollandiae***

We assessed the quality of the sequencing data generated by the Illumina HiSeq 2500 platform using FastQC [107]. Adapters were removed using Trimmomatic [108]. Genome assembly was performed with AllPaths-LG v52488 [109].

***Furnarius rufus* and *Trichoglossus moluccanus***

De novo genome assembly for both species was performed using the nf-core/neutronstar analysis pipeline implemented in Nextflow [110]. Supernova v2.1.0 [111] was used to construct initial assemblies from Illumina NovaSeq 6000 sequencing data derived from 10X Genomics libraries. For *Trichoglossus moluccanus*, additional scaffolding was performed using the 3D-DNA pipeline [112], leveraging Hi-C data to improve assembly contiguity.

## Annotation

**Modeling and masking repeats**

For the newly generated assemblies, we used RepeatModeler [113] to identify repeat families and RepeatMasker [114] to soft-mask repeats. For other bird genomes, we used the available repeat masker annotations provided by NCBI. All species names and their assemblies are listed in table S1.

**Generating pairwise genome alignments**

To compute pairwise genome alignments between chicken (galGal6 assembly), *T. guttata* (GCF_003957565.2) and *P. major* (GCF_001522545.3) as references and other birds as query species, we used LASTZ version 1.04.00 [115] with parameters K = 2,400, L = 3,000, Y = 9,400, H = 2,000 and the default scoring matrix, axtChain [116], chainCleaner [117], and RepeatFiller [118] (all with default parameters). For the downstream analyses except TOGA, we excluded chains with a score less than 100,000.

**Protein-coding gene annotation**

To achieve maximum completeness of the protein-coding gene annotation the following data were combined: 1) homology-based predictions, 2) transcriptome data, 3) protein data, and 4) *ab initio* gene predictions.

We started with generating pairwise alignments between all target genomes and three reference genomes: *G. gallus* (chicken), *T. guttata* (zebra finch), and *P. major* (great tit). We ran TOGA [23] with default parameters to project transcripts annotated for reference species onto the target genomes. As reference annotations, we used NCBI annotations of *T. guttata* (46,022 transcripts)

and *P. major* (41,530 transcripts). For *G. gallus* we combined NCBI annotation with APPRIS principal isoforms (resulting in 64,081 transcripts).

We prepared protein evidence aligning translated protein data to the target genomes using GenomeThreader (v1.7.0) [119] applying the Bayesian Splice Site Model (BSSM) trained for chicken and default parameters aside from those specified below. For protein alignments, we used parameters: a seed and minimum match length of 20 amino acids (preseedlength 20, prminmatchlen 20) and allowed a Hamming distance of 2 (prhdist 2). For the transcript alignments, we used parameters: a seed length and minimum match length of 32 nucleotides (seedlength 32, minmatchlen 32). At least 70% of the protein or mRNA sequence was required to be covered by the alignment (-gcmincoverage 70), and potential paralogous genes were also computed (-paralogs).

For five species (black jacobin, New Holland honeyeater, rainbow lorikeet, brown thornbill, and rufous hornero) we used transcriptomic data as an additional source of evidence and assembled transcripts with Stringtie (v2.1.2).

Next, we ran *de novo* gene prediction using Augustus (v3.3.3) [120] providing TOGA projections, mapped protein and RNA-seq data as hints. bam2hints module of Augustus with introns-only flag was used to produce RNA-seq hints. Priority 4 was given for RNA-seq hints. With the script align2hints.pl (provided with BRAKER distribution) protein hints from GenomeThreader alignments were prepared. Priority of 4 was given to protein-derived hints. Finally, RNA-seq- and protein-derived hints were merged using the join_mult_hints.pl script from Augustus. Prediction of additional splice sites was enabled with allow_hinted_splice_sites flag (--allow_hinted_splicesites=gcag,atac). Prediction of untranslated regions was disabled (--UTR=off). The resulting files were merged using the Augustus joingenes module. bedtools intersect was used to exclude all Augustus predictions having more than 10% overlap with a repeat region. Next, all predictions were converted to protein and queried with blastp against the Swissprot database with an E-value cut off 1e-10. An in-house script was used to filter blast results allowing only hits matching a sequence in the vertebrates database or having lengths greater than 200 amino acids.

We prepared a set of consensus transcripts running EVidenceModeler (v1.1.1) [121] providing combined aligned protein and RNA-seq data, TOGA projections, and de novo transcripts as input. We assigned weights prioritizing TOGA projections over protein data and RNA-seq data followed by de novo predictions (8, 2, 2, and 1 respectively). Since EVidenceModeler predicts only one consensus gene model per locus, the final set of gene models lacks information about 1) nested genes located in the intronic regions of other genes; 2) alternative isoforms. To rescue those, we added high-confidence gene predictions to the resulting set of consensus models by adding TOGA transcript projections that were identical for at least two out of the three reference species. The final annotation comprised between 33,507 (sun parakeet) and 131,520 (white-plumed honeyeater) transcripts and between 19,321 (sun parakeet) and 44,121 (rainbow lorikeet) genes (table S1).

Where RNA-seq data were available, we used a custom script to extend gene models into UTR regions based on the RNA-seq Stringtie assembly. The donor site of the first intron or the acceptor

site of the last intron of a predicted transcript was required to match perfectly by RNA-seq structure. For single-exon transcripts, the exon start and end were required to lie within the corresponding sequence evidence. If there were conflicts between several Stringtie models, the UTR regions were assigned based on the majority.

For the species with the available NCBI annotation, we downloaded the available annotations.

We assessed the completeness of all annotations (NCBI and ours) using BUSCO pipeline (v5.2.1) [17] with the set of 8338 conserved, single-copy avian genes.

## Positive selection analysis

To identify the one-to-one orthologs, we used TOGA [23] based on chicken gene annotation, keeping only projections classified as intact, partially intact or uncertain loss. Next, we aligned identified orthologs with MACSE_v2 [122] with default parameters, exon by exon, taking into account the codon structure of a gene. We filtered the alignments with HmmCleaner (v0.180750) [123] and then applied additional filtering with an in-house script, removing sequences entirely if less than 50% of the sequence was aligned and removing codon alignment columns if less than 50% of species were aligned. We inspected final alignments manually for potential misalignments. We analyzed positive selection with aBSREL package of HyPhy (v2.5.58) [24,124]: for each transcript we ran four models: 1) the default model; 2) model enabling synonymous rate variation (--srv Yes); 3) model enabling correction for multinucleotide substitution (--multiple-hits Double+Triple); and 4) model enabling both SRV and multihits (--multiple-hits Double+Triple --srv Yes). For every model we enabled tolerating numerical errors (ENV='TOLERATE_NUMERICAL_ERRORS=1). For each transcript, we selected the best model based on AICc (Akaike information criterion corrected) values, selecting a more complex model only if its AICc value was smaller than of a simpler model by at least 2.

Each gene we assigned a ‘relevance ratio’ based on the ratio of number of sugar branches to total number of branches with a significant signal of positive selection (branch-corrected p-value $< 0.05$); branches classified as ambiguous were not included in either counts. Genes having the relevance ratio $> 0.5$ were shortlisted and additionally examined for pathway enrichment.

We performed the enrichment analysis with the R package ClusterProfiler (v4.8.2) [125] using human OrgDb org.Hs.eg.db and Gene Ontology (GO) Biological Processes subontology. To be able to use the human database, we converted chicken gene IDs into human gene IDs with biomart (https://useast.ensembl.org/biomart/martview). The entire genome was considered as the background gene set. We selected significantly enriched terms applying p-value cutoff=0.05. To select terms specific for the sugar-feeders, we excluded the terms that were significant in at least two non-sugar control clades (full list of control GO enrichments is in table S5). Next, we ran a custom script find_representative_go.R to simplify the raw list of GO terms summarizing clusters of semantically similar terms with a representative ancestral term. Finally, we removed redundancy of the raw GO term lists excluding direct children terms of the terms that were already present in the list.

To test gene-level convergence, we used a two-sided Fisher's exact test. For each pair, we compared the ratio of convergent sugar-selected genes to lineage-specific genes with the corresponding ratio calculated for a core non-sugar control group. For example, to test convergence between hummingbirds and sunbirds, we calculated the number of genes convergently under selection in both groups, the number of genes under selection in either hummingbirds or sunbirds, and compared these gene counts to the ones is ratio to the ratio we calculated similarly for swifts and control Passerida.

To identify specific sites under selection in *MLXIPL* and *HK3* genes, we used branch-site model of codeml from the PAML (v4.9) package [126], selecting the ancestral honeyeater branch (for *HK3*) and all sugar-consuming branches (for *MLXIPL*) as the foreground. Sites under positive selection were defined as sites having Bayes Empirical Bayes (BEB) posterior probabilities >0.5.

## Homology modeling and protein structure manipulation

We performed homology modeling using the swissmodel.expasy.org modeling service [127] providing New Holland honeyeater sequence. Since no crystal structure of hexokinase 3 was solved, we used a human hexokinase 2 crystal structure (PDB ID: 5HEX) as a template. The structure visualization and manipulation were done in PyMOL.

## Convergent evolution in protein-coding sequences

We utilized CSUBST v1.0.0 [25] to identify convergent evolution in protein-coding sequences. CSUBST employs the $\omega_c$ metric, a ratio of non-synonymous to synonymous combinatorial substitution rates. This metric is designed to be robust against phylogenetic errors and can discern genuine adaptive convergence from false positives. Unlike the traditional ω measure, $\omega_c$ contrasts the ratio of observed to expected non-synonymous combinatorial substitutions with the synonymous counterpart between distinct phylogenetic branches. For convergence in a pairwise branch comparison, we set criteria: $\omega_c \geq 3$, $O_c^N \geq 3$ (any2spe category).

## CNEE alignments

We obtained vertebrate conserved non-exonic elements (CNEEs) from previously published sources [26–29]. These were then lifted over to Gallus gallus (galGal6 or GRCg6a assembly) coordinates using halLiftover [128].

To ensure the CNEEs met the general expectations for length and conservation, we performed a series of cleaning steps. Initially, we eliminated elements exceeding 5 kb in length and merged elements that overlapped or were within 5 bp of each other. Elements shorter than or equal to 5 bp were also removed. Subsequently, we excluded elements overlapping with any exonic regions in GRCg6a. Lastly, we discarded CNEEs that were not conserved as defined by phyloP [129] using the CON mode and LRT method, with adjusted p values greater than 0.1.

Following the pipeline established in Sackton et al. 2019, we generated CNEE alignments from the curated set of CNEEs. In brief, we extracted sequences for the curated CNEEs from all species included in the whole-genome alignment. We used the halLiftover tool from HAL tools v2.1 to generate psl files, which mapped chicken reference coordinates to each target species. Custom Perl scripts [28] were then used to convert halLiftover output into FASTA format files. Sequences for individual CNEEs were omitted if the target region was duplicated in a given taxon or exceeded twice the reference chicken length.

Individual CNEEs were aligned using MAFFT v7.304 [130] with the 'ginsi' option. The alignments were then trimmed using TrimAl v1.4.rev22 build[2015-05-21] and the '-noallgaps' option, which removes columns composed entirely of gaps. Subsequently, the trimmed fasta files were merged into a single file using catsequences [131], with the provision of a partition file defining the start and end positions of each locus. As a result, we obtained 363,476 CNEE alignments. All species demonstrated a liftover rate exceeding 96% relative to the curated set of CNEEs in the GRCg6a coordinates. We obtained 363,476 CNEE alignments. All species had a liftover rate exceeding 96% relative to the curated set of CNEEs in the GRCg6a coordinates. Furthermore, the majority (74.6%) of the CNEE alignments contained sequences from all species.

## PhyloAcc

We employed PhyloAcc v2.2.0 [30,132], a Bayesian method designed to model shifts in substitution rates within conserved elements across a phylogeny. Our objective was to determine whether a CNEE displays accelerated substitution rates in comparison to a neutral model across each branch of the avian phylogeny. To derive a neutral model, we first extracted fourfold degenerate sites from the entire genome alignment using the PHAST software package [133]. Subsequently, we utilized the 'phyloFit' program [134] to estimate a neutral model from these sites. We performed a phyloAcc analysis specifically for the sugar-consuming birds, designating these species as ‘TARGETSPECIES’. Further program settings are detailed on our GitHub repository (https://github.com/maggieMCKO/nectargenomics/). To identify sugar-specific acceleration, we conducted an additional phyloAcc run (referred to as the control run), setting the core non-sugar-consuming birds (defined above), as ‘TARGETSPECIES’.

## Defining accelerated CNEEs

We used PhyloAcc [30,132] to identify genomic elements exhibiting lineage-specific acceleration. This method relies on sequence alignments across a phylogeny to model substitution rate shifts using Bayes factors. PhyloAcc employs three nested models: 1) the Null model (M0), which assumes no shift in substitution rates; 2) the lineage-specific model (M1), which permits rate shifts in predefined lineages, such as sugar-consuming birds.; and 3) the full model (M2), which allows rate shifts in any branch. To detect CNEEs with lineage-specific accelerations, we compared the marginal likelihoods of these models, defining three Bayes factors: Bayes factor 1 (BF1) = $(P(Y|M_1))/(P(Y|M_0))$ and Bayes factor 2 (BF2) = $(P(Y|M_1))/(P(Y|M_2))$.

To identify CNEEs predominantly accelerated in sugar-consuming lineages, we applied a cutoff of logBF1 > 10 and a 'relevance ratio' > 0.5, which is conceptually similar to the ratio used for positively selected genes. For each CNEE, the relevance ratio was calculated as the number of

sugar branches, including internal branches, in the accelerated state (posterior probability under the full model > 0.9) divided by the number of background branches (excluding ambiguous branches) in the accelerated state. To categorize accelerated CNEEs within specific sugar-consuming or core non-sugar clades, we required at least one branch in each clade to have a posterior probability exceeding 0.9.

## CNEE statistical analysis

To assess CNEE-level convergence, we conducted a two-sided Fisher's exact test to examine if convergently accelerated CNEEs are notably more prevalent in sugar-consuming clades compared to core non-sugar clades, across pairwise, triple, and quadruple combinations. Additionally, for gene-level convergence analysis, we utilized GREAT (see below) to associate CNEEs with genes and performed a two-sided Fisher's exact test.

We employed the GREAT software [31] to associate CNEEs with genes, using its 'createRegulatoryDomains' utility set to the default 'basalPlusExtension' configuration. This configuration establishes a 'regulatory domain' for each protein-coding gene, extending 5 kb upstream and 1 kb downstream from the transcription start site. Furthermore, the regulatory domain of each gene is expanded to the basal regulatory regions of the nearest upstream and downstream genes, with a maximum extension limit of 1 Mb in each direction. This allows for the incorporation of distal binding sites into its analysis, thereby providing a broader perspective on the potential regulatory influences of CNEEs on gene expression.To enhance the annotation of chicken genes (GRCg6a) for the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) by leveraging human databases, we created custom GO and KEGG databases. We used the Biomart library (v.2.56.0, Ensembl ver106) in R to map chicken Entrez gene IDs to their corresponding Ensembl and human Ensembl gene IDs, based on Ensembl's ortholog database. For GO annotation, we employed the Biomart library to retrieve GO IDs for each human gene, thereby increasing the number of chicken genes linked to GO IDs from 2,203 to 7,586. For KEGG pathways, we converted human Ensembl gene IDs to Entrez IDs using the 'org.Hs.eg.db' R library (v.3.17.0) and then used the 'KEGGREST' R library (v.1.40.0) to identify associated KEGG pathways. This expanded the number of chicken genes linked to KEGG pathways from 3,482 to 6,680.

To focus on biologically meaningful enrichments and reduce computational noise, we first filtered our analysis to include only GO terms with at least 40 genes associated with CNEEs. For these filtered GO terms, we then conducted both binomial and hypergeometric tests. We utilized the 'calculateBinomialP' utility in GREAT for the binomial test-based enrichment analysis. Given that CNEEs are not randomly distributed across the genome and tend to cluster in specific loci, we also performed a hypergeometric test to correct for enrichment that may arise solely from the dense distribution of CNEEs. In our final results, we only show significantly enriched GO terms that had at least 3 genes from input gene lists. Additionally, we exclude enriched pathways that are shared by at least 2 control clades. This combined approach of pre-filtering, dual statistical testing, and post-test filtering ensures a rigorous identification of functionally relevant enrichments while accounting for the non-random distribution of CNEEs across the genome.

## Tissue collection and RNA sequencing

Tissues were excised, each piece was cut into aliquots of approximately 100 mg and then transferred into a Dewar with liquid nitrogen for several minutes, or placed on dry ice, or placed in 5-10 volumes of RNAlater (Thermofisher Scientific, AM7020) solution and incubated overnight at 4°C before storage at -80°C.Throughout the whole process, an RNAse-free environment was ensured and all surfaces and instruments were treated with RNase Zap solution (Thermofisher Scientific, AM9780).

Total RNA was isolated using the RNeasy mini kit (Qiagen). The concentration and the RNA integrity number (RIN) were analyzed using the Bioanalyzer (Agilent). For all samples, RINs were above 7. RNA-seq library preparation and the sequencing analysis were performed in the Dresden-Concept Genome Center (DcGC; TU Dresden, Germany). In detail, mRNA was enriched from 400 ng total RNA by poly-dT enrichment using the NEBNext Poly(A) mRNA Magnetic Isolation Module according to the manufacturer's instructions. The polyadenylated RNA fraction was eluted in 11.5ul 2x first-strand cDNA synthesis buffer (NEBnext, NEB). After chemical fragmentation, by incubating for 13 min at 94°C the samples were directly subjected to the workflow for strand-specific RNA-seq library preparation (Ultra II Directional RNA Library Prep, NEB). In short, the first and second strand synthesis followed an end repair and A-tailing of the fragments that were ligated with the universal NEB hairpin loop adapter. After ligation, adapters were depleted by an XP bead purification (Beckman Coulter) adding the beads solution in a ratio of 1:0.9. During the PCR enrichment with 12 cycles, unique dual index primers were incorporated carrying the sequence for i7 and i5 tails (Primer 1: AAT GAT ACG GCG ACC ACC GAG ATC TAC AC XXXXXXXX ACA TCT TTC CCT ACA CGA CGC TCT TCC GAT CT, Primer 2: CAA GCA GAA GAC GGC ATA CGA GAT XXXXXXXX GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T; X represents the different barcode sequences). After two more AMPure XP bead purifications (in a ratio of 1:0.9), libraries were quantified using the Fragment Analyzer (Agilent) and sequenced on an Illumina NovaSeq 6000 with 2x 100 bp reads using an S4 flowcell to an average depth of 30 mio. read pairs.

## Analysis of transcriptomic data

We trimmed adapters and low-quality sequences from the reads using Trimmomatic (v0.39) with parameters: PE ILLUMINACLIP:****:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. We quantified transcript-level expression with Kallisto (v 0.46.1) with default parameters and bootstrap 100 (-b 100). To run Kallisto, we generated species-specific transcriptome indexes based on our generated annotations. We quantified gene-level expression with tximport R package. We ran differential gene expression analysis with DESeq2 (R v4.1.1), specifying contrasts to perform between species pairwise and multifactor cross-tissue analyses. We performed the enrichment analysis with the R package ClusterProfiler (v4.8.2) [125] using human OrgDb 'org.Hs.eg.db' and Gene Ontology (GO) Biological Processes subontology. To be able to use the human database, we converted chicken gene IDs into human gene IDs with biomart (https://useast.ensembl.org/biomart/martview). For the pairwise comparisons, we specified the background set to only list genes having a non-zero expression in at least one of the samples in the tested tissue. For the cross-tissue comparison, we specified the background

set as genes having a non-zero expression in either liver or pectoralis. We selected significantly enriched terms applying p-value cutoff=0.05. Finally, we removed redundancy of the raw GO term lists excluding direct children terms of the terms that were already present in the list.

## Manual curation of hexokinase 3 and MLXIPL gene models

Reference exons from hexokinase 3 (*HK3*), *MLXIPL*, and *MLX* gene predictions were downloaded from NCBI (for HK3: chicken (XM_001231328.5), zebra finch (XM_030283802.2), great tit (XM_015642097.2), lance-tailed manakin (XM_032703318.1), and human (NM_002115.3); for *MLXIPL*, chicken (NM_001398117.1); for *MLX* chicken (NM_001030930.1) and Anna's hummingbird (XM_030465864.1)) and mapped to genome assemblies (for *HK3*, to our New Holland honeyeater assembly, as well as to the publically-available superb lyrebird assembly (GCA_013396355.1) and rufous treecreeper (GCA_013398175.1); *MLXIPL* and *MLX* were curated from our common swift assembly, as well as from the publically-available Anna's hummingbird assembly (GCA_003957555.2)). Transcriptomes from our RNA-seq data from New Holland honeyeater (liver, duodenum, heart, palate, and pectoralis), Anna's hummingbird and common swift (duodenum and intestine) were assembled using HISAT2 v2.1.0 and STRINGTIE v.2.2.1 to validate the predicted transcripts. In Anna's hummingbird, two corresponding splice variants of *MLXIPL* were identified, which differ solely by 279 base pairs in the first exon. Both variants were expressed in the duodenum, with the shorter variant also observed in one intestine sample. Similar, corresponding variants were observed in the common swift genome. The identified gene versions for *MLXIPL*, *MLX*, and *HK3* from various species (swift, hummingbird, honeyeater, lyrebird, and human) were synthesized by Genewiz for further experimental validation.

## Hexokinase purification

The synthesized *HK3* genes from New Holland honeyeater, rufous treecreeper, superb lyrebird, and human were cloned into the pCoofy6 vector [135] using the SLIC cloning method [136]. This created fusion proteins with His6-Sumo3 tags at the N-terminus of HK3, facilitating purification. The *E. coli* expression system was chosen based on previous successful purifications of human hexokinases (HK2 and HK3) [137,138].

Recombinant HK3 proteins were expressed in *Escherichia coli* Rosetta (DE3) T1 (Novagen N59205). Cultures were grown in auto-induction medium [139] at 37°C for 2 hours, in the presence of appropriate antibiotics. The temperature was then lowered to 20°C and cells were grown overnight. Cells were harvested by centrifugation at 6,000 rpm at 4°C for 10 minutes. The cell pellet was resuspended in lysis buffer (50 mM Sodium Phosphate, pH 8.0, 500 mM NaCl, 10 mM imidazole, 10% Glycerol, 1 mM TCEP, 0.1% protease inhibitor cocktail, and 2.4 U/ml Benzonase). Cells were lysed by LM20 microfluidizer, and the lysate was clarified by centrifugation at 20,500 rpm for 30 minutes at 4°C. The supernatant was loaded onto a Nickel-Chelating Resin pre-equilibrated with binding buffer (50 mM Sodium Phosphate, pH 8.0, 500 mM NaCl, 10 mM imidazole, 10% Glycerol and 1 mM TCEP) at 4°C. The resin was washed with binding buffer followed by washing buffer (same as binding buffer

but with 20 mM imidazole). The His-tagged HK3 was eluted with elution buffer (50 mM Sodium Phosphate, pH 8.0, 500 mM NaCl, 250 mM imidazole, 10% Glycerol and 1 mM TCEP). The eluted His-SUMO-HK3 was digested with SenP2 at 4°C in dialysis buffer (same as binding buffer) overnight. The Ni-NTA column was used again to remove His-SUMO, undigested His-SUMO-HK3, and SenP2 protease (in which the hexahistidine tag was also fused to the N-terminal), thus resulting in the native HK3. HK3 were further purified by size-exclusion chromatography using a HiLoad Superdex 200 column equilibrated with filtration buffer (50 mM Tris pH 7.4, 250 mM NaCl, 5% Glycerol and 1 mM TCEP). The purified protein concentration was determined by A280 measurement. Protein purity was assessed using SDS-PAGE. Finally, the purified HK3 in storage buffer (50 mM Tris, pH 7.5, 250 mM NaCl, 5% Glycerol, 1 mM TCEP) was snap-frozen in liquid nitrogen and stored at -80°C until further use.

## Hexokinase functional assay

The purified HK3 enzymes from New Holland honeyeater, rufous treecreeper, superb lyrebird, and human were diluted to a final concentration of 200 nM for activity measurements. To determine kinetics, hexokinase activity was assessed using the Abcam hexokinase activity assay kit (ab136957) according to the manufacturer's instructions, except the HK substrate was replaced by varying glucose concentrations (1, 0.5, 0.15, 0.075, 0.0375, 0.01875, and 0.009375 mM). The assay was performed at 25°C using a Flexstation3 microplate reader (Molecular Devices), with optical density (OD) readings taken every 30 seconds at 450 nm. Sample sizes were between 7 and 13 for each enzyme and glucose concentration combination. Kinetic parameters were estimated using R statistical software. Haldane's formula for substrate inhibition [140] was applied to fit the data:

$$v = Vmax*S / (Km + S + S^2/Ksi)$$

where v is the reaction velocity, S is the substrate concentration, Vmax is the maximum velocity, Km is the Michaelis constant, and Ksi is the substrate inhibition constant. The nonlinear least squares function in R was used to fit the model to the experimental data, with the PORT algorithm for optimization. Vmax, Km, and Ksi values were extracted from the model fit, and 95% confidence intervals were calculated, characterizing the hexokinase activity for HK3 enzymes (fig. S8). Statistical significance between species for each parameter was determined by non-overlapping 95% confidence intervals.

## Histological analysis

Tissues were excised, each piece was cut into aliquots of approximately 100 mg. Tissue aliquots were fixed in 4% PFA solution at 4°C overnight. The next day, tissue aliquots were placed in fresh sterile PBS and shipped with cooling packs at 4°C. The fixed tissues were stored in PBS for 48 hours. Next, tissues were incubated in 10% sucrose solution (24 hours) followed by incubation in

30% sucrose solution (24 hours). Then tissues were covered with OCT (Optimum Cutting Temperature compound, Thermofisher Scientific, 23730571) and placed in a labeled cryomold. Isopentane (2-methylbutane) was chilled in a steel beaker for 3-5 minutes using liquid nitrogen until white precipitates started to form on the bottom and walls of the beaker. The cryomolds with tissues were immersed in chilled isopentane for 6-12 seconds (depending on the tissue size). Cryomolds with tissues were transferred on dry ice for 15 mins to let isopentane evaporate. Cryomolds were filled up to the top with OCT and stored at -80°C until sectioning. Sectioning was performed on the cryostat NX70 (Thermofisher Scientific) with the following settings: -15°C temperature for the tissue and -29°C for the knife, 10m sections in a series of 2. The sections were collected on superfrost slides (Langenbrinck) and stored at -20°C until staining. Cryosections were stained for lipids with Oil Red O histological staining. In summary, sections were dried for 1 hour at room temperature. Next, slides were incubated in 60% isopropanol for 20 sec followed by the incubation in Oil Red O solution (Merck 1.02419.0250) for 10 mins. The sections were incubated in 60% isopropanol for 30 sec, rinsed with distilled water, counterstained with hematoxylin (Thermofisher Scientific, 7211) histological staining for 2 min, rinsed with distilled water, and mounted with Aquatex mounting media (Merck 1.08562.0050). Stained sections were imaged using Zeiss Imager.Z2 upright ApoTome, with Axiocam 512 color camera, at 10x, 40x, and 63x magnifications. For each section, at least three images were acquired. Imaging conditions were kept constant when acquiring images to be compared.

## Plasmid design *MLXIPL*

There are several *MLXIPL* isoforms annotated for the hummingbird (fig. S10C), so we manually curated the most likely functional isoforms based on the transcriptomic data. We found that two isoforms have a dominant expression (fig. S10D). These isoforms differ only in the N-terminus, the long isoform has an upstream start codon, while the short isoform uses an alternative downstream start codon (fig. S10C,D). The domain structure shows that the short isoform potentially lacks a part of a functional (LID) domain. Additionally, the short isoform possibly lacks a bird-specific nuclear localization signal (NLS) (fig. S10E).

## Cell culture and glucose-sensing

HEK293 cells were cultured in Dulbecco′s Modified Eagle′s Medium (DMEM, Thermo Fisher Scientific, REF: 41965-039) containing 10% FBS (gibco, REF: A3160801) and 1% Penicillin/Streptomycin (P/S, Thermo Fisher Scientific, REF: 15140122). Glucose Sensing was performed by incubation in 2.5mM glucose DMEM (Thermo Fisher Scientific, REF: 11966-0259) + FBS and P/S for 6 hours and following incubation in 2.5mM or 25mM glucose DMEM + FBS and P/S for 24 hours. Prior harvest cells were washed with ice cold PBS.

## Transfection

One day before transfection cells were seeded at a concentration of 12000 cells/well in Poly-L-Lysine (45µl/well, 2.5µg/ml for 30min @ room temperature; Bio&SELL, REF: BS.L 7250) coated 96 well plates (Greiner Bio-One, Ref: 655-180). Co-transfection of pCMV4-Mlxipl-5’FLAG (20ng/well) and pCMV4-Mlx (25ng/well) isoforms was carried out overnight at a cell confluence

of 60% using Lipofectamine 2000 (Thermo Fisher Scientific, REF: 11668019) according to the manufacturer's protocol.

## Immunoblotting

Cells were treated with lysis buffer (1:1 = 20% SDS : Non-Reducing Lane Marker Sample Buffer (Thermo Fisher Scientific, REF: 39001) + beta-Mercaptoethanol) and denatured at 95°C for 10min. Proteins were separated by SDS-PAGE and transferred to PVDF membranes by Western Blot overnight. Membranes were blocked in 4% milk. Anti-FLAG M2-Peroxidase (Sigma-Aldrich, REF: A8592, Lot: SLBK9652V), anti-beta-Actin (Santa Cruz, REF: sc-47778, Lot: #H1914), & anti-ChREBP (Novus Biologicals, REF: NB400-135, Lot: S2) were used as primary antibody and subsequently a corresponding secondary antibody when necessary. After incubation in chemiluminescent substrate (Thermo Fisher Scientific), bands were detected on a ChemiDocTM Imaging System (Bio-Rad).

## qPCR

RNA was extracted from cells using Total RNA Kit (peqGOLD, REF: 12-6834-02). After determination of RNA concentration per NanoDrop® Spectrophotometer ND-1000, cDNA was synthesized from equal RNA amounts using M-MLV Reverse Transcriptase (Promega, REF: M1708) followed by quantitative PCR with Takyon™ No ROX SYBR 2X MasterMix blue dTTP. Data analysis was performed using standard curve calculation and mRNA expression was normalized to human HPRT. Primer sequences are listed in (table S17).

## Luciferase Assay

Transcriptional activity was assessed using a Luciferase reporter driven by acetyl-CoA carboxylase (ACC1) Carbohydrate Response Element (ChoRE) (Tsatsos and Towle. Biochem Biophys Res Commun. 2006). The Dual-Luciferase® Reporter Assay System (Promega, REF: E1960) was used to assess transcriptional activity. The respective pCMV4-MLXIPL-5'FLAG isoform (20ng/well) was cotransfected with the corresponding pCMV4-MLX isoform (25ng/well), pGL3-ACC1-ChoRE firefly luciferase (100 ng/well) as well as pRL-Renilla luciferase (0.4ng/well) as internal control. Luminometry was performed as outlined in the manufacturers manual using Mithras LB 940 (Berthold technologies). Results were normalized to Renilla luciferase activity.

## Calculation of evolution score

We focused on the key biological processes where we observed the most significant evolutionary changes linked to sugar-feeding. For each of these pathways, we counted the number of sugar-consuming groups (up to a maximum of 4) that exhibited signs of selection for each annotated gene. We then calculated the regulatory and protein sequence evolution scores. The regulatory evolution score was based on the count of signals from phyloAcc and DESeq2, and the protein sequence evolution score was derived from aBSREL signals. Both scores were normalized by the pathway size (the total number of genes annotated for that pathway), the frequency of the signal from each method (the proportion of genes showing a signal anywhere in the phylogeny relative to the total number of genes tested), and the proportion of sugar branches tested (0.25 for aBSREL and phyloAcc, and 1 for DESeq2). Finally, we normalized the scores for each pathway by the protein and regulatory evolution score of calculated for a gene set of 2000 genes

randomly selected from the gneome. Since CSUBST showed signals in only 55 genes within sugar-consuming groups, we excluded it from this analysis.

# Competing interests

The authors have no competing interests.

# Acknowledgments

This work was supported by the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (LOEWE/1/10/519/03/03.001(0014)/52), the Max Planck Society, grants from the German Research Foundation (HI1423/4-1 and SCHU 2546/7-1). K.M.P. and M.S. were supported by the German Research Foundation (DFG, project ID:415542650. K.M.P. was also supported by the Sonnenfeld Foundation Berlin. We thank the support and expertise provided by the Protein Production Core Facility of the Max Planck Institute of Biochemistry. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and NAISS/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We thank the Macaulay Library at the Cornell Lab of Ornithology for providing bird images. We also thank Elena Gracheva, Madeleine Junkins, Paul Sunnucks, and Jane Reznick for comments on the manuscript, and Emily Wheeler for editorial assistance.

# Data and Code Availability

Genome assemblies are available on NCBI under the following accessions: *Acanthiza pusilla, Phylidonyris novaehollandiae, Furnarius rufus*, and *Trichoglossus moluccanus* are: PRJNA623276; *Florisuga fusca*: PRJNA622897; *Nymphicus hollandicus*: PRJNA819251; *Loriculus galgulus*: PRJNA1112892; *Ptilotula penicillata*: PRJNA1112893. Transcriptomic data are available on NCBI under the following accessions: *Calypte anna*: PRJNA1031337; *Apus apus*: PRJNA1031644; *Taeniopygia guttata*: PRJNA1032078; *Nymphicus hollandicus*: PRJNA1032113; *Phylidonyris novaehollandiae*: PRJNA1051659; *Trichoglossus moluccanus*: PRJNA1051684. TOGA annotations are available at https://genome.senckenberg.de/download/TOGA/. The scripts used in this study are available at https://github.com/maggieMCKO/nectargenomics.

# References

1. Nicolson, S. W. & Fleming, P. A. Drinking problems on a ‘simple’ diet: physiological convergence in nectar-feeding birds. *J. Exp. Biol.* **217**, 1015–1023 (2014).
2. Nicolson, S. W. & Fleming, P. A. Nectar as food for birds: the physiological consequences of drinking dilute sugar solutions. *Plant Syst Evol* **238**, 139–153 (2003).
3. Perret, M., Chautems, A., Spichiger, R., Peixoto, M. & Savolainen, V. Nectar sugar composition in relation to pollination syndromes in Sinningieae (Gesneriaceae). *Ann. Bot.* **87**, 267–273 (2001).
4. Stiles, F. G. Behavioral, ecological and morphological correlates of foraging for arthropods by the hummingbirds of a tropical wet forest. *Condor* **97**, 853–878 (1995).
5. Paton, D. C. & Collins, B. G. Bills and tongues of nectar-feeding birds: A review of morphology, function and performance, with intercontinental comparisons. *Aust. J. Ecol.* **14**, 473–506 (1989).
6. Cuban, D., Hewes, A. E., Sargent, A. J., Groom, D. J. E. & Rico-Guevara, A. On the feeding biomechanics of nectarivorous birds. *J. Exp. Biol.* **225**, jeb243096 (2022).
7. Hewes, A. E. *et al.* Variable evidence for convergence in morphology and function across avian nectarivores. *J. Morphol.* **283**, 1483–1504 (2022).
8. Suarez, R. K. Hummingbird flight: sustaining the highest mass-specific metabolic rates among vertebrates. *Experientia* **48**, 565–570 (1992).
9. Vitali, S. D., Withers, P. C. & Richardson, K. C. Standard metabolic rates of three nectarivorous meliphagid passerine birds. *Aust. J. Zool.* **47**, 385 (1999).
10. Suarez, R. K. & Welch, K. C. Sugar Metabolism in Hummingbirds and Nectar Bats. *Nutrients* **9**, (2017).
11. Osipova, E. *et al.* Loss of a gluconeogenic muscle enzyme contributed to adaptive metabolic traits in hummingbirds. *Science* **379**, 185–190 (2023).
12. Workman, R. E. *et al.* Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated hummingbird Archilochus colubris. *Gigascience* **7**, 1–12 (2018).
13. Gershman, A. *et al.* Genomic insights into metabolic flux in hummingbirds. *Genome Res* (2023) doi:10.1101/gr.276779.122.
14. Baldwin, M. W. *et al.* Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* **345**, 929–933 (2014).
15. Toda, Y. *et al.* Early origin of sweet perception in the songbird radiation. *Science* **373**, 226–231 (2021).
16. Wilman, H. *et al.* EltonTraits 1.0: Species-level foraging attributes of the world’s birds and mammals. *Ecology* **95**, 2027–2027 (2014).
17. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
18. Kirilenko, B. M. *et al.* TOGA integrates gene annotation with orthology inference at scale. *bioRxiv* 2022.09.08.507143 (2022) doi:10.1101/2022.09.08.507143.
19. Feng, S. *et al.* Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
20. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
21. Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).

22. Thibaud-Nissen, F. *et al.* P8008 The NCBI Eukaryotic Genome Annotation Pipeline. *J. Anim. Sci.* **94**, 184–184 (2016).
23. Kirilenko, B. M. *et al.* Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107 (2023).
24. Smith, M. D. *et al.* Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
25. Fukushima, K. & Pollock, D. D. Detecting macroevolutionary genotype–phenotype associations using error-corrected rates of protein convergence. *Nature Ecology & Evolution* **7**, 155–170 (2023).
26. Craig, R. J., Suh, A., Wang, M. & Ellegren, H. Natural selection beyond genes: Identification and analyses of evolutionarily conserved elements in the genome of the collared flycatcher (Ficedula albicollis). *Mol. Ecol.* **27**, 476–492 (2018).
27. Lowe, C. B., Clarke, J. A., Baker, A. J., Haussler, D. & Edwards, S. V. Feather development genes and associated regulatory innovation predate the origin of Dinosauria. *Mol. Biol. Evol.* **32**, 23–28 (2015).
28. Sackton, T. B. *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
29. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
30. Hu, Z., Sackton, T. B., Edwards, S. V. & Liu, J. S. Bayesian Detection of Convergent Rate Changes of Conserved Noncoding Elements on Phylogenetic Trees. *Mol. Biol. Evol.* **36**, 1086–1100 (2019).
31. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
32. Whittle, A. J. *et al.* BMP8B increases brown adipose tissue thermogenesis through both central and peripheral actions. *Cell* **149**, 871–885 (2012).
33. Quiros, P. M. *et al.* Loss of mitochondrial protease OMA1 alters processing of the GTPase OPA1 and causes obesity and defective thermogenesis in mice. *EMBO J.* **31**, 2117–2133 (2012).
34. Abdul-Wahed, A., Guilmeau, S. & Postic, C. Sweet Sixteenth for ChREBP: Established Roles and Future Goals. *Cell Metab* **26**, 324–341 (2017).
35. Yamashita, H. *et al.* A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9116–9121 (2001).
36. Regnier, M., Carbinatti, T., Parlati, L., Benhamed, F. & Postic, C. The role of ChREBP in carbohydrate sensing and NAFLD development. *Nat. Rev. Endocrinol.* **19**, 336–349 (2023).
37. Iizuka, K., Bruick, R. K., Liang, G., Horton, J. D. & Uyeda, K. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7281–7286 (2004).
38. Gordon, W. E. *et al.* Integrative single-cell characterization of a frugivorous and an insectivorous bat kidney and pancreas. *Nat. Commun.* **15**, 12 (2024).
39. McWhorter, T. J. *et al.* Sucrose digestion capacity in birds shows convergent coevolution with nectar composition across continents. *iScience* **24**, 102717 (2021).

40. Napier, K. R., McWhorter, T. J., Nicolson, S. W. & Fleming, P. A. Sugar preferences of avian nectarivores are correlated with intestinal sucrase activity. *Physiol. Biochem. Zool.* **86**, 499–514 (2013).
41. Santos, J. L. *et al.* Copy number polymorphism of the salivary amylase gene: implications in human nutrition research. *J. Nutrigenet. Nutrigenomics* **5**, 117–131 (2012).
42. Elferink, H., Bruekers, J. P. J., Veeneman, G. H. & Boltje, T. J. A comprehensive overview of substrate specificity of glycoside hydrolases and transporters in the small intestine : 'A gut feeling': 'A gut feeling'. *Cell. Mol. Life Sci.* **77**, 4799–4826 (2020).
43. Galand, G. Brush border membrane sucrase-isomaltase, maltase-glucoamylase and trehalase in mammals. Comparative development, effects of glucocorticoids, molecular mechanisms, and phylogenetic implications. *Comp. Biochem. Physiol. B* **94**, 1–11 (1989).
44. Brun, A. *et al.* Duplications and Functional Convergence of Intestinal Carbohydrate-Digesting Enzymes. *Mol. Biol. Evol.* **37**, 1657–1666 (2020).
45. Cao, D., Fan, S. T. & Chung, S. S. Identification and characterization of a novel human aldose reductase-like gene. *J. Biol. Chem.* **273**, 11429–11435 (1998).
46. Iizuka, K. Recent Progress on Fructose Metabolism-Chrebp, Fructolysis, and Polyol Pathway. *Nutrients* **15**, (2023).
47. Chandel, N. S. Mitochondria. *Cold Spring Harb. Perspect. Biol.* **13**, (2021).
48. Aleshin, A. E. *et al.* The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose and glucose-6-phosphate. *Structure* **6**, 39–50 (1998).
49. White, T. K. & Wilson, J. E. Rat brain hexokinase: Location of the allosteric regulatory site in a structural domain at the N-terminus of the enzyme. *Archives of Biochemistry and Biophysics* **259**, 402–411 (1987).
50. Heumann, S., Falkenberg, F. & Pfleiderer, G. Purification and immunological characterization of the human hexokinase isoenzymes I and III (ATP-d-hexose 6-phosphotransferase EC 2.7.1.1). *Biochimica et Biophysica Acta (BBA) - Enzymology* **334**, 328–342 (1974).
51. Cayanis, E. & Balinsky, D. Comparative kinetic properties of human hexokinases. *Int. J. Biochem.* **6**, 741–749 (1975).
52. González, C., Ureta, T., Sánchez, R. & Niemeyer, H. Multiple molecular forms of ATP:hexose 6-phosphotransferase from rat liver. *Biochem. Biophys. Res. Commun.* **16**, 347–352 (1964).
53. Radojković, J. & Ureta, T. Hexokinase isoenzymes from the Novikoff hepatoma. Purification, kinetic and structural characterization, with emphasis on hexokinase C. *Biochem. J* **242**, 895–903 (1987).
54. Balinsky, D. & Fromm, H. J. Hexokinase III from Rana catesbeiana. *Comp. Biochem. Physiol. B* **60**, 71–75 (1978).
55. Reed, M. C., Lieb, A. & Nijhout, H. F. The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays* **32**, 422–429 (2010).
56. Wilson, J. E. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J. Exp. Biol.* **206**, 2049–2057 (2003).
57. Perez-Pozo, S. E. *et al.* Excessive fructose intake induces the features of metabolic syndrome in healthy adult men: role of uric acid in the hypertensive response. *Int J Obes (Lond)* **34**, 454–461 (2010).

58. Barbagallo, M., Shan, J., Pang, P. K. & Resnick, L. M. Glucose-induced alterations of cytosolic free calcium in cultured rat tail artery vascular smooth muscle cells. *J Clin Invest* **95**, 763–767 (1995).
59. Nandadasa, S., Nelson, C. M. & Apte, S. S. ADAMTS9-Mediated Extracellular Matrix Dynamics Regulates Umbilical Cord Vascular Smooth Muscle Differentiation and Rotation. *Cell Rep.* **11**, 1519–1528 (2015).
60. Bround, M. J. *et al.* Cardiomyocyte ATP production, metabolic flexibility, and survival require calcium flux through cardiac ryanodine receptors in vivo. *J. Biol. Chem.* **288**, 18975–18986 (2013).
61. Bround, M. J. *et al.* Cardiac Ryanodine Receptor (Ryr2)-mediated Calcium Signals Specifically Promote Glucose Oxidation via Pyruvate Dehydrogenase. *J. Biol. Chem.* **291**, 23490–23505 (2016).
62. Yamamoto, W. R. *et al.* Endoplasmic reticulum stress alters ryanodine receptor function in the murine pancreatic β cell. *J. Biol. Chem.* **294**, 168–181 (2019).
63. Harvey, K. E. *et al.* RyR2/IRBIT regulates insulin gene transcript, insulin content, and secretion in the insulinoma cell line INS-1. *Sci. Rep.* **12**, 7713 (2022).
64. Xu, F. *et al.* Disruption of a mitochondrial RNA-binding protein gene results in decreased cytochrome b expression and a marked reduction in ubiquinol-cytochrome c reductase activity in mouse heart mitochondria. *Biochem. J* **416**, 15–26 (2008).
65. Baleva, M. V. *et al.* Pentatricopeptide Protein PTCD2 Regulates COIII Translation in Mitochondria of the HeLa Cell Line. *Int. J. Mol. Sci.* **23**, (2022).
66. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
67. Bottomley, W. E. *et al.* IRS2 variants and syndromes of severe insulin resistance. *Diabetologia* **52**, 1208–1211 (2009).
68. White, M. F. IRS2 integrates insulin/IGF1 signalling with metabolism, neurodegeneration and longevity. *Diabetes Obes. Metab.* **16 Suppl 1**, 4–15 (2014).
69. Sélénou, C., Brioude, F., Giabicani, E., Sobrier, M.-L. & Netchine, I. IGF2: Development, Genetic and Epigenetic Abnormalities. *Cells* **11**, (2022).
70. Ullrich, S. *et al.* Effects of I(Ks) channel inhibitors in insulin-secreting INS-1 cells. *Pflugers Arch.* **451**, 428–436 (2005).
71. Fleming, P. A. & Nicolson, S. W. Osmoregulation in an avian nectarivore, the whitebellied sunbird Nectarinia talatala: response to extremes of diet concentration. *J. Exp. Biol.* **206**, 1845–1854 (2003).
72. Fleming, P. A., Gray, D. A. & Nicolson, S. W. Osmoregulatory response to acute diet change in an avian nectarivore: rapid rehydration following water shortage. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **138**, 321–326 (2004).
73. Werner, H. The IGF1 signaling pathway: From basic concepts to therapeutic opportunities. *Int. J. Mol. Sci.* **24**, (2023).
74. Saha, B. *et al.* Potassium Activates mTORC2-dependent SGK1 Phosphorylation to Stimulate Epithelial Sodium Channel: Role in Rapid Renal Responses to Dietary Potassium. *J. Am. Soc. Nephrol.* **34**, 1019–1038 (2023).

75. Li, M. V., Chang, B., Imamura, M., Poungvarin, N. & Chan, L. Glucose-dependent transcriptional regulation by an evolutionarily conserved glucose-sensing module. *Diabetes* **55**, 1179–1189 (2006).
76. Singh, P. & Irwin, D. M. Contrasting Patterns in the Evolution of Vertebrate MLX Interacting Protein (MLXIP) and MLX Interacting Protein-Like (MLXIPL) Genes. *PLoS One* **11**, e0149682 (2016).
77. Kim, M. *et al.* Intestinal, but not hepatic, ChREBP is required for fructose tolerance. *JCI Insight* **2**, (2017).
78. Jang, C. *et al.* The small intestine shields the liver from fructose-induced steatosis. *Nat. Metab.* **2**, 586–593 (2020).
79. Herman, M. A. & Birnbaum, M. J. Molecular aspects of fructose metabolism and metabolic disease. *Cell Metab.* **33**, 2329–2354 (2021).
80. Hannou, S. A., Haslam, D. E., McKeown, N. M. & Herman, M. A. Fructose metabolism and metabolic disease. *J Clin Invest* **128**, 545–555 (2018).
81. Tsatsos, N. G., Davies, M. N., O'Callaghan, B. L. & Towle, H. C. Identification and function of phosphorylation in the glucose-regulated transcription factor ChREBP. *Biochem. J.* **411**, 261–270 (2008).
82. Minn, A. H., Hafele, C. & Shalev, A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. *Endocrinology* **146**, 2397–2405 (2005).
83. Pashkov, V. *et al.* Regulator of G protein signaling (RGS16) inhibits hepatic fatty acid oxidation in a carbohydrate response element-binding protein (ChREBP)-dependent manner. *J Biol Chem* **286**, 15116–15125 (2011).
84. Potter, J. H. T. *et al.* Nectar-feeding bats and birds show parallel molecular adaptations in sugar metabolism enzymes. *Curr. Biol.* **31**, 4667–4674.e6 (2021).
85. Hickey, A. J. *et al.* Birds and longevity: does flight driven aerobicity provide an oxidative sink? *Ageing Res. Rev.* **11**, 242–253 (2012).
86. Butler, P. J. The physiological basis of bird flight. *Philos Trans R Soc Lond B Biol Sci* **371**, (2016).
87. Robledo-Ruiz, D. A. *et al.* Chromosome-length genome assembly and linkage map of a critically endangered Australian bird: the helmeted honeyeater. *Gigascience* **11**, (2022).
88. Pauw, A. A Bird's-Eye View of Pollination: Biotic Interactions as Drivers of Adaptation and Community Change. *Annu. Rev. Ecol. Evol. Syst.* **50**, 477–502 (2019).
89. Sin, S. Y. W., Lu, L. & Edwards, S. V. De Novo Assembly of the Northern Cardinal (Cardinalis cardinalis) Genome Reveals Candidate Regulatory Regions for Sexually Dichromatic Red Plumage Coloration. *G3* **10**, 3541–3548 (2020).
90. Sin, S. Y. W., Cloutier, A., Nevitt, G. & Edwards, S. V. Olfactory receptor subgenome and expression in a highly olfactory procellariiform seabird. *Genetics* **220**, (2022).
91. Monte, A. *et al.* Testosterone treatment unveils testosterone-insensitive song in an early-branched hummingbird. *Ethol. Ecol. Evol.* 1–29 (2023).
92. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566–568 (2014).
93. Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**, 1050–1054 (2016).

94. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
95. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
96. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
97. Ghurye, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol* **15**, e1007273 (2019).
98. Ludwig, A., Pippel, M., Myers, G. & Hiller, M. DENTIST-using long reads for closing assembly gaps at high accuracy. *Gigascience* **11**, (2022).
99. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
100. Garg, S. *et al.* Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat. Biotechnol.* **39**, 309–312 (2021).
101. He, S. *et al.* Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun Biol* **3**, 361 (2020).
102. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
103. Koren, S. *et al.* Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).
104. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
105. Kolmogorov, M., Raney, B., Paten, B. & Pham, S. Ragout—a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* **30**, i302–i309 (2014).
106. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
107. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
108. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
109. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1513–1518 (2011).
110. Ewels, P. A. *et al.* nf-core: Community curated bioinformatics pipelines. *bioRxiv* 610741 (2019) doi:10.1101/610741.
111. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
112. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
113. Smit, A. F. A. & Hubley, R. RepeatModeler Open-1.0. www.repeatmasker.org (2008-2010).
114. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-3.0. www.repeatmasker.org (1996-2010).
115. Harris, R. S. Improved pairwise alignment of genomic DNA. (The Pennsylvania State University, 2007).

116. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**, 11484–11489 (2003).
117. Suarez, H. G., Langer, B. E., Ladde, P. & Hiller, M. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596–1603 (2017).
118. Osipova, E., Hecker, N. & Hiller, M. RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**, (2019).
119. Gremme, G., Brendel, V., Sparks, M. E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
120. Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
121. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
122. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).
123. Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19**, 21 (2019).
124. Lucaci, A. G., Zehr, J. D., Enard, D., Thornton, J. W. & Kosakovsky Pond, S. L. Evolutionary Shortcuts via Multinucleotide Substitutions and Their Impact on Natural Selection Analyses. *Mol Biol Evol* **40**, (2023).
125. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
126. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591 (2007).
127. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).
128. Hickey, G., Paten, B., Earl, D., Zerbino, D. & Haussler, D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).
129. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
130. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
131. Creevey, C. & Weeks, N. *ChrisCreevey/catsequences: Version 1.4*. (Zenodo, 2023). doi:10.5281/ZENODO.4312908.
132. Yan, H. *et al.* PhyloAcc-GT: A Bayesian Method for Inferring Patterns of Substitution Rate Shifts on Targeted Lineages Accounting for Gene Tree Discordance. *Mol. Biol. Evol.* **40**, (2023).
133. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
134. Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004).

135. Addgene: pCoofy6. https://www.addgene.org/43990/.
136. Scholz, J., Besir, H., Strasser, C. & Suppmann, S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnol.* **13**, 12 (2013).
137. Ferreira, J. C. *et al.* Linker residues regulate the activity and stability of hexokinase 2, a promising anticancer target. *J. Biol. Chem.* **296**, 100071 (2021).
138. Wyatt, E. *et al.* Regulation and cytoprotective role of hexokinase III. *PLoS One* **5**, e13823 (2010).
139. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
140. Leow, J. W. H. & Chan, E. C. Y. Atypical Michaelis-Menten kinetics in cytochrome P450 enzymes: A focus on substrate inhibition. *Biochem. Pharmacol.* **169**, 113615 (2019).