Loss of a gluconeogenic muscle enzyme contributed to adaptive metabolic traits in hummingbirds

Ekaterina Osipova[1,2,3,4,5,6], Rico Barsacchi[1], Tom Brown[1,3,7], Keren Sadanandan[8], Andrea H. Gaede[9,10], Amanda Monte[11], Julia Jarrells[1], Claudia Moebius[1], Martin Pippel[1,3], Douglas L. Altshuler[9], Sylke Winkler[1,7], Marc Bickle[12], Maude W. Baldwin[8], Michael Hiller[1,2,3,4,5,6*]

[1] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany,
[2] Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
[3] Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany
[4] LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
[5] Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany
[6] Goethe-University, Faculty of Biosciences, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
[7] DRESDEN concept Genome Center, Technische Universität Dresden, 01062 Dresden, Germany
[8] Evolution of Sensory Systems Research Group, Max Planck Institute for Ornithology, Seewiesen, Germany
[9] University of British Columbia, Vancouver, Vancouver, BC, V6T 1Z4 Canada
[10] Structure and Motion Laboratory, Royal Veterinary College, University of London, London, England, United Kingdom
[11] Department of Behavioural Neurobiology, Max Planck Institute for Ornithology, Seewiesen, Germany
[12] Roche Institute for Translational Bioengineering, Grenzacherstrasse 124, 4070 Basel, Switzerland

*To whom correspondence should be addressed:
Michael Hiller
LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
Tel: +49 69 7542-1398
E-Mail: michael.hiller@senckenberg.de

Running title: Metabolic adaptations in hummingbird muscle

Keywords: gene loss, metabolic adaptations, hummingbird muscle, hovering flight, positive selection

Abstract:
Hummingbirds possess remarkable metabolic adaptations to fuel their energy-demanding hovering flight, but the underlying genomic changes are largely unknown. Here, we generated a chromosome-level genome assembly of the long-tailed hermit and screened for genes that have been specifically inactivated in the ancestral hummingbird lineage. We discovered that *FBP2*, a gluconeogenic muscle enzyme, was lost during a time period when hovering flight evolved. We show that *FBP2* knockdown in an avian muscle cell line upregulates glycolysis and enhances mitochondrial respiration, coincident with an increased mitochondria number. Furthermore, genes involved in mitochondrial respiration and organization have upregulated expression in hummingbird flight muscle. Together, these results suggest that *FBP2* loss was likely a key step in the evolution of metabolic muscle adaptations required for true hovering flight.

One-Sentence Summary: *FBP2* loss in hummingbirds coincided with the evolution of true hovering flight and likely contributed to muscle adaptations

Hummingbirds are the only birds capable of true (sustained) hovering flight and exhibit exceptional maneuverability including backward flight. These feats are enabled by shoulder girdle adaptations that produce lift both with the up- and downstroke. Requiring rapid wingbeats, hovering flight is the most energy-demanding locomotion type and hummingbirds have the highest mass-specific metabolic rate measured for vertebrates (*1, 2*). Hummingbirds fuel this energy-demanding activity largely with sugars, obtained by feeding on flower nectar. Within minutes after feeding, hummingbirds can fuel almost their entire metabolism with newly-ingested sugars (*3, 4*).

These metabolic adaptations are facilitated by high sugar uptake rates and increased rates of carbohydrate metabolizing enzymes (*4-6*). Hummingbirds have the unique ability to directly metabolize dietary fructose as efficiently as glucose (*7, 8*) and their sweet taste perception evolved by co-option of an umami taste receptor (*9*). Sequence analyses revealed signatures of positive selection in hummingbird glycolytic enzymes (*10*). However, the genomic basis of metabolic muscle adaptations in hummingbirds is largely unknown.

Assembly of the long-tailed hermit genome

True hovering flight and nectarivory are ancestral features of hummingbirds that evolved after the hummingbird and swift lineage split ~54 Mya and before the beginning of the hummingbird radiation ~22 Mya (*11, 12*). To reveal the genomic underpinnings of adaptations that evolved during that period, we generated PacBio continuous long reads, 10x genomics linked reads, Bionano optical maps and chromosome conformation capture (Hi-C) read pairs to assemble a highly-contiguous, chromosome-level genome of the long-tailed hermit (*Phaethornis superciliosus*), a representative of the hermit subfamily that is a sister clade to nearly all other hummingbirds (fig. S1A). Together with two other sequenced hummingbirds (*13-15*), this allowed us to infer ancestral hummingbird mutations based on parsimony (fig. S1B). Our hermit assembly has contig and scaffold N50 values of 35.2 Mb and 72.7 Mb, respectively, and thus a higher contiguity than previous assemblies of two hummingbirds and even chicken (Fig. 1A). We estimated a high base accuracy with only ~3 remaining errors per 1 Mb (QV 55.3).

To compare assembly quality, we used a previously-developed approach (*16*) to determine how many of 11,611 ancestral bird genes have (i) an intact reading frame, (ii) gene-inactivating mutations (in-frame stop codons, frameshifting insertions and deletions, splice site mutations, and exon deletions) or (iii) missing sequence due to assembly incompleteness (Tables S1-S2). Considering genomes of 50 birds that cover major clades of the avian phylogeny (*11*), our long-tailed hermit assembly is among those with the highest number of intact genes (Fig. 1B). Compared to previous hummingbird assemblies, our hermit assembly has fewer genes with inactivating mutations (Fig. 1B, fig. S2). Together, this supports not only a high level of assembly completeness but also a high base accuracy.

Identification of hummingbird-specific gene losses

Loss of ancestral coding genes can be relevant for adaptation (*16-18*). To investigate whether gene losses may have contributed to hummingbird adaptations, we performed a genome-wide screen for genes that were specifically inactivated in the ancestral hummingbird lineage. We identified genes likely lost in all three sequenced hummingbird species and then excluded those genes that are also lost in swifts (the closest insectivorous sister lineage of hummingbirds) or any of 45 other birds (Table S1). This screen identified five genes (Fig. 1C). Three of these five genes (*FBP2*, *PTER*, *CAPN13*) have inactivating mutations shared between the three hummingbird species, which parsimoniously suggests that these mutations, and thus the gene losses, occurred on the ancestral hummingbird branch (Fig. 1D, fig. S1B,3).

***FBP2* loss coincided with the evolution of hovering flight**

Whereas four of these five genes have no known function related to metabolism and physiology, the fifth gene is *FBP2* (fructose-bisphosphatase 2) (Table S3). This gene encodes a muscle FBPase enzyme that catalyzes a rate-limiting step in gluconeogenesis and converts fructose-1,6-bisphosphate to fructose-6-phosphate (*19*). *FBP2* exhibits numerous inactivating mutations, resulting in a non-functional protein. Six mutations are shared among assemblies of the three hummingbird species (Fig. 1D), which excludes the possibility of base errors.

To further investigate when the loss of *FBP2* happened on the ancestral hummingbird branch, we used a molecular dating approach based on the rate of nonsynonymous to synonymous substitutions and species divergence times (*11, 12*). We estimated that *FBP2* inactivation happened 34-46 Mya (Table S4). Two key fossils illuminate when hovering flight evolved in the hummingbird lineage. An insectivorous apodiform bird fossil (*Parargornis messelensis*) found in 48-million-year-old strata in Germany's Messel pit provides an upper bound (*20*). The oldest hummingbird-like fossil (*Eurotrochilus inexpectatus*), possessing key anatomical features for nectar feeding and hovering flight, dated 30-35 Mya (*21*). These fossils suggest that true hovering flight likely evolved between 48 and ~30 Mya, and our data indicate that *FBP2* loss coincided with this period (Fig. 1D).

Hummingbird muscle expresses no FBPases

In tetrapods, FBPase enzymes are encoded by two paralogous genes: *FBP1* that is mostly expressed in the liver and kidneys and *FBP2* that is predominantly expressed in skeletal muscle (*22*). Since none of the investigated bird species including hummingbirds have lost *FBP1* (Table S3), we tested whether hummingbird *FBP1* is expressed in muscle to compensate for *FBP2* loss. To this end, we performed RNA-sequencing on biological triplicates of the major flight muscle (pectoralis) and liver tissue of Anna's hummingbird and the common swift. As in mammals, the common swift expressed *FBP1* in liver and *FBP2* in muscle (Fig. 2). In the hummingbird, no relevant exon expression of the *FBP2* remnant was detected (fig. S4), confirming its loss. Importantly, the intact hummingbird *FBP1* gene is expressed in liver but not muscle (Fig. 2). This indicates that the consequences of *FBP2* loss are likely restricted to muscle, as this tissue expresses no FBPase enzymes in hummingbirds.

FBP2 knockdown upregulates glycolysis and mitochondrial respiration

Gluconeogenic FBPase enzymes oppose the action of phosphofructokinase, one of the rate-limiting enzymes in glycolysis (*23*). Thus, loss of the muscle enzyme encoded by *FBP2* may increase glycolytic flux. To test this experimentally, we used QM7 cells, a quail muscle (myoblast) cell line that expresses *FBP2* (fig. S5), and the CRISPR-Cas9 system to knockout *FBP2*. We designed four guide RNAs that specifically target intronic regions around exon 5 (Fig. 3A), which is required for substrate binding and encodes the nuclear localization signal required for FBP2's mitochondrial function (*24, 25*). While neither wild-type nor knockout QM7 cells survive in isolation, not allowing for the generation of a clonal knockout lineage, we achieved deletions of *FBP2* exon 5 in pools of cells (Fig. 3B, fig. S6). In these *FBP2* knockdown cells, we measured a 65% (P=0.0007) reduction of *FBP2* expression without a compensatory upregulation of *FBP1* and a 20% (P=0.0025) lower FBPase activity (Fig. 3B-D).

To test whether *FBP2* knockdown increases glycolytic flux, we used the Seahorse Glycolytic Rate Assay to measure in real time the glycolytic proton efflux rate, a direct readout of glycolytic flux. Both basal glycolysis and compensatory glycolysis, induced by mitochondrial inhibition, were significantly increased in *FBP2* knockdown cells in comparison with mock-treated wild-type cells (Fig. 3E,F, fig. S7).

Glycolysis produces pyruvate, which can be catabolized by mitochondrial oxidative phosphorylation (OXPHOS) under aerobic conditions. Analyzing the Seahorse real time measurements of the oxygen consumption rate, a readout of OXPHOS, showed that *FBP2* knockdown significantly upregulates mitochondrial respiration (Fig. 3G,H), which we estimate to generate most ATP in QM7 cells (fig. S8). Together, this shows that *FBP2* knockdown increases both major energy-producing pathways in this avian myoblast cell line, potentially explaining the small but significant increase in lipid accumulation (fig. S9).

FBP2 knockdown increases mitochondria number

The observed increase in mitochondrial respiration could be due to increased mitochondrial biogenesis. Indeed, using the ratio of mitochondrial to nuclear DNA as a proxy for mitochondria number, we observed a ~50% higher ratio in *FBP2* knockdown compared to mock-treated wild-type cells (Fig. 3I, fig. S10). This indicates that *FBP2* downregulation enhances mitochondrial biogenesis, which is in line with observations that hummingbird flight muscles exhibit a high mitochondrial density (*26*).

Upregulation of genes for mitochondrial biosynthesis and function

We next explored whether signatures of increased mitochondrial biogenesis and respiration are also evident in the transcriptomic data of hummingbirds that naturally lost *FBP2*. We first focused on mitochondria-encoded genes and found a ~30% increased expression in pectoralis muscle but not liver of Anna's hummingbird in comparison to the common swift (Fig. 3J, fig. S11). We next compared the expression of 597 nuclear-encoded genes that function in mitochondria.

Genes significantly up- or downregulated in hummingbird muscle showed distinct functional enrichments. Downregulated genes are enriched for fatty acid and amino acid metabolism and include important enzymes such as *ACADSB*, *ACSS3*, *GCAT*, and *BCKDHB* (Fig. 3K, Table S5). This is consistent with fatty acid and amino acid metabolism being less important for obligate nectarivores. In contrast, upregulated genes are enriched for carbohydrate metabolism, the tricarboxylic acid (TCA) cycle, regulation of electron transport chain function, and the formation of inner mitochondrial membrane folds (cristae) (Fig. 3K, Table S5). Upregulated genes include *COX6A1*, encoding a subunit of cytochrome c oxidase, and *IDH3A* and *IDH3B*, encoding two subunits of mitochondrial NAD^+-specific isocitrate dehydrogenase. IDH3 catalyzes a key reaction in the TCA cycle, and its upregulation promotes a metabolic switch from glycolysis to OXPHOS (*27*). Overall, these expression differences are consistent with muscle-specific FBP2 negatively regulating mitochondria number and with previous observations that hummingbird flight muscle exhibit a high mitochondrial density and high levels of mitochondrial respiration (*26*).

Glycolytic genes evolved under positive selection

To investigate whether, in addition to the loss of *FBP2*, other genes involved in glucose metabolism exhibit changes in the ancestral hummingbird lineage, we screened for signatures of positive selection in 46 genes with roles in glycolysis and gluconeogenesis, glycogen synthesis, and monosaccharide uptake (Fig. 4A,B, Table S6). We detected positive selection on the ancestral hummingbird branch for the less characterized inositol (a sugar alcohol) transporter *SLC2A13* and five glycolytic genes (*PFKP*, *ALDOB*, *PDHA2*, *PDHB, ENO1*). *PFKP* encodes a phosphofructokinase enzyme that is expressed in many tissues including muscle (*22*) and catalyzes the committed step in glycolysis. *ALDOB* encodes the liver-expressed aldolase, a key enzyme in both glycolysis and fructolysis, whose overexpression in cancer cells increases fructose metabolism (*28*). *PDHA2* and *PDHB* encode subunits of the pyruvate dehydrogenase enzyme complex that converts pyruvate to acetyl-CoA, and thus connects glycolysis with the TCA cycle. *ENO1* encodes a ubiquitously-expressed glycolytic enzyme whose upregulation enhances glycolysis in tumors (*29*). Consistent with positive selection, we observed amino acid substitutions shared between all three hummingbird species, and some of them are potentially involved in regulating protein function or stability (Fig. 4C, fig. S12). Interestingly, *ALDOB* and *PDHB* also evolved under positive selection in nectar-feeding bats (*10*). Finally, compared to swift, expression of *ALDOB* and *ENO1* is significantly upregulated in hummingbird liver and expression of *PDHA2* and *PDHB* is significantly upregulated in hummingbird muscle and liver (Fig. 4D), indicating additional regulatory changes.

Discussion

Here, we discovered that the gluconeogenic muscle enzyme *FBP2* was inactivated in ancestral hummingbirds, coinciding with the time period during which energy-demanding hovering flight evolved. Experiments in an avian myoblast cell line showed that *FBP2* knockdown upregulated glycolysis, mitochondrial biogenesis, and mitochondrial respiration, which concurs with characteristics of hummingbird flight muscles. The evolutionary loss of *FBP2* in hummingbirds provides interesting parallels to human cancer cells, where downregulation of *FBP2* (and *FBP1*) is frequently observed, because FBPases have tumor suppressing functions (*23, 25, 30*). Consistent with our knockdown experiments in quail myoblast cells, downregulation of *FBP1* or *FBP2* in cancer cells enhances glycolysis and contributes to the Warburg effect, describing the preference of cancer cells for aerobic glycolysis to generate three-carbon precursors for rapid growth (*25, 30*). Furthermore, restoring *FBP2* expression in cancer cells inhibits mitochondrial biogenesis and respiration (*25*).

While signatures of selection in more broadly-expressed glycolytic enzymes, detected here and previously (*10*), highlight additional genomic changes with relevance for systemic metabolic adaptations in hummingbirds, metabolic shifts caused by losing the muscle-expressed *FBP2* gene were likely restricted to muscle tissue. Since downregulation or inactivation of *FBP2* provides a mechanism to increase metabolic capacity in muscle by upregulating sugar metabolism, the loss of the *FBP2* gene was likely beneficial for hummingbirds and could have been a key step in the co-evolution of nectarivory and energy-demanding hovering flight.

Methods summary

To assemble the hermit genome, we used Falcon and Falcon Unzip (*31*), purged haplotypic duplications with purge-dups (*32*), performed scaffolding with Scaff10X, Bionano Solve and salsa2 (*35*), and polished the assembly using 10X linked-reads. Merqury (*36*) with the 10X data was used to estimate base accuracy. RepeatModeler (*37*) and RepeatMasker (*38*) were used to soft-mask repeats. Pairwise genome alignments between chicken and other birds was computed using LASTZ (*39*), axtChain (*40*), chainCleaner (*41*), and RepeatFiller (*42*). Ancestral bird genes were defined as chicken genes annotated by Ensembl that have an intact reading frame in at least one of ostrich, kiwi, or tinamou, representing one of the earliest diverging avian clades (*11*). To detect hummingbird-specific gene losses, we used a previously-developed approach (*16*) to screen for gene-inactivating mutations in hummingbirds and excluded genes lost in other birds. To detect genes evolving under positive selection, we used TOGA (*48*) to obtain coding sequences orthologous to chicken Ensembl genes, generated multiple codon alignments with MACSE v2 (*56*), applied the branch-site model of PAML codeml (*51*), and corrected for multiple testing with the Benjamini–Hochberg method (*57*). Homology modeling was performed with the https://swissmodel.expasy.org/ modeling service (*58*). Additional details and methods concerning sample collection, genome and RNA sequencing, gene loss dating, analysis of transcriptomic data and all experiments in QM7 cells can be found in the supplementary materials.

References

1. T. Weis-Fogh, Energetics of Hovering Flight in Hummingbirds and in Drosophila. *J Exp Biol* **56**, 79-104 (1972).
2. R. K. Suarez, Hummingbird flight: sustaining the highest mass-specific metabolic rates among vertebrates. *Experientia* **48**, 565-570 (1992).
3. K. C. Welch, Jr., B. H. Bakken, C. Martinez del Rio, R. K. Suarez, Hummingbirds fuel hovering flight with newly ingested sugar. *Physiol Biochem Zool* **79**, 1082-1087 (2006).
4. R. K. Suarez, K. C. Welch, Sugar Metabolism in Hummingbirds and Nectar Bats. *Nutrients* **9**, (2017).
5. T. J. McWhorter, B. H. Bakken, W. H. Karasov, C. M. del Rio, Hummingbirds rely on both paracellular and carrier-mediated intestinal glucose absorption to fuel high metabolism. *Biol Lett* **2**, 131-134 (2006).
6. J. E. Schondube, C. Martinez del Rio, Sugar and protein digestion in flowerpiercers and hummingbirds: a comparative test of adaptive convergence. *J Comp Physiol B* **174**, 263-273 (2004).
7. C. C. W. W. Chen, K. C., Hummingbirds can fuel expensive hovering flight completely with either exogenous glucose or fructose. *Functional Ecology* **28**, 589-600 (2013).
8. A. M. Myrka, K. C. Welch, Jr., Evidence of high transport and phosphorylation capacity for both glucose and fructose in the ruby-throated hummingbird (Archilochus colubris). *Comp Biochem Physiol B Biochem Mol Biol* **224**, 253-261 (2018).
9. M. W. Baldwin *et al.*, Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* **345**, 929-933 (2014).
10. J. H. T. Potter *et al.*, Nectar-feeding bats and birds show parallel molecular adaptations in sugar metabolism enzymes. *Curr Biol*, (2021).
11. R. O. Prum *et al.*, A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569-573 (2015).
12. J. A. McGuire *et al.*, Molecular phylogenetics and the diversification of hummingbirds. *Curr Biol* **24**, 910-916 (2014).
13. E. D. Jarvis *et al.*, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331 (2014).
14. A. Rhie *et al.*, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737-746 (2021).
15. S. Feng *et al.*, Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252-257 (2020).
16. V. Sharma *et al.*, A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun* **9**, 1215 (2018).
17. R. Albalat, C. Canestro, Evolution by gene loss. *Nat Rev Genet* **17**, 379-391 (2016).
18. M. Blumer *et al.*, Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Sci Adv* **8**, eabm6494 (2022).
19. F. Marcus, B. Gontero, P. B. Harrsch, J. Rittenhouse, Amino acid sequence homology among fructose-1,6-bisphosphatases. *Biochem Biophys Res Commun* **135**, 374-381 (1986).
20. G. Mayr, A new Eocene swift-like bird with a peculiar feathering. *Ibis* **145**, 382-391 (2003).
21. G. Mayr, Old World fossil record of modern-type hummingbirds. *Science* **304**, 861-864 (2004).
22. M. Uhlen *et al.*, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
23. L. B. Tanner *et al.*, Four Key Steps Control Glycolytic Flux in Mammalian Cells. *Cell systems* **7**, 49-62.e48 (2018).
24. A. Gizak, E. Maciaszczyk-Dziubinska, M. Jurowicz, D. Rakus, Muscle FBPase is targeted to nucleus by its 203KKKGK207 sequence. *Proteins* **77**, 262-267 (2009).

25. P. Huangyang *et al.*, Fructose-1,6-Bisphosphatase 2 Inhibits Sarcoma Progression by Restraining Mitochondrial Biogenesis. *Cell Metab* **31**, 174-188 e177 (2020).
26. R. K. Suarez, J. R. Lighton, G. S. Brown, O. Mathieu-Costello, Mitochondrial respiration in hummingbird flight muscles. *Proc Natl Acad Sci U S A* **88**, 4870-4873 (1991).
27. D. Zhang *et al.*, Metabolic reprogramming of cancer-associated fibroblasts by IDH3alpha downregulation. *Cell Rep* **10**, 1335-1348 (2015).
28. P. Bu *et al.*, Aldolase B-Mediated Fructose Metabolism Drives Metabolic Reprogramming of Colon Cancer Liver Metastasis. *Cell Metab* **27**, 1249-1262 e1244 (2018).
29. T. Yang *et al.*, Enolase 1 regulates stem cell-like properties in gastric cancer cells by stimulating glycolysis. *Cell death & disease* **11**, 870 (2020).
30. C. Dong *et al.*, Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell* **23**, 316-331 (2013).
31. C. S. Chin *et al.*, Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**, 1050-1054 (2016).
32. D. Guan *et al.*, Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896-2898 (2020).
33. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
34. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
35. J. Ghurye *et al.*, Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol* **15**, e1007273 (2019).
36. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
37. A. F. A. H. Smit, R., RepeatModeler Open-1.0. (2008-2010).
38. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0. (1996-2010).
39. R. S. Harris, The Pennsylvania State University, (2007).
40. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**, 11484-11489 (2003).
41. H. G. Suarez, B. E. Langer, P. Ladde, M. Hiller, chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596-1603 (2017).
42. E. Osipova, N. Hecker, M. Hiller, RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**, (2019).
43. V. Sharma, A. Elghafari, M. Hiller, Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation. *Nucleic Acids Res* **44**, e103 (2016).
44. Y. Toda *et al.*, Early origin of sweet perception in the songbird radiation. *Science* **373**, 226-231 (2021).
45. R. T. Kimball, Carl H. Oliveros, Ning Wang, Noor D. White, F. K. Barker, Daniel J. Field, Daniel T. Ksepka, R. T. Chesser, Robert G. Moyle, Michael J. Braun, Robb T. Brumfield, Brant C. Faircloth, Brian T. Smith, and Edward L. Braun, A Phylogenomic Supertree of Birds. *Diversity* **11**, 109 (2019).
46. C. H. Oliveros *et al.*, Earth history and the passerine superradiation. *Proc Natl Acad Sci U S A* **116**, 7916-7925 (2019).
47. P. A. Hosner, B. C. Faircloth, T. C. Glenn, E. L. Braun, R. T. Kimball, Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). *Mol Biol Evol* **33**, 1110-1125 (2016).
48. B. M. Kirilenko *et al.*, Integrating gene annotation with orthology inference at scale. *Science* (in press) https://www.biorxiv.org/content/10.1101/2022.09.08.507143v1, (2022).

49. D. G. MacArthur *et al.*, A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823-828 (2012).
50. R. W. Meredith, J. Gatesy, W. J. Murphy, O. A. Ryder, M. S. Springer, Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet* **5**, e1000634 (2009).
51. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).
52. G. Mayr, Skeletal morphology of the middle Eocene swift Scaniacypselus and the evolutionary history of true swifts (Apodidae). *Journal of Ornithology* **156**, 441-450 (2015).
53. D. T. Ksepka, J. A. Clarke, S. J. Nesbitt, F. B. Kulp, L. Grande, Fossil evidence of wing shape in a stem relative of swifts and hummingbirds (Aves, Pan-Apodiformes). *Proc Biol Sci* **280**, 20130580 (2013).
54. G. J. Dyke, A Primitive Swift from the London Clay and the Relationships of Fossil Apodifom Birds. *Journal of Vertebrate Paleontology* **21**, 195-200 (2001).
55. J. M. Rodriguez *et al.*, APPRIS: selecting functionally important isoforms. *Nucleic Acids Res* **50**, D54-D59 (2022).
56. V. Ranwez, E. J. P. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol Biol Evol* **35**, 2582-2584 (2018).
57. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300 (1995).
58. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
59. A. Dobin, T. R. Gingeras, Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* **51**, 11 14 11-11 14 19 (2015).
60. M. Bernt *et al.*, MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* **69**, 313-319 (2013).
61. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
62. S. Rath *et al.*, MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res* **49**, D1541-D1547 (2021).
63. Y. Zhou *et al.*, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
64. T. F. Scientific, Neon Transfection System for transfecting mammalian cells, including primary and stem cells, with high transfection efficiency. *Waltham, MA, Thermo Fisher Scientific, Inc.*, (2014).
65. I. D. techonologies, Alt-R CRISPR-Cas9 System:User guide for cationic lipid delivery of CRISPR-Cas9 ribonucleoprotein into mammalian cells. *Coraville, IA, Integrated DNA techonologies, Inc.*, (2017).
66. M. Wu *et al.*, Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. *Am J Physiol Cell Physiol* **292**, C125-136 (2007).
67. S. A. Mookerjee, R. L. S. Goncalves, A. A. Gerencser, D. G. Nicholls, M. D. Brand, The contributions of respiration and glycolysis to extracellular acid production. *Biochimica et biophysica acta* **1847**, 171-181 (2015).
68. S. A. Mookerjee, M. D. Brand, Measurement and Analysis of Extracellular Acid Production to Determine Glycolytic Rate. *J Vis Exp*, e53464 (2015).
69. A. E. Carpenter *et al.*, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).
70. M. R. Berthold *et al.* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 319-326.

71. Y. S. Hwang *et al.*, The first whole transcriptomic exploration of pre-oviposited early chicken embryos using single and bulked embryonic RNA-sequencing. *Gigascience* **7**, 1-9 (2018).
72. K. Okajima *et al.*, Mutations of the E1beta subunit gene (PDHB) in four families with pyruvate dehydrogenase deficiency. *Mol Genet Metab* **93**, 371-380 (2008).

Acknowledgment

We thank Ninon Ballerstädt, Gabriel Low, Maria Luisa da Silva, Manfred Gahr, and Carolina Frankl-Vilches for assistance with tissue collection, Henrike Indrischek, Ksenia Kuznetsova, Juliana Roscito, Mihail Sarov, Ilka Reinhard, Arjun Narayanan, Felix Lansing, Aksana Schneider, Anna Shevchenko, Nikolai Hecker and Leon Hilgers for reagents and advice on analyses and experiments. We also thank Gerald Mayr for advice on hummingbird and swift fossil data and the Computer Service Facilities of the MPI-CBG and MPI-PKS and Christoph Sinai for technical support.

Funding:

This work was supported by a sequencing grant from the German Research Foundation (HI1423/4-1), the Dresden Concept Genome Center supported by the German Research Foundation (INST 269/768-1), the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2016-05381), the Max Planck Society, and the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK).

Author contributions: EO generated and analyzed data and performed the experiments. SW sequenced and TB and MP assembled the genome. RB, JJ, CM and MB contributed to experiments. KS, AHG, AM, DLA and MWB collected tissues. MH conceived and supervised the study. EO and MH wrote the initial manuscript. All authors edited and approved the final manuscript.

Competing interests: The authors declare no competing interests.

Data and materials availability: The genome assembly and all sequencing data has been submitted to NCBI under the BioProjects PRJNA785144 (primary assembly) and PRJNA785156 (alternate assembly). RNA-sequencing data is uploaded to NCBI (PRJNA785156). Accession codes and identifiers of publicly available genomic data are listed in Table S1. Please contact the corresponding author for any other materials.

Supplementary Materials

- Materials and Methods
- Figs. S1 to S12
- Notes S1 to S2
- Tables S1-S6

Figure legends

Figure 1: Loss of *FBP2* coincided with the evolution of hovering flight.
(A) Comparison of genome assembly contiguity. N(x%) graphs show contig (left) and scaffold (right) sizes on the Y-axis for which x percent of the assembly consists of contigs and scaffolds of at least that size. Photo: Mats Lindberg.
(B) Status of 11,611 ancestral avian genes in 51 bird assemblies.
(C) Overlap of gene losses identified in four assemblies of three hummingbird species.
(D) Left: Timetree with the estimated time period of *FBP2* loss, which overlaps the period during which hovering flight evolved according to two dated fossils. Right: Visualization of the seven *FBP2* coding exons shows that hummingbirds share several inactivating mutations. Insets show the sequences of the six assemblies.

Figure 2: Hummingbird muscle does not express FBPase enzymes.
(A, B) UCSC Genome Browser screenshot showing the neighboring *FBP1* and *FBP2* genes in the common swift (A) and Anna's hummingbird (B) together with the expression level in liver (blue) and pectoralis muscle (red).
(C, D) Quantification of *FBP1* (C) and *FBP2* (D) expression. A two-sided t-test was used. Error bars represent standard deviation; n = 3 birds.

Figure 3. *FBP2* knockdown upregulates glycolysis and mitochondrial respiration.
(A) Positions of guide RNAs around exon 5 of the Japanese quail *FBP2*. *FBP2* knockdown in quail QM7 myoblast cells was achieved by electroporating the four guide RNAs with Cas9 protein.
(B-C) Quantitative reverse transcription PCR of *FBP2* and *FBP1* in mock-treated wild-type (WT^{mock}, electroporated with a negative control RNA) and *FBP2* knockdown (KD) cells; n = 9.
(D) FBPase activity is reduced in *FBP2* KD compared to WT^{mock} cells.
(E) Cell number-normalized glycolytic Proton Efflux Rate (PER) values at different time points comparing *FBP2* KD with WT^{mock} cells.
(F) Glycolytic PER values at two timepoints representing basal and compensatory glycolysis. Box plots show the first quartile, median and third quartile with whiskers extending up to 1.5 times the interquartile distance; n = 16.
(G) Cell number-normalized Oxygen Consumption Rate (OCR) values comparing *FBP2* KD with WT^{mock} cells. The OCR drop after adding Rot/AA shows that mitochondrial respiration is the main OCR contributor.
(H) Box plots of OCR values before inhibiting mitochondrial respiration; n = 16.
(I) Box plots of qPCR measured ratios of mitochondrial to nuclear DNA indicate a higher number of mitochondria in *FBP2* KD cells; n = 27.
(J) Sum of normalized expression values of mitochondria-encoded genes in pectoralis and liver of Anna's hummingbird and common swift; n = 3 birds.
(K) Volcano plot shows the significance (Y-axis) vs. the magnitude (X-axis) of expression differences between hummingbird and swift muscle for 597 genes that function in mitochondria.

Genes that are significantly up- or downregulated (adjusted p-value < 0.01 and >2-fold expression difference) are in red and blue, respectively. Genes mentioned in the text are underlined.
A two-sided Mann-Whitney U-test was used in panels C, D, F, H, and I; a two-sided t-test was used in panels D and J. Error bars in panels C, D, E, G represent the 95% confidence interval. Error bars in panel J represent standard deviation.

Figure 4: Positive selection and expression upregulation in glycolytic genes in hummingbirds

(A) Main glucose metabolism pathways.
(B) Enzymes that catalyze reactions in the glycolysis and gluconeogenesis. Genes under significant positive selection (false discovery rate (FDR) < 0.05) in the ancestral hummingbird branch are in red. *ENO1* has an FDR just above the threshold (0.0505) and is also shown in red.
(C) Protein alignments of PFKP, ALDOB, PDHB, and ENO1 show amino acid substitutions that likely evolved under positive selection in the ancestral hummingbird branch (Bayes empirical Bayes posterior probabilities >0.5).
(D) Expression level of *PFKP*, *ALDOB*, *PDHA2*, *PDHB*, and *ENO1* in hummingbird tissues compared to swift. Error bars represent standard deviation; n = 3 birds.

Science
AAAS

Supplementary Materials for

Loss of a gluconeogenic muscle enzyme contributed to adaptive metabolic traits in hummingbirds

Ekaterina Osipova[1,2,3,4,5,6], Rico Barsacchi[1], Tom Brown[1,3,7], Keren Sadanandan[8], Andrea H. Gaede[9,10], Amanda Monte[11], Julia Jarrells[1], Claudia Moebius[1], Martin Pippel[1,3], Douglas L. Altshuler[9], Sylke Winkler[1,7], Marc Bickle[12], Maude W. Baldwin[8], Michael Hiller[1,2,3,4,5,6*]

[1] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany,
[2] Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
[3] Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany
[4] LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
[5] Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany
[6] Goethe-University, Faculty of Biosciences, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
[7] DRESDEN concept Genome Center, Technische Universität Dresden, 01062 Dresden, Germany
[8] Evolution of Sensory Systems Research Group, Max Planck Institute for Ornithology, Seewiesen, Germany
[9] University of British Columbia, Vancouver, Vancouver, BC, V6T 1Z4 Canada
[10] Structure and Motion Laboratory, Royal Veterinary College, University of London, London, England, United Kingdom
[11] Department of Behavioural Neurobiology, Max Planck Institute for Ornithology, Seewiesen, Germany
[12] Roche Institute for Translational Bioengineering, Grenzacherstrasse 124, 4070 Basel, Switzerland

Correspondence to: michael.hiller@senckenberg.de

This PDF file includes:

Materials and Methods
Figs. S1 to S12
Notes S1 to S2
Captions for Tables S1 to S6

Other Supplementary Materials for this manuscript include the following:

Tables S1-S6 as sheets in an Excel file

Materials and Methods

Sample collection

For genome sequencing, two adult male long-tailed hermits (*Phaethornis superciliosus*) were captured in the Gunma Ecological Park, an Amazon Rainforest fragment in Brazil (1º13'S, 48º41'W), and surroundings using mist nets. After capture, the birds were deeply anesthetized with an intramuscular injection of ketamine hydrochloride (Cetamin, Rhobifarma Indústria Farmacêutica Ltda, SP, Brazil). Completely unconscious birds were euthanized by decapitation and dissected. Samples of the brain, syrinx and blood plasma were collected and flash frozen at -80℃. All procedures realized in Brazil were conducted in conformity with the instruction n° 03/2014 of the Chico Mendes Institute for Biodiversity Conservation (ICMBio) and with the permission of the Brazilian Institute of Environment and Renewable Natural (IBAMA) under the Biodiversity Information and Authorization System (SISBio) license number 41794-1. The genetic data of the long-tailed hermit presented in this study was obtained in compliance with the Law nº 13.123/2015 of the Ministry of the Environment and Genetic Heritage Management Council and the activities were registered in the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen) under the registration number A8E6064.

For transcriptomics, three common swifts (*Apus apus*) were obtained from an avian rehabilitator in Bavaria, Germany. The three individuals used for the study were terminally injured and unable to be rehabilitated, and were euthanized by an isoflurane overdose 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. After euthanasia, the birds were dissected. Tissue samples were collected from the liver and pectoralis and flash frozen at -80˚C.

Three adult male Anna's hummingbirds (*Calypte anna*) were caught in Vancouver, BC, Canada and housed individually. All experimental procedures were approved by the University of British Columbia Animal Care Committee in accordance with the guidelines set out by the Canadian Council on Animal Care. Birds were anesthetized with isoflurane and euthanized by N_2 asphyxiation. After euthanasia, the birds were dissected and the tissue samples of the liver and pectoralis were collected and stored in RNAlater. Since individuals of common swifts and Anna's hummingbird are wild animals, we cannot control for variation in diet, activity or other environmental factors.

All animal care and user permissions were obtained from the appropriate institutions.

Extraction of ultra-long genomic DNA

Ultra-long genomic DNA (gDNA) from snap-frozen brain tissues of one long-tailed hermit individual was extracted following the Bionano Prep Animal Tissue DNA Isolation Soft Tissue Protocol (Document Number 30077, document revision C). In brief, snap frozen brain tissue was homogenized on ice in a tissue grinder and nuclei were mildly fixed for 1 hour on ice with 100% ice cold ethanol. The fixed homogenized tissue-nuclei mix was embedded into agarose plugs and an enzymatic treatment with Proteinase K and RNAse was applied. Genomic DNA was recovered from agarose plugs by agarase treatment and further purified by drop dialysis against 1x TE buffer. The integrity of the high molecular weight (HMW) gDNA was determined by pulse field gel

electrophoresis using the Pippin PulseTM device (SAGE Science). The majority of the gDNA was between 50 and 500 kb in length. All pipetting steps involving gDNA were done very carefully with wide-bore pipette tips.

PacBio continuous long read (CLR) library preparation and sequencing

Long insert libraries were prepared as recommended by Pacific Biosciences 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 5 µg sheared gDNA was used for library preparation. The PacBio SMRTbell library was size-selected for fragments larger than 26 kb with the BluePippinTM device according to the manufacturer's instructions. The size selected library was run on one Sequel II SMRT cell. This generated a total of 103 Gb of unique insert reads, representing about 85X effective genome coverage.

10x genomic linked read sequencing

Ultra-long gDNA was used for 10x Genomic linked read sequencing following the manufacturer's instructions (10X Genomics Chromium Reagent Kit v2, revision B). Briefly, 0,84 ng of HMW gDNA was loaded into 10X genome GEM droplets (Gel Bead-In-EMulsions = GEM) 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 used for Illumina sequencing library preparation, as described in the protocol. Sequencing was conducted on a NovaSeq 6000 S1 flow cell using the 2x150 cycles paired-end regime plus 8 cycles of i7 index.

Bionano optical mapping

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 322 Gb in molecules larger than 150 kb.

Chromatin conformation capturing (HiC)

Chromatin conformation capturing was done using the ARIMA-HiC High Coverage Kit (Article Nr. A101030-ARI), following the user guide for animal tissues (ARIMA Document, Part Number: A160162 v00). Briefly, 48 mg flash-frozen powdered syrinx tissue was crosslinked chemically. The crosslinked genomic DNA was digested with a cocktail of four 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 used 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 library ran on an S4 flow cell of an Illumina NovaSeq 6000 with 2x150 cycles.

Genome assembly

Initial contigs were assembled from PacBio subreads using Falcon (falcon-kit v1.8.1) and Falcon Unzip (v1.3.7) (*31*). The Config files are provided in Supplementary Note 1 and 2. Remaining haplotypic duplications in the primary contig set were then removed using purge-dups (v.1.2.3) (*32*).

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). 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). Finally, HiC reads were mapped to the resulting scaffolds using bwa-mem (v0.7.17-r1188) (*33, 34*). We then followed the VGP scaffolding pipeline (https://github.com/VGP/vgp-assembly/tree/master/pipeline/salsa) (*14*), which uses salsa2 (v2.2) (*35*). Finally, we manually curated the scaffolds to join those contigs missed by salsa2 and break those joins which were spuriously created.

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". Based on the read-piles created by reads spanning across gap regions, we created a consensus sequence to replace the N sequences in our genome. We used gcpp (version 2.0.2) 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) (*14*). 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, which changed in the first round 232,004 bp and in the second round 40,098 bp. Finally, we used Merqury (v1.0) (*36*) with the 10x data, which estimated the QV of the final assembly to be 55.3.

Modeling and masking repeats

For the newly generated long-tailed hermit assembly, we used RepeatModeler (*37*) to identify repeat families and RepeatMasker (*38*) 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 and other birds, we used LASTZ version 1.04.00 (*39*) with parameters K = 2,400, L = 3,000, Y = 9,400, H = 2,000 and the default scoring matrix, axtChain (*40*), chainCleaner (*41*), and RepeatFiller (*42*) (all with default parameters). For the downstream analyses, we excluded chains with a score less than 100,000.

Ancestral bird genes and gene classification

To obtain ancestral bird genes, we considered 25,218 isoforms of 17,911 genes that are annotated by Ensembl v92 for the *Gallus gallus* assembly (NCBI GCA_000002315.3, UCSC galGal5). To detect gene-inactivating mutations, we ran a previously developed pipeline (*16*) using default parameters. We classified a gene as having an intact reading frame in another bird assembly if the middle 80% of the ORF is present (not overlapping assembly gaps) and lacks inactivating mutations. Genes with inactivating mutations have at least one inactivating mutation in the middle 80% of the ORF, where such mutations rarely occur in conserved genes (*43*). Genes with missing exonic sequence lack at least a part of the middle 80% of the ORF in a bird assembly. We then defined ancestral bird genes as genes that are annotated in the chicken genome (used as the reference) and that have an intact reading frame in a member of Palaeognathae, representing one of the earliest diverging avian clades (*11*). Palaeognathae are represented by the ostrich, kiwi, and tinamou genomes. This resulted in 11,611 genes. The phylogeny shown in Fig. 1B represents the current understanding of evolutionary relationships within birds (*44*) and was compiled from several sources, including a recent phylogenomic supertree (*45*), family-level relationships of passerines (*46*) and data from studies focusing on specific groups (*12, 47*).

Detecting gene losses in hummingbirds

To identify hummingbird-specific gene losses, we applied a stricter filter and defined a gene as lost if at least two inactivating mutations are present in the middle 80% of the ORF. We do not consider mutations as gene-inactivating that are located in the first or last 10% of the ORF, because frameshift or stop codon mutations that occur in truly-conserved genes are heavily biased towards the first and last 10% of the ORF (*43, 48*), where alternative start codons or different stop codons result in truncated or different termini. Since loss of function variants in human populations exhibit a similar pattern (*49*), this suggests that protein termini are under less evolutionary constraint. By considering only mutations in the middle 80% of the ORF, our approach is conservative. Using the same criteria, we then excluded genes that are lost in any other 47 bird assemblies outside of the hummingbird clade. For *FBP2*, we confirmed that the gene has an intact ORF in all 47 non-hummingbird species (Table S3).

Gene loss dating

To date the loss of *FBP2* on the ancestral hummingbird branch, we used the method described in reference (*50*), which estimates the proportion of the loss branch (the branch along which the gene was inactivated) where a gene was functional and where it evolved neutrally. This method requires estimates of the rate of nonsynonymous to synonymous substitutions and lineage divergence times representing the start and the end of the loss branch. The rate of nonsynonymous to synonymous substitutions for functional and pseudogenic branches was estimated with codeml from PAML package (*51*). As described in (*50*), we assumed that the synonymous substitution rate on functional branches is 70% of the synonymous substitution rate on pseudogene branches. The split

of hummingbirds and swifts (Trochilidae – Apodidae) is estimated to have happened ~54 Mya (*11*), which is consistent with fossil data suggesting that the earliest swifts appeared in the early Eocene (56 - 49 Mya) (*52-54*). Multi-locus data from 284 hummingbird species estimated that the split of extant hummingbirds happened ~22 Mya (*12*). The upper and lower bounds of species divergence times, the estimated length of the loss branch, and respective sources are given in Table S4.

Positive selection analysis

Of 50 genes generally involved in glucose metabolism, we identified 46 in the chicken Ensembl v92 annotation (Table S6). For these 46 genes, we extracted the principal isoform from APPRIS (Annotating principal splice isoforms) (*55*). In case APPRIS provided no or multiple principal isoforms for a gene, we used the Ensembl transcript with the longest coding sequence. We used TOGA (Tool to infer Orthologs from Genome Alignments) (*48*) (https://github.com/hillerlab/TOGA) to obtain the orthologous coding sequence in other birds. Multiple codon alignments of orthologous exons were generated with MACSE v2 (*56*) using default parameters. For each gene, we joined all exon alignments, including potential codons that are split at exon boundaries to obtain a full codon alignment. We removed sequences of species that have missing sequence for ≥50% of the alignment columns and removed codon alignment columns that lack a codon for ≥50% of the species. Positive selection was analyzed with the branch-site model of codeml from the PAML package (*51*), selecting the ancestral hummingbird branch as the foreground branch. Model fit was assessed with a likelihood ratio test. We corrected for multiple testing over the 46 genes with the Benjamini–Hochberg method (*57*) and used a false discovery rate threshold of 0.05. We manually inspected the alignments of the four selected genes to confirm that the selection signal is not due to misalignments. Sites shown in Fig. 4C have Bayes empirical Bayes posterior probabilities >0.5.

Homology modeling and protein structure manipulation

Homology modeling was performed with the https://swissmodel.expasy.org/ modeling service (*58*), using the protein sequences of the long-tailed hermit. Structure visualization and manipulation was done in PyMOL v2.5.2. Residues involved in ligand binding were inferred based on the sequence alignment with the corresponding human proteins.

RNA sequencing

Total RNA was isolated from muscle and liver tissues of Anna's hummingbird and the common swift using the RNeasy mini kit (Qiagen). Concentration and the RNA integrity number (RIN) were analyzed using a Bioanalyzer (Agilent 2100). For all samples, RIN values were >7. Using 400 ng total RNA, we enriched mRNA 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 11.5 µl 2x first-strand cDNA synthesis buffer (NEBnext, NEB). Samples were incubated for 13 min at 94°C for chemical fragmentation and directly afterwards used for strand-specific RNA-seq library preparation (Ultra II Directional RNA Library Prep, NEB). 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 using a ratio of 1:0.9, libraries were quantified using the Fragment Analyzer (Agilent) and sequenced on an Illumina NovaSeq 6000 with 2x100 bp reads using an S4 flow cell to an average depth of 30 million read pairs.

Analysis of transcriptomic data

Adapters and low-quality sequences were trimmed 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. Next, reads were mapped to the genomes with STAR (version 2.7.3a) (*59*) with parameters --outFilterMultimapNmax 20 --outFilterMismatchNoverLmax 0.04. Annotations of nuclear-encoded genes were generated with TOGA (*48*) and annotations of the 13 mitochondria-encoded genes were generated with MITOS (*60*). Gene count tables were prepared with htseq-count (v0.13.5) with parameters -m intersection-strict --stranded=reverse. Differential gene expression analysis was done using DESeq2 (R v4.1.1) (*61*) with default parameters, followed by testing for significant differences in the normalized count values with the python package scipy.stats. Of the 13 mitochondria-encoded genes, we analyzed the 12 genes that are polyadenylated and thus reliably detected in our RNA-seq data. We did not analyze *nad6* (*MT-ND6*), as this mitochondrial gene is not polyadenylated. Nuclear genes that have a known function in mitochondria were taken from MitoCarta3.0 (*62*) and filtered for genes with an intact reading frame in both Anna's hummingbird and common swift, resulting in 597 genes. The enrichment analysis was done using metascape.org (*63*). Since expression differences were only analyzed for mitochondrial genes, we used the set of MitoCarta3.0 annotated genes (instead of all genes) as the background and the sets of significantly up- or downregulated genes (corrected p-value < 0.01) as the foreground.

Cell culture

QM7 myoblasts (ATCC, CRL-1962TM) were cultured in 199M Medium with Earl's salts (ATCC, LOT11150059) supplemented with 10% FBS, 10% TPB (Thermofisher Scientific, 18050039) according to the manufacturer's protocol. The incubation conditions were 37°C in an incubator with 5% CO_2. Cells were split and passaged every third day (reaching around ~90% confluency), the passage number for the experiments was kept below 12.

Generation of FBP2 CRISPR/Cas9 knockout in QM7 myoblasts

Guide RNAs for target genome *Coturnix japonica* RefSeq V2.0 (GCF_001577835.2) were designed with CRISPOR tool, targeting the intronic regions around the functionally critical *FBP2* exon 5. The designed sequences were: guide a: 5'-TTTGCTTTCAATAGCTAGCTTGG-3'; guide b: 5'-TACACAAGACTGGTGCTGAAAGG-3'; guide c: 5'-CCGGCCAGCAAATAGTATACTGG-3'; and guide d: 5'-GTCAGCCCCTTGTAAACCTTTGG-3'. Cells were detached using 0.05% Trypsin, washed once with PBS, and electroporated using the Neon Transfection System (Invitrogen, Germany, 10 µl kit, 1300V, 30 ms, one pulse). 50,000 cells per reaction were used. The guide RNA complex was formed by mixing the crRNAs and tracrRNAs followed by forming the ribonuclear protein complex with the Cas9 protein (Alt-R S.p. Cas9 Nuclease V3, 100 µg) based on two protocols of Alt-R CRISPR-Cas9 Neon Transfection System (*64, 65*). The Alt-R CRISPR-Cas9 Negative

Control crRNA #1 served as a negative control, which we used as a comparison in all downstream phenotyping experiments (referred to as WT^{mock} throughout the manuscript). After electroporation, cells were seeded in 6 well plates and incubated for 5 days. The Seahorse Glycolytic Rate Assay, qPCR, mitochondrial number estimation, and lipid quantification experiments were performed at passage numbers below 5 after the electroporation. The FBPase activity assay and doubling time estimation were performed at passage number 13.

Genotyping of QM7 cell pools

Genomic DNA was extracted according to the standard phenol-chloroform DNA extraction protocol. Gene fragments flanking the target site of the respective CRISPR-Cas9 constructs were amplified using Phusion High-Fidelity DNA polymerase (Thermofisher Scientific) with primers forward (5'-GTCAGGGCACCAAAAACCAAC-3') and reverse (5'-GGGCAGTCAAGGGGAGTTCA-3'). To verify amplicon sizes, PCR products were analyzed by 1%-agarose gel electrophoresis.

Reverse transcription qPCR

Reverse transcription qPCR was performed on a LightCycler96. Primers were designed with NCBI Primer blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The NCBI mRNA RefSeq annotation of *Coturnix japonica* (GCF_001577835.2) and *Mus musculus* (GCF_000001635.20) were used as references for quail cells (QM7) and embryos and mouse cells (C2C12, ESC) and tissues respectively. *TBP* (TATA-box binding protein) and *Rpl13a* (ribosomal protein 13a) were used as normalizer genes for the quail and mouse experiments, respectively. Primers designed for the quail were: *TBP* forward primer: 5'-CAGAAACTGGGTTTTCCTGCG-3'; *TBP* reverse primer: 5'- ACAATTCTGGCTCATAGCTGC-3'; *FBP1* forward primer: 5'-GCAATCTTGTGGCAGCAGGT -3'; *FBP1* reverse primer: 5'-TCTCCAATTGCCGGATCAAG -3'; *FBP2* forward primer 2: 5'-CCCTGAGGATGGAAGTTCCCC -3'; *FBP2* reverse primer: 5'-AGGACTCTTCTGACTGGCAGG-3'. Quail embryos were used as the control cDNA to test the expression of *TBP*, *FBP1*, and *FBP2*. Primers designed for the mouse samples were: *Rpl13a* forward primer: 5'- CTGCTCTCAAGGTTGTTCGGCT -3'; *Rpl13a* reverse primer: 5'-CCTTCCGTTTCTCCTCCAGAGT -3'; *FBP1* forward primer: 5'-GCATCGCACAGCTCTATGGT -3'; *FBP1* reverse primer: 5'-ACACAGGTAGCGTAGGACGA -3'; *FBP2* forward primer: 5'-GAAATTCCCCGAGGATGGCA -3'; *FBP2* reverse primer: 5'-CCGGAGCTTGCCATTAGGAC -3';

Seahorse glycolytic rate assay

ECAR and OCR measurements were performed with XF96 Extracellular Flux analyzer (Seahorse Bioscience, North Billerica, MA) as described (*66*). Briefly, cells were seeded in XF96 (V3) polystyrene cell culture plates (Seahorse Bioscience, North Billerica) at 15,000 cells/well density and incubated for 24 hours. Prior to performing an assay, the growth medium in the wells of an XF cell plate was exchanged with the warm standard Glycolytic Rate Assay Medium (AM). Cell plates were incubated in a 37°C/non-CO2 incubator for 60 minutes prior to the start of the assay. All compounds were prepared at appropriate concentrations in the AM and added to the appropriate injection port. All experiments were performed at 37°C. Each measurement cycle consisted of a mixing time of 4 minutes and a data acquisition period of 4 minutes. OCR and

ECAR data points refer to the average rates during the measurement cycles. The resulting ECAR and OCR values were normalized by the number of cells for each well, as described below.

ECAR values were then converted into a proton efflux rate (PER, pmol H^+/min) and the contribution of mitochondrial respiration was subtracted from PER. Following the equation developed in (*67, 68*), the sum of glycolytic and residual PER was calculated using the equation:

$$PER_{glyc+residual} = ECAR_{total} / BP - (10^{(pH-pK1)}) / (1+10^{(pH-pK1)}) \times (\max H^+/O_2) \times (OCR_{total} - OCR_{Rot/AA})$$

Since this PER value still includes the small contribution of exogenous pyruvate and glutamine, and Na^+/H^+ antiport, the residual PER (PER after inhibition of both pathways) was subtracted, resulting in glycolytic PER values that are plotted in Fig. 3. The contribution of glycolysis and mitochondrial respiration to cellular ATP production was calculated using the approach described in https://www.agilent.com/cs/library/whitepaper/public/whitepaper-quantify-atp-production-rate-cell-analysis-5991-9303en-agilent.pdf.

Cell number quantification

To count the number of cells for normalization, cells were fixed in 3.7% paraformaldehyde (PFA)/PBS for 15 mins at room temperature. To stain the nuclei, cells were incubated with 1 µg/mL 4′,6-diamidino-2-phenylindole (DAPI) solution at 4°C overnight. Imaging was performed on an automated spinning disc confocal microscope (Yokogawa CV7000) using a 20x 0.45NA objective. DAPI fluorescence was acquired with laser excitation at 405 nm and an emission bandpass filter BP445/45. Image analysis was performed using the CellProfiler (*69*) software to identify and quantify nuclei number and staining intensities. Data analysis was then performed with KNIME (*70*).

Mitochondrial number estimation

The number of mitochondria was estimated using the ratio of mitochondrial DNA (mtDNA) to nuclear DNA (nDNA). Total DNA was extracted according to the standard phenol-chloroform DNA extraction protocol. To quantify mtDNA and nDNA copy number, the real-time qPCR was performed on a LightCycler96. The primers used to amplify mtDNA were: forward primer: 5'-CGTAGAATACGCTGCCGGAC-3'; reverse primer: 5'-AGCGGAATCGTGGGTATGAA-3'. The primers used to amplify nDNA target the GAPDH gene and the sequences were: forward primer: 5'-GAGGGTAGTGAAGGCTGCTG-3'; reverse primer: 5'-ACCAGGAAACAAGCTTGACG-3'. The relative mtDNA content was calculated using the formula: mtDNA content = $1/2^{\Delta Ct}$, where $\Delta Ct = Ct_{mtDNA} - Ct_{GAPDH}$.

Doubling time estimation

Cells were detached with 0.05% trypsin and the number of cells was estimated with a Luna-II Automated Cell Counter. Cells were then seeded in 6-well culture plates at the same density and cultured under normal incubation conditions for 5 days. The number of cells was estimated every 24 hours. The growth constant µ was estimated using the equation: $\mu = (\log_{10} N1 - \log_{10} N0) / (t1 - t0)$, where t0 and t1 are 24 and 96 hours and N0 and N1 are the number of cells estimated from linear regression at t0 and t1, respectively. The doubling time was estimated from the growth constant using the equation: $G = \log_{10} 2 / \mu$.

FBPase activity assay

Cells were detached with 0.05% trypsin and washed with PBS, followed by centrifugation at 100 rcf (relative centrifugal field) for 10 mins. The supernatant was discarded and pellets were used to measure FBPase activity using the colorimetric Fructose-1,6-Bisphosphatase Activity Assay Kit (abcam, ab273329) according to the manufacture's protocol. The absorbance was measured at OD=450 nm in kinetic mode at 37°C on the Tecan's Spark 20M microplate reader. All measurements were normalized by the amount of total protein, which was measured using the Pierce BCA (bicinchoninic acid) Protein Assay Kit (23225, Thermofisher Scientific) following the manufacture's protocol. The absorbance was measured at OD=562 nm on the Perkin Elmer EnVision plate reader.

Lipid quantification in cells

Lipid deposits were analyzed by lipid staining, using BODIPY and LipidTOX. Cells were seeded in 96-well Griner cell culture plates at a density of 15,000 cells/well and incubated for 24 hours in the standard growth medium supplemented with 4.5 mg/mL fructose. Cells were fixed in 3.7% paraformaldehyde (PFA)/PBS for 15 mins at room temperature. To stain lipid droplets, cells were incubated with BODIPY (4,4-Difluoro-1,3,5,7,8-Pentamethyl-4-Bora-3a,4a-Diaza-s-Indacen) (D3922, Thermofisher Scientific) or LipidTOX Deep Red (H34477, Thermofisher Scientific) solution at a final concentration of 0.5 µg/ml for 30 mins at room temperature. To stain the nuclei, cells were incubated with 1 µg/mL 4',6-diamidino-2-phenylindole (DAPI) solution at 4°C overnight. Imaging was performed on an automated spinning disc confocal microscope (Yokogawa CV7000) using a 60x 1.2NA objective. BODIPY fluorescence was acquired with laser excitation at 488 nm and an emission filter BP525/50. LipidTOX fluorescence was acquired with laser excitation at 647 nm and an emission bandpass filter BP676/29. DAPI fluorescence was acquired with laser excitation at 405 nm and an emission bandpass filter BP445/45. Image analysis was performed using CellProfiler (*69*) to quantify the area of lipid droplets in the cells.

Supplementary Note 1: Config file for running the falcon assembler

```
#### Input
[General]
input_fofn=input.fofn
input_type=raw
pa_DBdust_option=
pa_fasta_filter_option=streamed-median
target=assembly
skip_checks=False
LA4Falcon_preload=false

#### Data Partitioning
pa_DBsplit_option=-x4000 -s200
ovlp_DBsplit_option=-s200

#### Repeat Masking
pa_HPCTANmask_option=-k14 -e0.75 -s100 -l2500 -h240 -w8
#no-op repmask param set
pa_REPmask_code=1,10;0,20;0,300
pa_REPmask_option=-k14 -e0.75 -s100 -l2500 -h240 -w8

####Pre-assembly
# adjust to your genome size
genome_size = 1000000000
seed_coverage = 50
length_cutoff = -1
pa_HPCdaligner_option=-v -B128 -M24
pa_daligner_option= -k14 -e0.75 -s100 -l2500 -h240 -w8
falcon_sense_option=--output-multi --min-idt 0.70 --min-cov 4 --max-n-read 200
falcon_sense_greedy=False

#### Repeat Masking
ovlp_HPCTANmask_option=-k24 -e.90 -s100 -l1000 -h600
#no-op repmask param set
ovlp_REPmask_code=1,10;3,20;0,300
ovlp_REPmask_option=-k24 -e.90 -s100 -l1000 -h600

####Pread overlapping
ovlp_HPCdaligner_option=-v -B128 -M24
ovlp_daligner_option= -k24 -e.90 -s100 -l1000 -h600

####Final Assembly
length_cutoff_pr=2000
overlap_filtering_setting=--max-diff 40 --max-cov 80 --min-cov 2
fc_ovlp_to_graph_option=

[job.defaults]
job_type=slurm
pwatcher_type=blocking
JOB_QUEUE=batch
MB=48768
NPROC=8
njobs=200
submit = srun --wait=0 \
  -p ${JOB_QUEUE} \
  -J ${JOB_NAME} \
  -o "${JOB_STDOUT}" \
```

```
  -e "${JOB_STDERR}" \
  --mem=${MB}M \
  --cpus-per-task=${NPROC} \
  --time=24:00:00 \
  "${JOB_SCRIPT}"

[job.step.da]
NPROC=8
MB=60000
njobs=200
[job.step.la]
NPROC=4
MB=48768
njobs=480
[job.step.cns]
NPROC=8
MB=48036
njobs=480
[job.step.pda]
NPROC=4
MB=60768
njobs=1000
[job.step.pla]
NPROC=4
MB=32768
njobs=480
[job.step.asm]
NPROC=24
MB=196608
njobs=1
```

Supplementary Note 2: Config file for running falcon-unzip

```
[General]
max_n_open_files = 1000

[Unzip]
input_fofn=input.fofn
input_bam_fofn=input_bam.fofn
polish_include_zmw_all_subreads = true

[job.defaults]
job_type=slurm
pwatcher_type=blocking
JOB_QUEUE=batch
MB=48768
NPROC=8
njobs=200
JOB_TIME=24:00:00
submit = srun --wait=0 \
  -p ${JOB_QUEUE} \
  -J ${JOB_NAME} \
  -o "${JOB_STDOUT}" \
  -e "${JOB_STDERR}" \
  --mem=${MB}M \
  --cpus-per-task=${NPROC} \
  --time=${JOB_TIME} \
  "${JOB_SCRIPT}"

[job.step.unzip.track_reads]
JOB_QUEUE=long
njobs=100
NPROC=24
MB=193216
JOB_TIME=96:00:00
# uses minimap2 now
[job.step.unzip.blasr_aln]
njobs=100
NPROC=4
MB=32000
[job.step.unzip.phasing]
njobs=200
NPROC=4
MB=16384
[job.step.unzip.hasm]
njobs=100
NPROC=24
MB=48216
# uses arrow now
[job.step.unzip.quiver]
JOB_QUEUE=long
njobs=200
NPROC=1
MB=36304
JOB_TIME=200:00:00
```

Fig. S1: Strategic selection of a hermit representative and parsimony inference of mutations that occurred in ancestral hummingbirds.
(A) Topology of the hummingbird phylogeny from (*12*), showing all nine strongly-supported principal clades (red font). Swifts represent the direct sister lineage of hummingbirds. The ancestral hummingbird branch is highlighted in red. Species with sequenced genomes are listed in parentheses. We decided to sequence a representative of the hermit clade, since the last common ancestor of a hermit and the already-sequenced Anna's hummingbird and black-breasted hillstar is the ancestor of all hummingbirds.
(B) The parsimony principle infers that a mutation that is shared between all three sequenced hummingbirds was likely already present in their common ancestor, which is the ancestor of all hummingbirds. If that mutation is not shared with swifts and other outgroups, it likely evolved on the ancestral hummingbird branch (red). We illustrate this principle here for the premature stop codon in hummingbird *FBP2* exon 1.

Fig. S2: Putative base errors in Anna's hummingbird PacBio-based assembly.
Anna's hummingbird is represented by two different genome assemblies, one based on short but accurate Illumina reads and one that is based on long but error-prone PacBio Continuous Long Reads (CLR). Our data in Fig. 1B shows that, in comparison to the Illumina assembly, the PacBio assembly possesses less genes with missing sequence, consistent with a higher assembly completeness, but more genes with inactivating mutations. We show the exon structure of three representative genes that possess numerous inactivating mutations in many exons only in the PacBio assembly. Since none of these putative mutations occur in the Illumina assembly or in other hummingbirds, these mutations are likely base errors in the PacBio assembly that escaped polishing with Illumina 10X read data (*14*). Supporting that, most mutations are single base insertion or deletion errors, which is the most frequent base error in PacBio CLR.

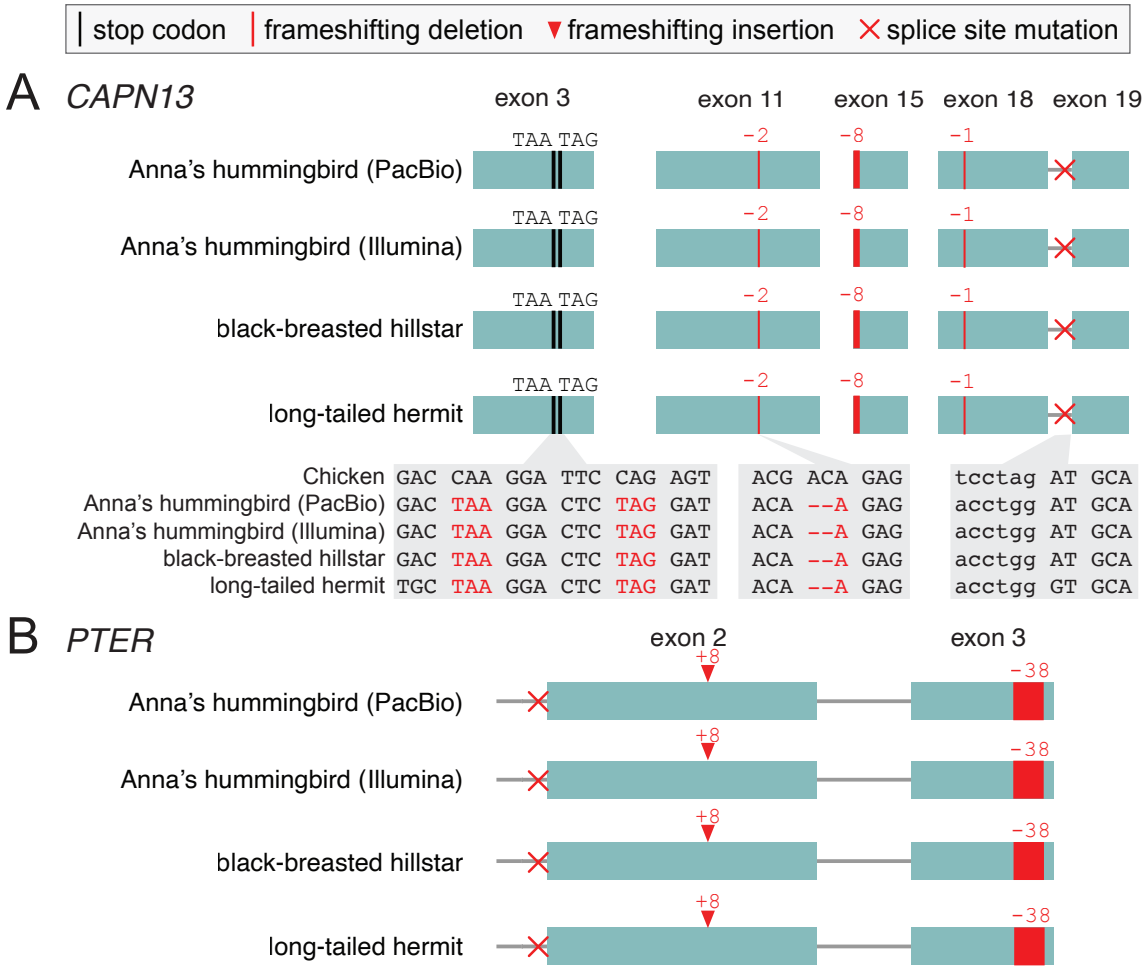

Fig. S3: Shared gene-inactivating mutations in genes identified in our screen.
(A) *CAPN13* has two stop codons in exon 3, a -2 bp frameshifting deletion in exon 11 and a splice site mutation in exon 19 that are shared between the four assemblies of the three hummingbird species. Insets show the mutations (red font). Exonic bases are in upper case, intronic bases are in lower case.
(B) *PTER* exhibits a shared splice site mutation in exon 2 and a shared -38 bp frameshifting deletion in exon 3.

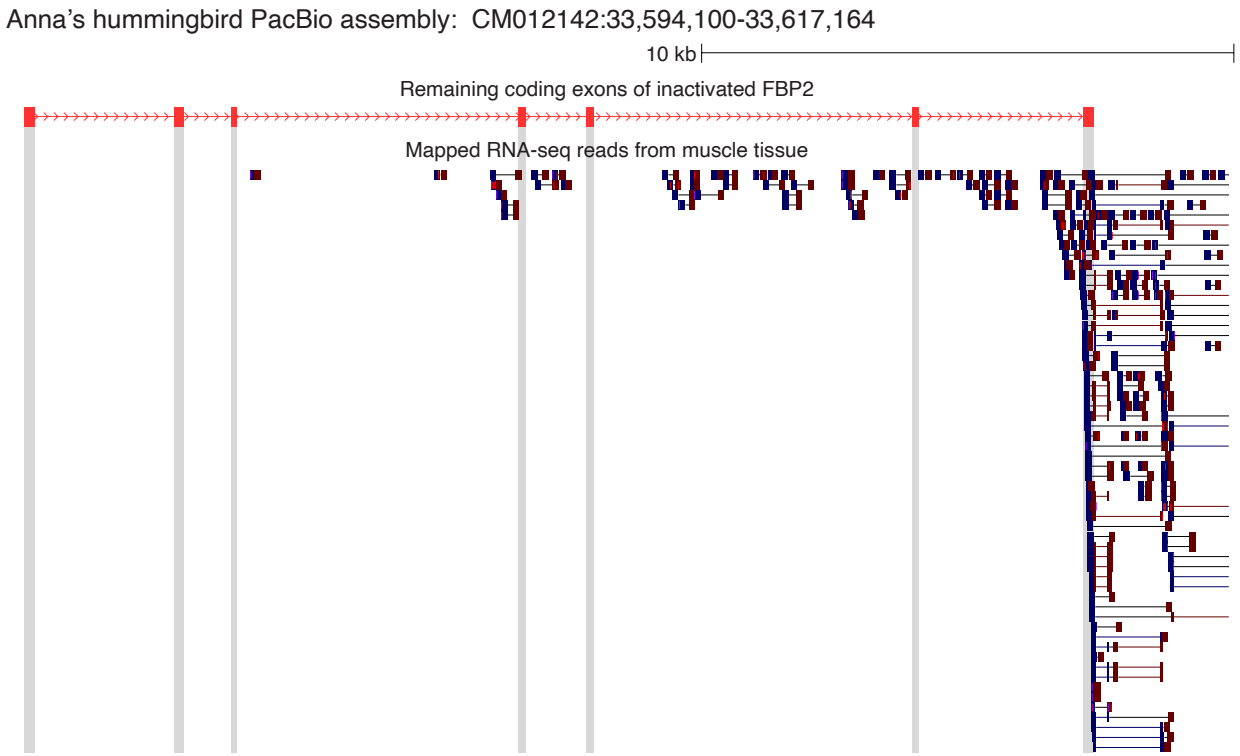

Fig. S4: *FBP2* exhibits no relevant exon expression in Anna's hummingbird muscle tissue. UCSC genome browser view showing the coding exons of the inactivated *FBP2* gene in the genome of Anna's hummingbird. Mapped RNA-seq reads from muscle tissue show that the last (exon 7) *FBP2* exon overlaps a number of reads. However, these reads do not show splicing to any of the upstream exons (highlighted with gray boxes), which are not expressed. Most RNA-seq reads in the gene body upstream of exon 7 are also intronic. Given that large portions of vertebrate genomes are transcribed, we attributed these patterns to background transcription and concluded that *FBP2* exhibits no relevant exon expression in Anna's hummingbird.

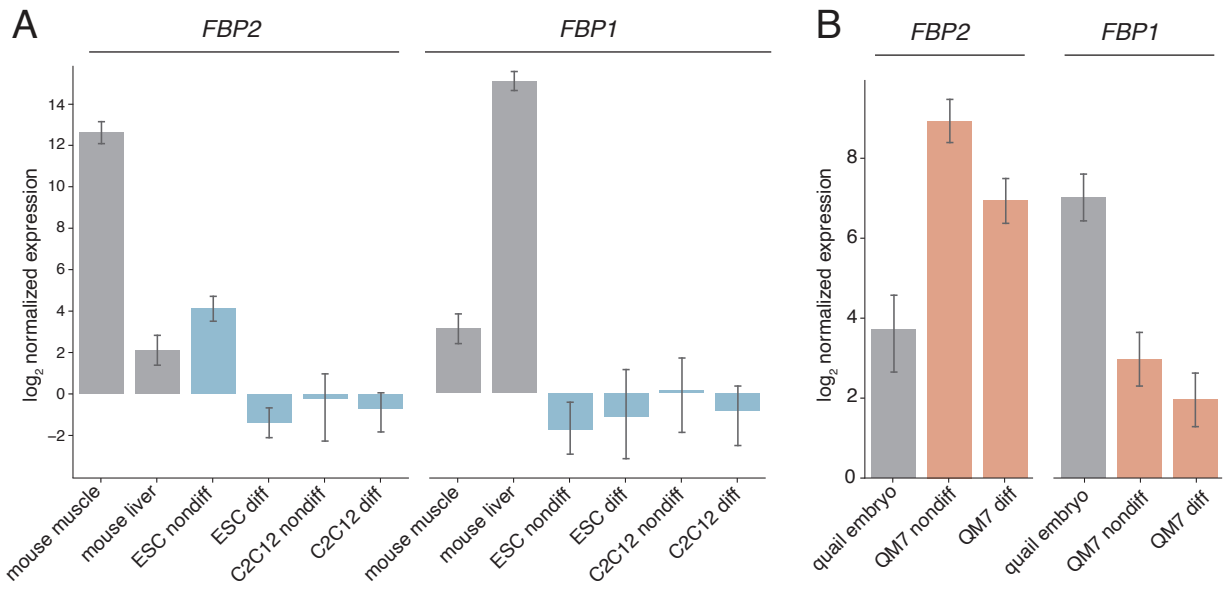

Fig. S5: Expression of *FBP1* and *FBP2* in cell lines, mouse liver and muscle and quail embryos.

(A) Whereas *FBP2* is expressed in mouse muscle and *FBP1* is expressed in mouse liver, as expected, both genes are substantially less expresses in differentiated (diff) and non-differentiated (nondiff) mouse myoblast cells (C2C12) and mouse embryonic stem cells (ESC).

(B) Since bird embryos are known to express both *FBP1* and *FBP2* (*71*), we used quail embryos as a control for expression. Differentiated and in particular non-differentiated quail myoblast cells (QM7) express *FBP2* at a 16-times higher level than *FBP1*. Furthermore, *FBP2* expression in QM7 cells is around 16-times higher than in quail embryos, and comparable or higher than *FBP1* expression in quail embryos.

Error bars represent 95% confidence intervals.

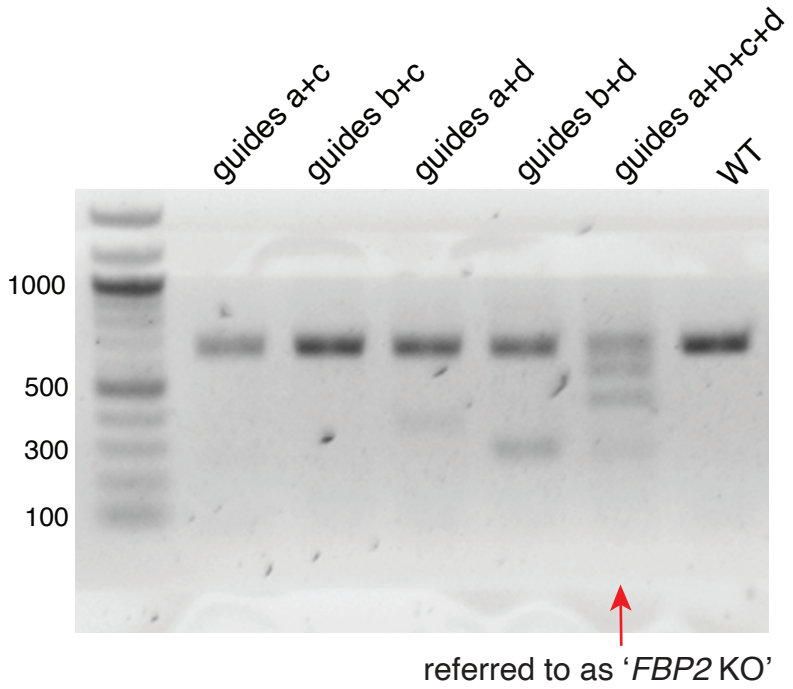

Fig. S6: Genotyping of the pools of *FBP2* knockdown QM7 cells.
Representative gel electrophoresis image shows that the CRISPR-Cas9 electroporation strategy resulted in *FBP2* knockdown (KD), evident by smaller bands representing different-sized deletions spanning exon 5. The highest editing efficiency was observed with the mixture of all four guide RNAs. The size of the lowest band (around 300 bp) corresponds to the deletion induced by the guides a and d.

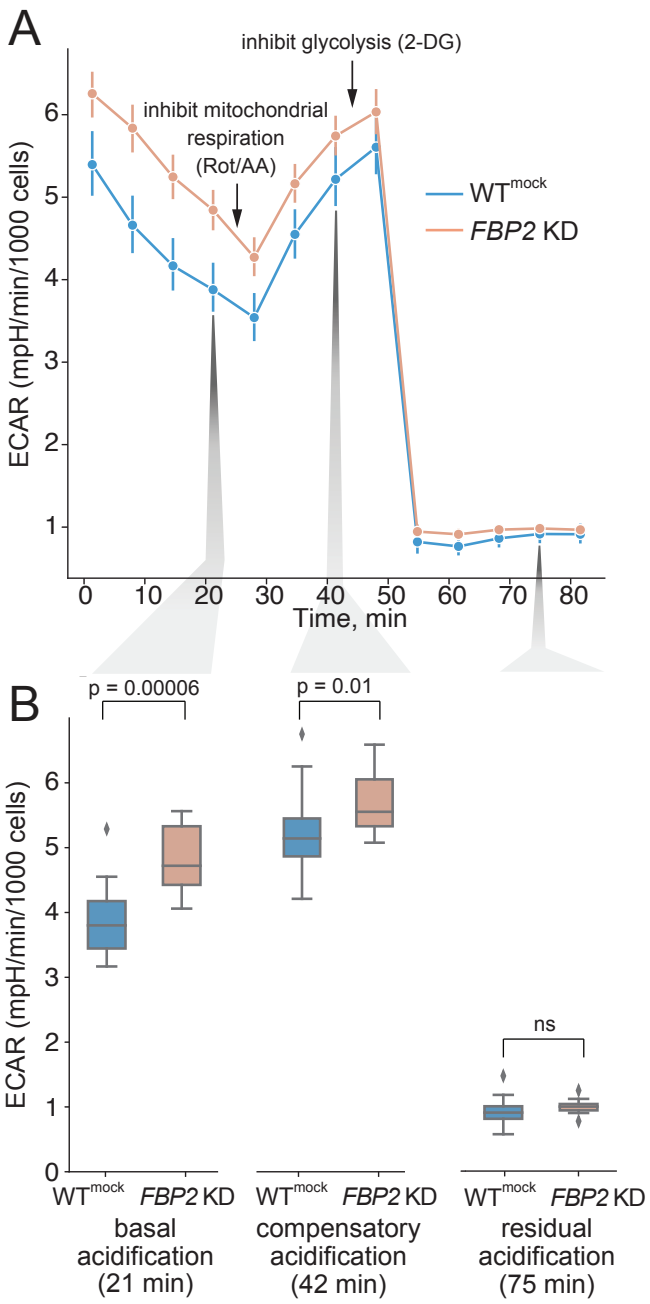

Fig. S7: *FBP2* knockdown upregulates Extracellular Acidification Rate (ECAR).
We converted ECAR into the glycolytic proton efflux rate (PER), which is shown in Fig. 3. This figure shows the Seahorse-measured ECAR values.
(A) Cell number-normalized Extracellular Acidification Rate values at different time points comparing *FBP2* KD with WT^{mock} cells.
(B) ECAR values at three timepoints representing basal, compensatory, and residual acidification. Box plots show the first quartile, median and third quartile with whiskers extending up to 1.5 times the interquartile distance. n = 16. Residual acidification is not significantly different between *FBP2* KD with WT^{mock} cells. A two-sided Mann-Whitney U-test was used.

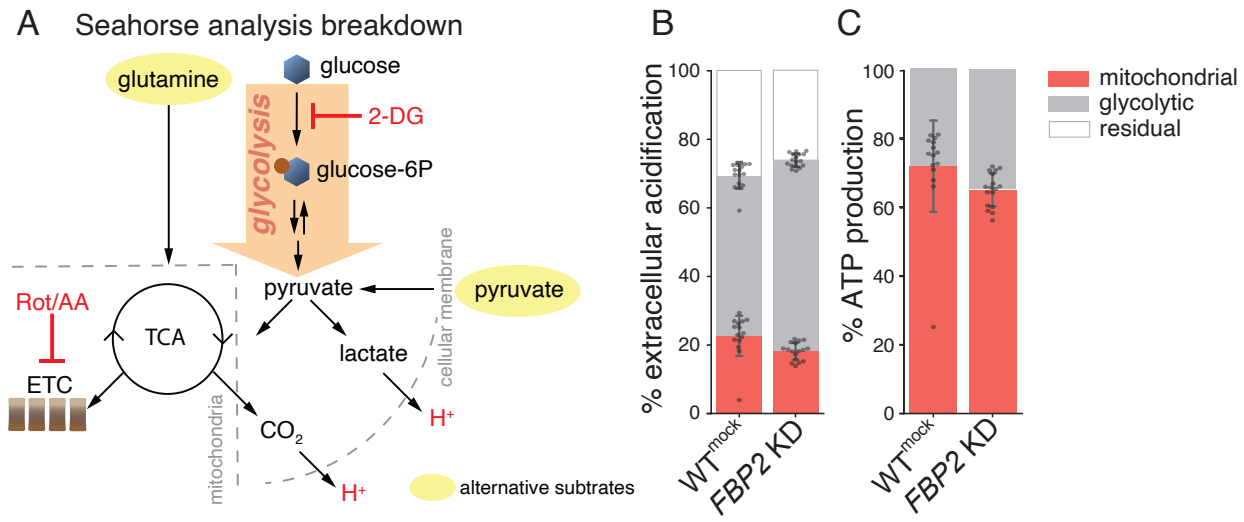

Fig. S8: Mitochondrial respiration contributes significantly to ATP production in QM7 cells.
(A) In the Seahorse assay, the measured extracellular acidification value reflects contributions from multiple sources, including glycolysis and mitochondrial respiration (target pathways of the assay), other sources like glutamine and exogenous pyruvate, and Na+/H+ antiport proteins that transport Na+ into and H+ out of the cell.
(B) Mitochondrial respiration (red) contributes ~20% to the basal extracellular acidification (21 min) in both mock-treated wild-type (WT^{mock}) and *FBP2* knockdown (KD) cells. Residual acidification (white) incorporates all sources of extracellular acidification except glycolysis and mitochondrial respiration.
(C) Mitochondrial respiration is estimated to contribute ~70% and ~65% of the cellular ATP in WT^{mock} and in *FBP2* KD cells, respectively.

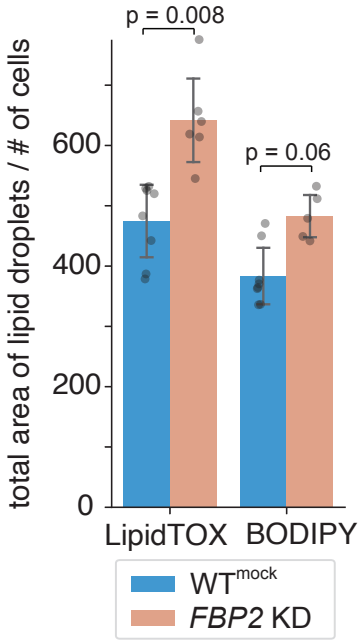

Fig. S9: Lipid staining in QM7 cells.
Total area of lipid droplets, estimated using two lipid dyes (LipidTOX and BODIPY), is higher in *FBP2* knockdown (KD) cells compared to mock-treated wild-type (WT^{mock}), suggesting upregulated lipogenesis. A two-sided Mann-Whitney U-test was used.

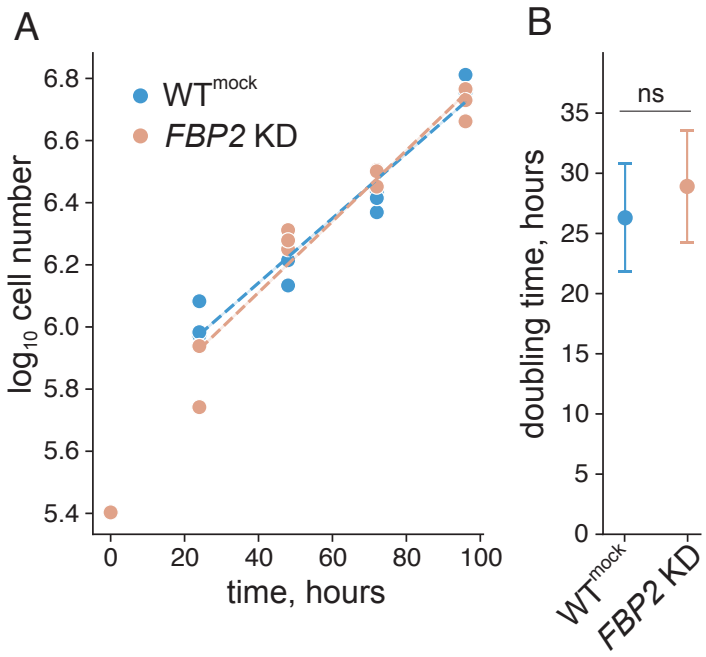

Fig. S10: Cell replication rate is similar between mock-treated wild-type (WT^{mock}) and *FBP2* knockdown (KD) cells.
(A) Measured number of QM7 cells on a log axis. We used the time period 24-96 hours to estimate growth constants by linear regression.
(B) Doubling time estimated from the growth curves does not differ between mock-treated wild-type (WT^{mock}) and *FBP2* knockdown (KD) cells.

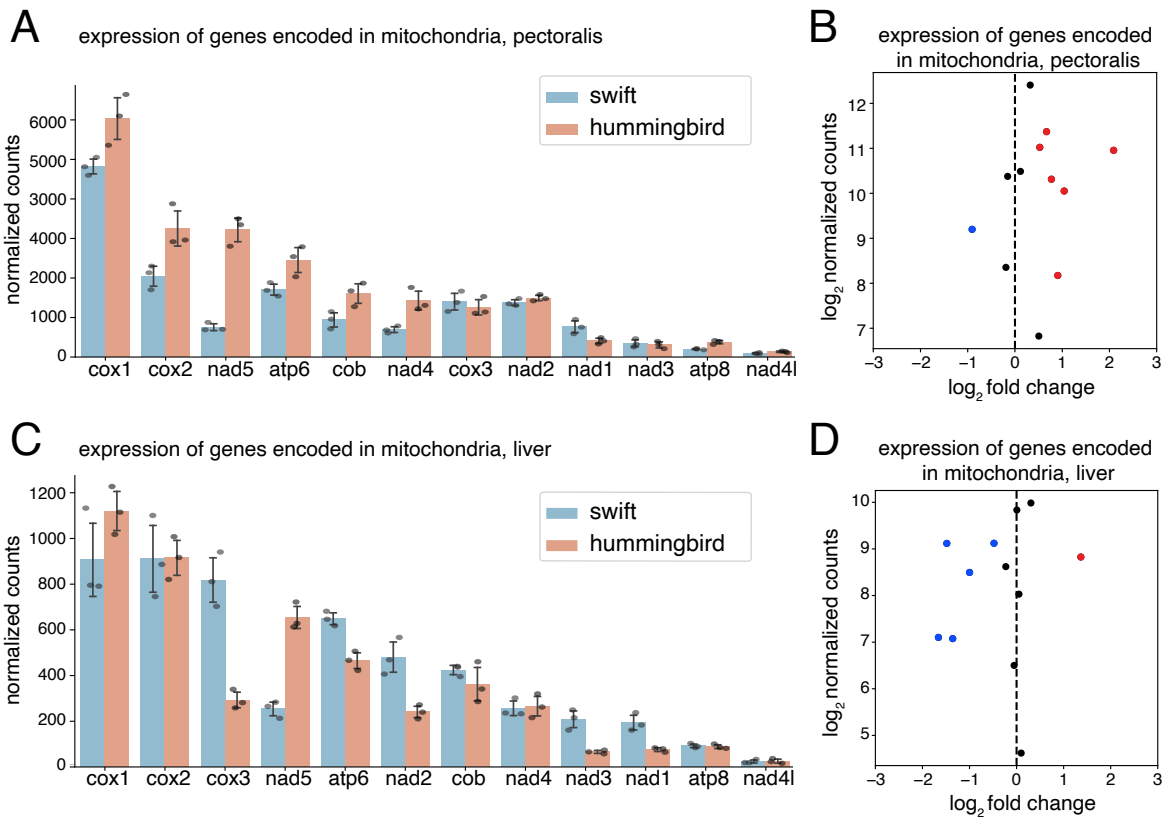

Fig. S11: Mitochondrial gene expression in hummingbird and swift tissues.
(A, C) Expression of 12 mitochondrial-encoded genes in hummingbird pectoralis muscle (A) and liver (C) in comparison to swift. Error bars represent standard deviation; n = 3 birds.
(B, D) Volcano plots show significance vs. magnitude of expression differences in hummingbird pectoralis muscle (B) and liver (D). Genes significantly upregulated in hummingbirds (corrected p-value < 0.01) are colored red. Significantly downregulated genes are in blue.

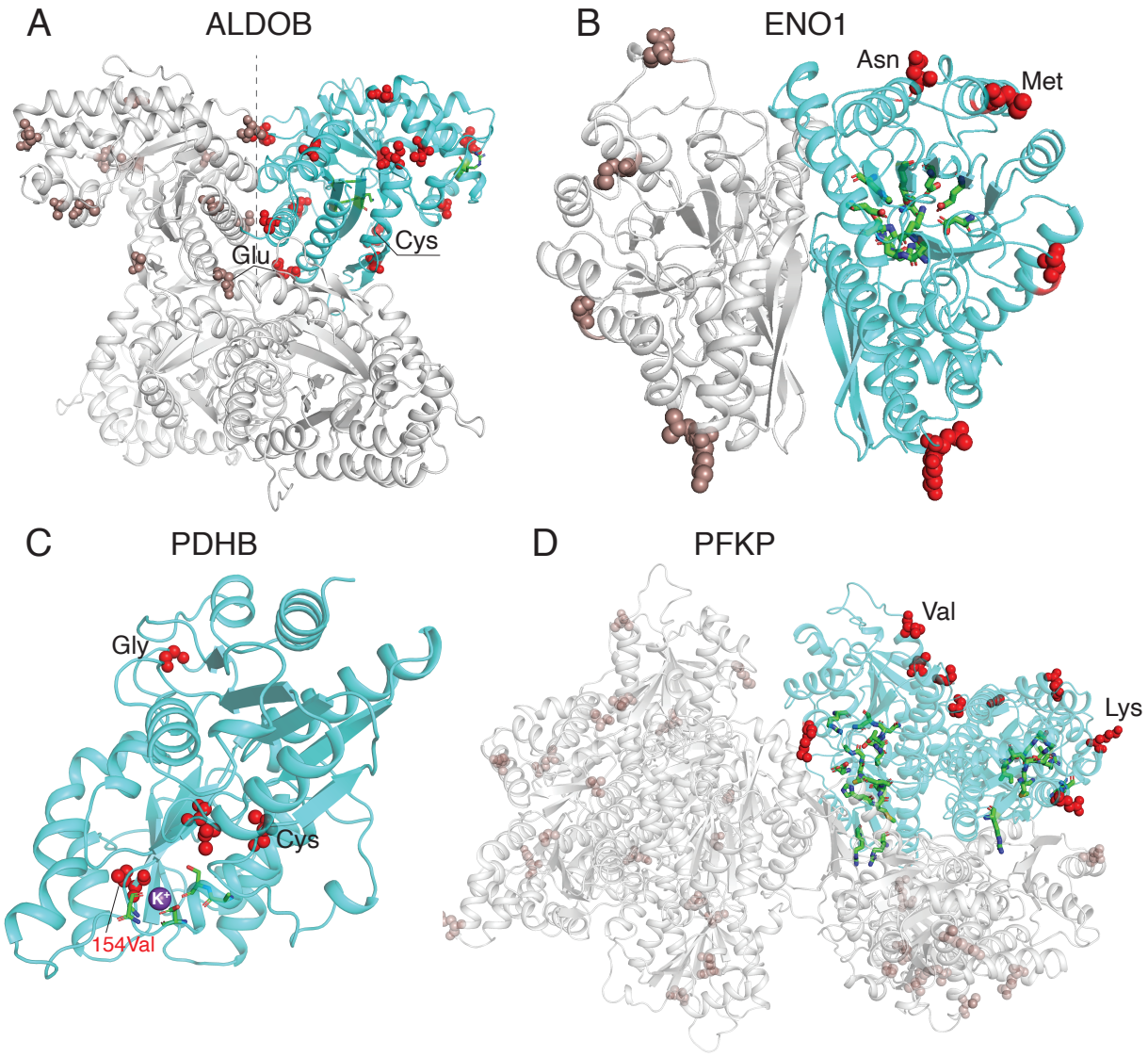

Fig. S12: Structural modeling of products of genes under positive selection.
(A-D) Sites that are likely under positive selection (BEB posterior probabilities >0.5, red spheres) were mapped on the structure models of the long-tailed hermit proteins. Sites involved in ligand binding are depicted with green sticks. The amino acids shown in Fig. 4C are named.
(A) In ALDOB, three amino acids are located on the intersection between the subunits, which could potentially contribute to changes in allosteric regulation.
(B, D) Most residues under selection in ENO1 (B) and PFKP (D) are located on the surface of the proteins.
(C) In PDHB, residue 154 (V in hermit, highlighted in red font) is under positive selection and located in close proximity to a potassium ion (purple); mutating this residue (I142M) in human was reported to effect metal binding and as a consequence, protein stability (*72*).

Table S1. (separate file)
Information of the species and assemblies used in the analysis

Table S2. (separate file)
Classification of ancestral bird genes for all assemblies

Table S3. (separate file)
Status of FBP1 and FBP2 in all species

Table S4. (separate file)
Dating the loss of FBP2 in the stem hummingbird lineage

Table S5. (separate file)
Enrichment analysis of upregulated/downregulated genes with metascape

Table S6. (separate file)
Positive selection analysis with PAML