

1 **Table of Contents of Supplementary Information:**

2

3 **Figures:**

4 **Supplementary Figure S1. Genome sequencing and assembling strategies.....2**

5 **Supplementary Figure S2. K-mer distribution of the *Fulvetta ruficapilla***  
6 **genome.....3**

7 **Supplementary Figure S3. BUSCO completeness of the *Fulvetta ruficapilla***  
8 **assembly in comparison with other avian genomes.....4**

9 **Supplementary Figure S4. Demographic history of *Fulvetta ruficapilla* inferred by**  
10 **the Pairwise Sequentially Markovian Coalescent model.....6**

11 **Supplementary Figure S5. Function enrichment of rapidly evolving genes and**  
12 **positively selected genes identified in the *Fulvetta ruficapilla* genome.....7**

13 **Supplementary Figure S6. Continuity of the *Fulvetta ruficapilla* assembly in**  
14 **comparison with other avian genomes.....9**

15

16 **Tables:**

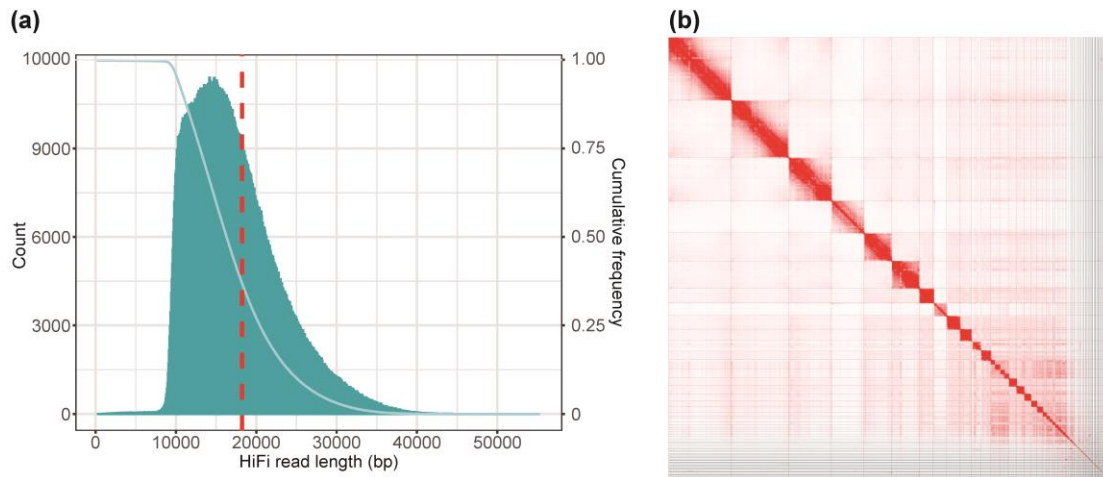
17 **Supplementary Table S1. Summary information of the RNA-seq runs used in gene**  
18 **annotation for the *Fulvetta ruficapilla* assembly. (Please see the supplementary**  
19 **Excel file “Supplementary Table S1.xlsx”)**

20 **Supplementary Table S2. Function enrichment of 87 rapidly evolving genes**  
21 **identified in the *Fulvetta ruficapilla* genome. The enrichment analysis was**  
22 **performed on the online g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>). (Please see the**  
23 **supplementary Excel file “Supplementary Table S2.xlsx”)**

24 **Supplementary Table S3. Function enrichment of 50 positively selected genes**  
25 **identified in the *Fulvetta ruficapilla* genome. The enrichment analysis was**  
26 **performed on the online g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>). (Please see the**  
27 **supplementary Excel file “Supplementary Table S3.xlsx”)**

28

29 **Figures:**



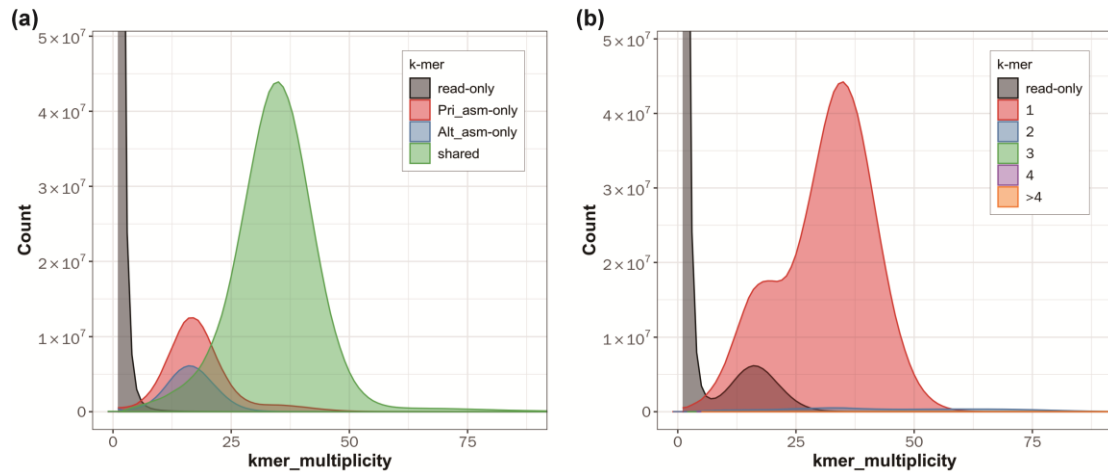
30

31 **Supplementary Figure S1. Genome sequencing and assembling strategies. (a)**

32 Length distribution of HiFi sequencing reads for the *Fulvetta ruficapilla* genome. The  
33 blue histogram indicates the numbers of reads in length. The light blue curve represents  
34 the cumulative length distribution of reads in terms of frequency from the long to short  
35 reads (from the right to left in the figure). The vertical dashed line in red represents the  
36 HiFi read N50 in length. (b) Hi-C contact heatmap of *F. ruficapilla*. Both axes indicate  
37 genomic coordinates from the top left to bottom right. Red dots indicate genomic  
38 regions with a high density of paired Hi-C sequencing reads mapped onto the genome.  
39 Each large square compartment of red dots along the diagonal represents a potential  
40 well-organized scaffold as suggested by Hi-C contact information.

41

42

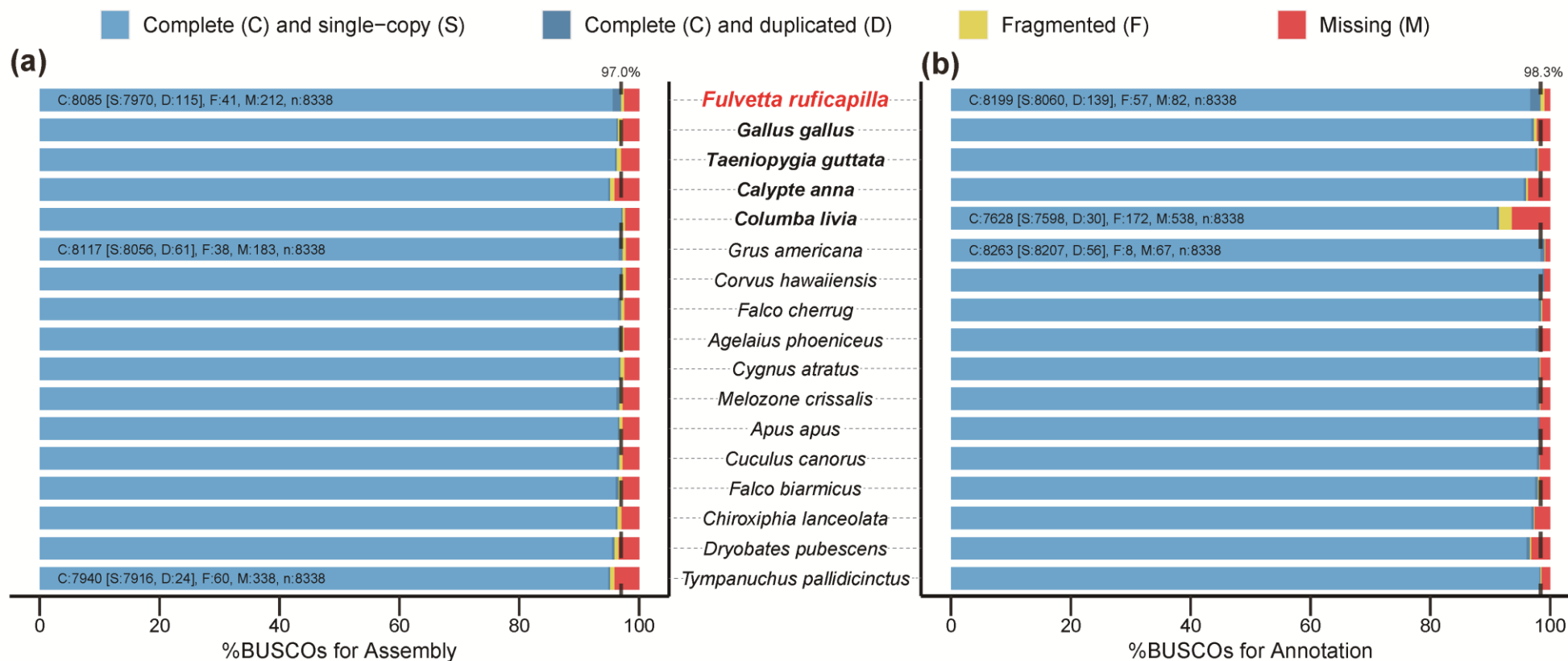


43

44 **Supplementary Figure S2. K-mer distribution of the *Fulvetta ruficapilla* genome.**

45 (a) K-mer evaluation for the primary and alternate assemblies combined. This has  
 46 resulted in a k-mer completeness of 92.93% for the primary assembly while 99.74% for  
 47 the primary and alternate assemblies when considered together. It thus indicates a very  
 48 high-level of assembly completeness for the *Fulvetta ruficapilla* genome in terms of k-  
 49 mer distribution. (b) K-mer evaluation for the primary assembly only after scaffolding.  
 50 It indicates an extremely low level of assembly duplication. The k-mer evaluation also  
 51 indicates barely noticeable assembling errors, as shown by the near-zero bars on the left  
 52 most of each figure (k-mers of zero-depth with counts close to 0), which are reflected  
 53 by the high QV (base quality value) of 62.32 and 59.12 for the primary and alternate  
 54 assemblies, respectively. According to the Merqury pipeline v1.3, a k-mer size of 20  
 55 has been assumed for the evaluation.

56



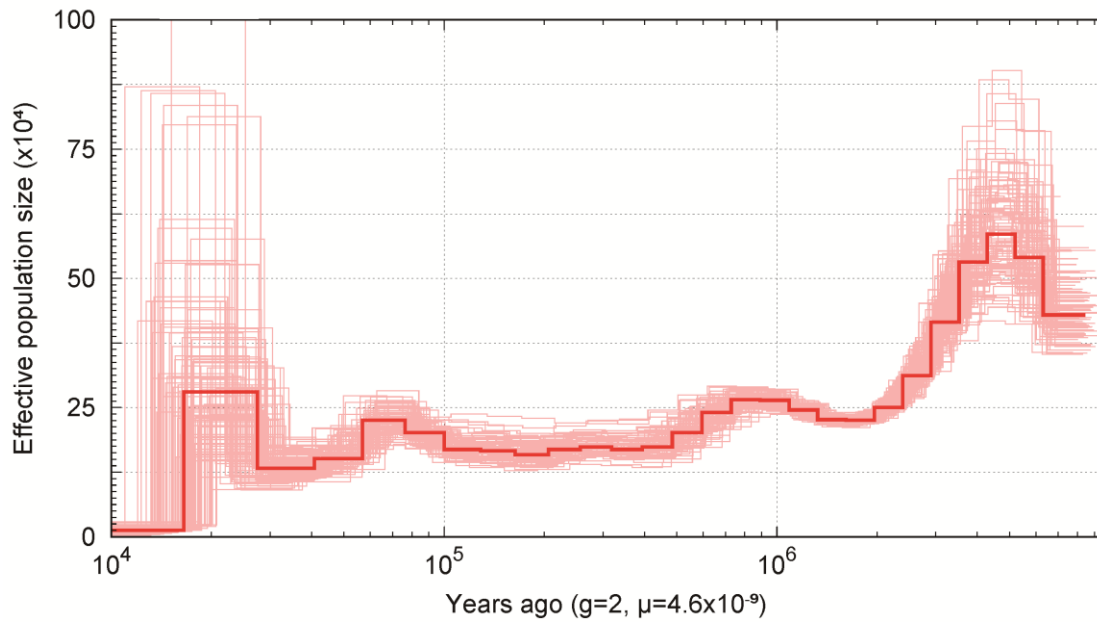
57

58 **Supplementary Figure S3. BUSCO completeness of the *Fulvetta ruficapilla* assembly in comparison with other avian genomes.** (a) is the  
 59 completeness evaluation of the genome assembly while (b) is for the genome annotation of protein-coding genes. All the BUSCOs are assessed  
 60 with the aves\_odb10 lineage database. The vertical dashed lines indicate the BUSCO scores of *F. ruficapilla* (97.0% and 98.3% for assembly and  
 61 annotation, respectively). For clarity, only the highest and lowest BUSCOs along with that of *Fulvetta ruficapilla* are labelled with detailed statistics.

62 The four birds with scientific names in black and bold represent species with their own best-assembled genomes most frequently used as references  
63 in avian genetics and evolutionary analyses. The other 12 birds are selected from the recently released genomes with high quality of VGP and  
64 other projects. Except for *Fulvetta ruficapilla*, all other genomes have been downloaded from the NCBI with accession listed as follows: *Gallus*  
65 *gallus* (chicken, GCF\_016699485.2)<sup>1</sup>, *Taeniopygia guttata* (zebra finch, GCF\_003957565.2)<sup>1</sup>, *Calypte anna* (Anna's hummingbird,  
66 GCF\_003957555.1)<sup>1</sup>, *Columba livia* (rock pigeon, GCA\_032206205.1)<sup>2</sup>, *Grus americana* (whooping crane, GCF\_028858705.1)<sup>1</sup>, *Corvus*  
67 *hawaiiensis* (Hawaiian crow, GCF\_020740725.1)<sup>1</sup>, *Falco cherrug* (saker falcon, GCF\_023634085.1)<sup>1</sup>, *Agelaius phoeniceus* (red-winged blackbird,  
68 GCF\_020745825.1)<sup>3</sup>, *Cygnus atratus* (black swan, GCF\_013377495.2)<sup>4</sup>, *Melospiza crissalis* (California towhee, GCF\_028551555.1)<sup>5</sup>, *Apus apus*  
69 (common swift, GCF\_020740795.1)<sup>1</sup>, *Cuculus canorus* (common cuckoo, GCF\_017976375.1)<sup>1</sup>, *Falco biarmicus* (lanner falcon,  
70 GCF\_023638135.1)<sup>1</sup>, *Chiroxiphia lanceolata* (lance-tailed manakin, GCF\_009829145.1)<sup>1</sup>, *Dryobates pubescens* (downy woodpecker,  
71 GCF\_014839835.1)<sup>1</sup>, *Tympanuchus pallidicinctus* (lesser prairie-chicken, GCF\_026119805.1)<sup>6</sup>.

72

73



74

75 **Supplementary Figure S4. Demographic history of *Fulvetta ruficapilla* inferred by**

76 **the Pairwise Sequentially Markovian Coalescent model.** The generation time and

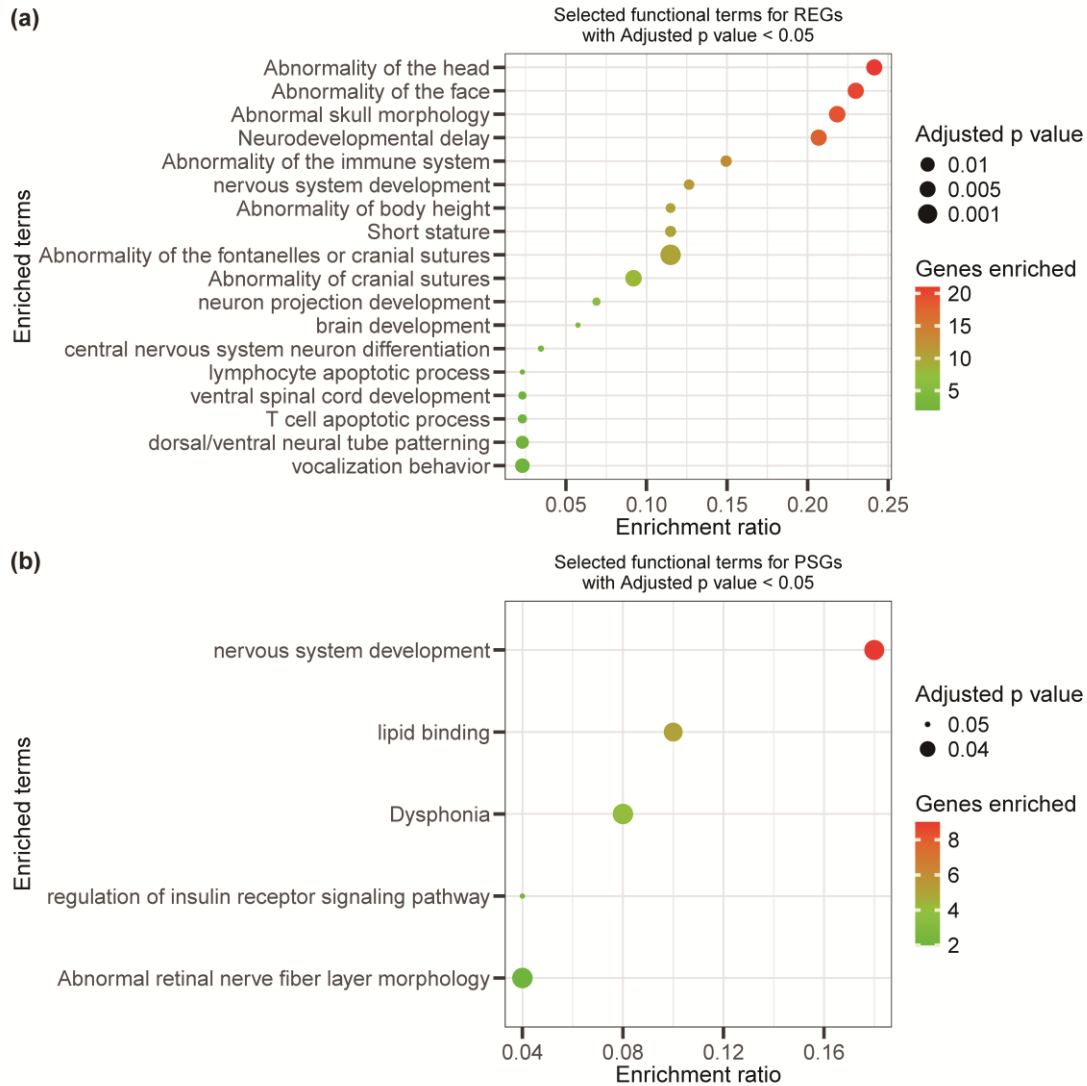
77 mutation rate employed to scale PSMC<sup>7</sup> inference into real years and sizes are 2 years

78 and  $4.6 \times 10^{-9}$  substitutions per site per generation, respectively, which are referenced

79 from the same settings in Leroy et al. (2021)<sup>8</sup>.

80

81



82

83 **Supplementary Figure S5. Function enrichment of rapidly evolving genes and**

84 **positively selected genes identified in the *Fulvetta ruficapilla* genome. (a) selected**

85 **function enrichment results for rapidly evolving genes (REGs). (b) selected function**

86 **enrichment results for positively selected genes (PSGs). It has to be stated that only a**

87 **subset of function terms enriched significantly (vocal learning, nervous system**

88 **development, immune system, etc.) are shown in this figure. The full list of enriched**

89 **function terms and associated genes is available in **Supplementary Tables S2 & S3.****

90 **The enrichment ratio indicates the proportion of REGs/PSGs in that term out of the**

91 **total REGs/PSGs. All the REGs (n=87) and PSGs (n=50) have been identified from**

92 **one-to-one orthologous genes in a 12-way genome alignment, which includes *Fulvetta***

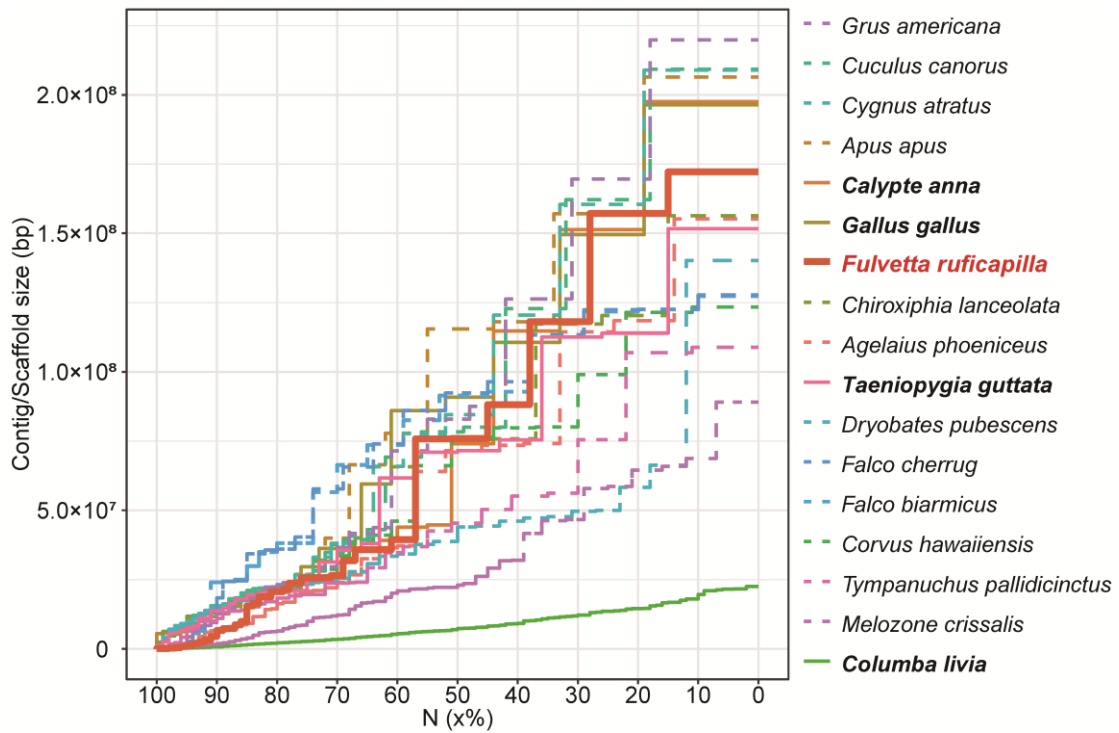
93 ***ruficapilla* ([https://identifiers.org/ncbi/insdc.gca:GCA\\_042477295.1](https://identifiers.org/ncbi/insdc.gca:GCA_042477295.1)), *Apus apus***

94 (common swift, GCF\_020740795.1)<sup>1</sup>, *Colius striatus* (speckled mousebird,  
95 GCF\_028858725.1)<sup>1</sup>, *Corvus moneduloides* (New Caledonian crow,  
96 GCF\_009650955.1)<sup>1</sup>, *Diglossa brunneiventris* (black-throated flowerpiercer,  
97 GCA\_019023105.1)<sup>9</sup>, *Gallus gallus* (chicken, GCF\_016699485.2)<sup>1</sup>, *Lichenostomus*  
98 *cassidix* (helmeted honeyeater, GCA\_008360975.2)<sup>10</sup>, *Melopsittacus undulatus*  
99 (budgerigar, GCF\_012275295.1)<sup>1</sup>, *Myiozetetes cayanensis* (rusty-margined flycatcher,  
100 GCF\_022539395.1)<sup>11</sup>, *Oenanthe melanoleuca* (eastern black-eared wheatear,  
101 GCF\_029582105.1)<sup>12</sup>, *Rhea pennata* (lesser rhea, GCF\_028389875.1)<sup>1</sup>, *Taeniopygia*  
102 *guttata* (zebra finch, GCF\_003957565.2)<sup>1</sup>.

103

104





105

106 **Supplementary Figure S6. Continuity of the *Fulvetta ruficapilla* assembly in**

107 **comparison with other avian genomes. N (x%) indicates the contig/scaffold length**

108 **on the y-axis for which x% of the assembly consists of contigs/scaffolds of at least this**

109 **size. The figure legend on the right shows the lines from top to bottom according to the**

110 **descending length of the longest scaffold in each assembly. Therefore, the N (x%) lines**

111 **start from the right of the x-axis. The genome assemblies evaluated here are same as**

112 **those in Supplementary Fig. S3.**

113

114 **References**

- 115 1. Rhie, A. et al. Towards complete and error-free genome assemblies of all vertebrate  
116 species. *Nature* **592**, 737-746 (2021).
- 117 2. Holt, C. et al. Improved genome assembly and annotation for the rock pigeon  
118 (*Columba livia*). *G3-Genes Genomes Genet.* **8**, 1391-1398 (2018).
- 119 3. *NCBI GenBank* [https://identifiers.org/ncbi/insdc.gca:GCA\\_020745825.3](https://identifiers.org/ncbi/insdc.gca:GCA_020745825.3) (2023).
- 120 4. *NCBI GenBank* [https://identifiers.org/ncbi/insdc.gca:GCA\\_013377495.2](https://identifiers.org/ncbi/insdc.gca:GCA_013377495.2) (2022).
- 121 5. *NCBI GenBank* [https://identifiers.org/ncbi/insdc.gca:GCA\\_028551555.1](https://identifiers.org/ncbi/insdc.gca:GCA_028551555.1) (2023).
- 122 6. Black, A.N. et al. A highly contiguous and annotated genome assembly of the lesser  
123 prairie-chicken (*Tympanuchus pallidicinctus*). *Genome Biol. Evol.* **15**, evad043  
124 (2023).
- 125 7. Li, H. & Durbin, R. Inference of human population history from individual whole-  
126 genome sequences. *Nature* **475**, 493-496 (2011).
- 127 8. Leroy, T. et al. Island songbirds as windows into evolution in small populations.  
128 *Curr. Biol.* **31**, 1303-1310 (2021).
- 129 9. Hiller, A.E., Brumfield, R.T. & Faircloth, B.C. A reference genome for the nectar-  
130 robbing Black-throated Flowerpiercer (*Diglossa brunneiventris*). *G3-Genes*  
131 *Genomes Genet.* **11**, jkab271 (2021).
- 132 10. Robledo-Ruiz, D.A. et al. Chromosome-length genome assembly and linkage map  
133 of a critically endangered Australian bird: the helmeted honeyeater. *Gigascience* **11**,  
134 giac025 (2022).
- 135 11. *NCBI GenBank* [https://identifiers.org/ncbi/insdc.gca:GCA\\_022539395.1](https://identifiers.org/ncbi/insdc.gca:GCA_022539395.1) (2022).
- 136 12. Peona, V. et al. An annotated chromosome-scale reference genome for Eastern  
137 black-eared wheatear (*Oenanthe melanoleuca*). *G3-Genes Genomes Genet.* **13**,  
138 jkad088 (2023).

139