



## Abundant Phenotypic Parallelism, Incomplete Lineage Sorting, and Introgression in Open-Habitat Chats

Journal:	<i>Molecular Biology and Evolution</i>
Manuscript ID	MBE-22-0548
Manuscript Type:	Article
Date Submitted by the Author:	23-Jun-2022
Complete List of Authors:	Alaei Kakhki, Niloofar; Friedrich Schiller University Jena, Department of Population Ecology Schweizer, Manuel; Natural History Museum of Bern; University of Bern Institute of Ecology and Evolution Lutgen , Dave; University of Bern Institute of Ecology and Evolution; Friedrich Schiller University Jena, Department of Population Ecology; Swiss Ornithological Institute Bowie, Rauri ; University of California Berkeley, Museum of Vertebrate Zoology; University of California Berkeley, Department of Integrative Biology Shirihai , Hadoram ; Natural History Museum of Bern Suh, Alexander; University of East Anglia, School of Biological Sciences; Uppsala University Evolutionary Biology Centre, Department of Organismal Biology – Systematic Biology (EBC), Science for Life Laboratory Schielzeth, Holger; Friedrich Schiller University Jena, Department of Population Ecology; German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig Burri, Reto; Swiss Ornithological Institute; University of Bern Institute of Ecology and Evolution; Friedrich Schiller University Jena, Department of Population Ecology
Key Words:	birds, colouration, gene tree heterogeneity, hybridization, phenotypic evolution, phylogenomics
Subject Section:	Discoveries

SCHOLARONE™  
Manuscripts

1  
2     1 Article, Discoveries  
3  
4  
5

## 6     2 **Abundant Phenotypic Parallelism, Incomplete Lineage 7     3 Sorting, and Introgression in Open-Habitat Chats**

8  
9     5 **Niloofar Alaei Kakhki<sup>1</sup>, Manuel Schweizer<sup>2,3</sup>, Dave Lutgen<sup>3,1,4</sup>, Rauri C. K. Bowie<sup>5,6</sup>,**  
10    6 **Hadoram Shirihai<sup>2</sup>, Alexander Suh<sup>7,8</sup>, Holger Schielzeth<sup>1,9</sup>, Reto Burri<sup>4,3,1</sup>**

11    7  
12    8  
13    9     <sup>1</sup> Department of Population Ecology, Institute of Ecology and Evolution, Friedrich-Schiller-  
14    10 University Jena, Jena, Germany

15    11     <sup>2</sup> Natural History Museum Bern, Bern, Switzerland

16    12     <sup>3</sup> Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

17    13     <sup>4</sup> Swiss Ornithological Institute, Sempach, Switzerland

18    14     <sup>5</sup> Museum of Vertebrate Zoology, University of California, Berkeley, USA

19    15     <sup>6</sup> Department of Integrative Biology, University of California, Berkeley, USA

20    16     <sup>7</sup> School of Biological Sciences, University of East Anglia, Norwich, United Kingdom

21    17     <sup>8</sup> Department of Organismal Biology – Systematic Biology (EBC), Science for Life Laboratory,  
22    18     Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

23    19     <sup>9</sup> German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Leipzig,  
24    20     Germany

### 34    21 35    22 36    23 **Correspondence**

37    24     Niloofar Alaei Kakhki, Department of Population Ecology, Institute of Ecology and Evolution,  
38    25     Friedrich-Schiller-University Jena, Dornburger Str. 159, D-07743 Jena, Germany  
39    26     Email: [niloofar.alaei.kakhki@uni-jena.de](mailto:niloofar.alaei.kakhki@uni-jena.de)

40    27     Reto Burri, Swiss Ornithological Institute, Seerose 1, CH-6204 Sempach.  
41    28     Email: [reto.burri@vogelwarte.ch](mailto:reto.burri@vogelwarte.ch)

## 29 Abstract

30 Insights into the processes underpinning the evolution of phenotypic parallelism contribute to  
31 our understanding of the contributions of ancestral, introgressed, and novel genetic variation to  
32 phenotypic evolution. Phylogenomic analyses characterizing genome-wide gene tree  
33 heterogeneity can provide first clues about the extent of ILS and of introgression and thereby into  
34 the potential of these processes or (in their absence) the need to invoke novel mutations to  
35 underpin phenotypic parallelism. Here, we were interested in understanding the trajectories  
36 along which phenotypic parallelism may have evolved in open-habitat chats (wheatears of the  
37 genus *Oenanthe* and their relatives). To this end, based on whole-genome resequencing data from  
38 50 taxa of 44 species, we established the species tree, characterized gene tree heterogeneity, and  
39 investigated the footprints of ILS and introgression within the latter. The species tree  
40 corroborates the evidence for abundant phenotypic parallelism, especially in wheatears. The high  
41 levels of gene tree heterogeneity in wheatears are explained by ILS alone only for 30% of internal  
42 branches. For multiple branches with high gene tree heterogeneity, D-statistics and phylogenetic  
43 networks identified footprints of introgression. Finally, long branches without extensive ILS  
44 between clades sporting similar phenotypes provide suggestive evidence for a role of novel  
45 mutations in the evolution of these phenotypes. Together, our results suggest that phenotypic  
46 parallelism in open-habitat chats evolved along diverse trajectories and highlight that phenotypic  
47 diversification is often complex and best depicted as a network of interacting lineages.

## 48 Introduction

49 Molecular phylogenetics has unveiled many previously unknown examples of phenotypic  
50 parallelism – here defined as a pattern in which non-sister species spread across the phylogeny  
51 are phenotypically more similar to each other than to their respective sister species. Under such  
52 an evolutionary outcome, species relationships based on morphometrics, coloration, behavior, or  
53 other ecological traits are discordant with the history of descent reflected in the species tree  
54 (Aliabadian et al. 2012; Elmer and Meyer 2011; Jarvis et al. 2014; Martin and Orgogozo 2013;  
55 Paterson et al. 2020; Schweizer et al. 2019a; Schweizer et al. 2019b; Stern 2013). While the many  
56 observations of such discordances across diverse species assemblages witness of the abundance  
57 of phenotypic parallelism, insights into the underlying processes remain more elusive.

58 Phylogenetic information from genome sequencing data now provides unprecedented  
59 power to consolidate patterns of phenotypic parallelism and obtain insights into the processes  
60 underlying them. Many published examples of putative phenotypic parallelism are yet derived  
61 from interpretations of traits mapped across phylogenies using a single or a restricted number of  
62 genetic markers (Aliabadian et al. 2012; Brusatte et al. 2015; Colosimo et al. 2005; Cresko et al.  
63 2004; Stern 2013). Since phylogenetic relationships at different positions in the genome, referred  
64 to as 'gene trees', can vary substantially, many gene trees inevitably deviate from the species'  
65 history of descent reflected in the species tree (Degnan and Rosenberg 2006; Toews and Brelsford  
66 2012). Hence, the mismatch of single gene trees with phenotypic similarities alone does not  
67 provide conclusive evidence for phenotypic parallelism (Degnan and Rosenberg 2006; Doyle  
68 1997; Lamers et al. 2012). Confirming instances of phenotypic parallelism, therefore, requires  
69 species tree reconstructions from genome-wide variation. Once the evidence for phenotypic  
70 parallelism is corroborated by the species tree, we can move on to investigate processes  
71 underlying gene tree heterogeneity that, in part, can underpin phenotypic parallelism.

72 Phenotypic parallelism can evolve in three ways: First, under phenotypic parallelism *sensu*  
73 *stricto*, phenotypic similarities evolve through independent mutations in the same or different  
74 genes (Arendt and Reznick 2008; Christin et al. 2010; Gompel and Prud'homme 2009; Martin and  
75 Orgogozo 2013; Stern 2013). In Mexican cavefish (*Astyanax mexicanus*), for instance, the  
76 phenotypic parallelism of decolored brown phenotypes and albinism in separate caves occurred  
77 through different mutations in the MC1R and OCA2 genes, respectively (Gross et al. 2009; Protas  
78 et al. 2006; Stahl and Gross 2015). Similarly, isoforms of PEPC found in C4 photosynthesis  
79 (Besnard et al. 2009; Christin et al. 2007), and similar floral traits important for pollination  
80 (Hoballah et al. 2007; Preston and Hileman 2009; Whittall et al. 2006) have evolved multiple times  
81 independently in plants.

82 The second, and likely most frequent process leading to phenotypic parallelism that also  
83 accounts for most gene tree heterogeneity is incomplete lineage sorting (ILS), that is, the retention

of alleles and traits that were already present in the ancestral lineage (Colosimo et al. 2005; Cresko et al. 2004; Stern 2013; Van Belleghem et al. 2018). ILS is especially prevalent in radiations characterized by large effective population sizes and fast succession of speciation events, such as, for instance, in the evolution of neoavian birds (Jarvis et al. 2014; Suh 2016; Suh et al. 2015). In such cases, a high proportion of ancestral variation may be retained over subsequent species splits and segregate in the independently evolving gene pools of daughter species (Maddison 1997). Selection or drift in non-sister species may fix the same genotype (and phenotype), while sister species may fix a different genotype/phenotype. For instance in Humans, at least 1% of the genome is genetically more similar to orangutans than to chimps due to ILS, even though these primates are characterized by small effective population sizes (Hobolth et al. 2011).

Third, in hybridizing lineages, phenotypic parallelism and gene tree heterogeneity may be underpinned by introgression that mingles genotypes and phenotypes among species (Heliconius Genome Consortium 2012; Malinsky et al. 2018; Song et al. 2011; Stryjewski and Sorenson 2017). In particular, introgression between non-sister species may result in these species being phenotypically more similar than they are to their respective sister species, such as exemplified by wing-pattern mimicry in *Heliconius* butterflies (Edelman et al. 2019; Pardo-Diaz et al. 2012) and plumage coloration of *Munia* finches (Stryjewski and Sorenson 2017) and of members of the Black-eared Wheatear (*Oenanthe hispanica*) complex (Schweizer et al. 2019a). Importantly, in an increasing number of instances, such as in *Heliconius* butterflies, Yellowstone wolves, Darwin's finches and cichlid fish in Lake Malawi and Lake Victoria, introgression has exchanged alleles between species and resulted in the formation of beneficial phenotypes (Enciso-Romero et al. 2017; Genner and Turner 2012; Grant et al. 2005; Lamers et al. 2012; Wallbank et al. 2016). Given that over the last decade genomic studies have contributed increasing evidence for the abundance of such adaptive introgression, hybridization may underpin phenotypic parallelism more often than previously appreciated (Campagna et al. 2017; Han et al. 2017; Marques et al. 2019a; Meier et al. 2018).

Multiple factors influence which of these routes specific cases of phenotypic parallelism have most likely taken. These factors include the evolutionary time scale under consideration, the speed at which successive speciation events occurred, effective population sizes, and the opportunity for genetic exchange according to biogeographic history. Waiting times for beneficial mutations are long (Barrett and Schlüter 2008; Hedrick 2013; Hermisson and Pennings 2005), such that independent mutations with the same phenotypic effect are expected to be exceedingly rare (Eyre-Walker and Keightley 2007) and only over the course of millions of years may occur in sufficient number to be a source of phenotypic parallelism (Hedrick 2013). Therefore, at short evolutionary timescales, phenotypic parallelism may more often involve the recruitment of standing genetic variation (Barrett and Schlüter 2008), notably from the pool of ancestral

1  
2  
3 120 variation segregating in extant species, or variation introgressed from other species (Stern 2013).  
4 121 Among young evolutionary lineages, ancestral variation is still segregating, especially if speciation  
5 122 events occurred rapidly. Phenotypic parallelism in such cases may occur via ILS (Jones et al. 2012;  
6 123 Wu et al. 2018). Furthermore, among young species, reproductive isolation may still be  
7 124 incomplete. Provided species overlap in time and space, the exchange of genetic variation through  
8 125 introgression may lead to phenotypic parallelism, too.  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

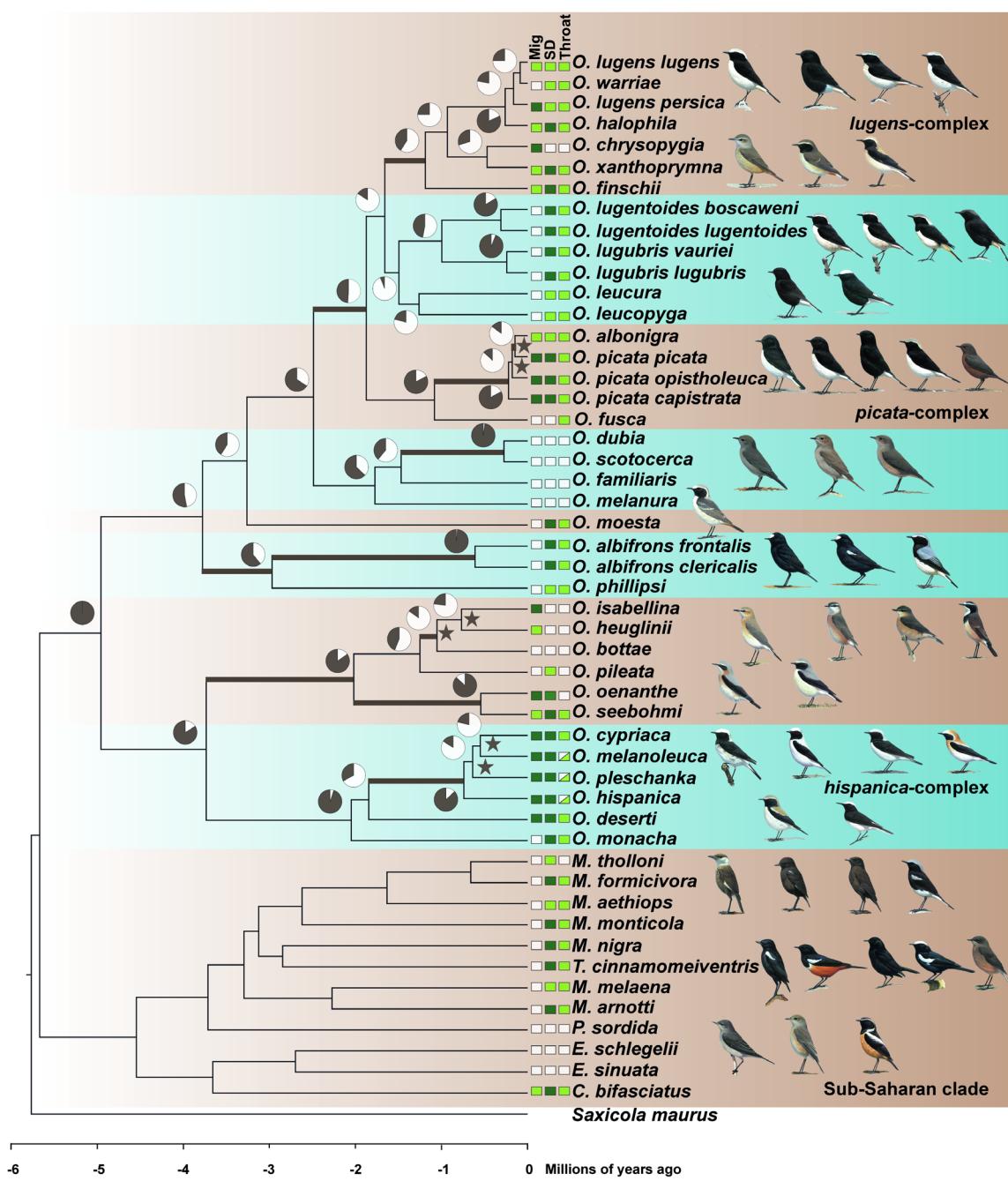
126 Although convincing examples of phenotypic parallelism include all three evolutionary  
127 trajectories, the incidence at which phenotypic parallelism occurs along each of these remains  
128 largely unknown; not least, because identifying the process involved in a particular case of  
129 phenotypic parallelism is difficult. Doing so requires mapping the genomic region underlying the  
130 phenotype in question and tracing the evolution of this genomic region across the phylogeny of  
131 the species involved. In the wild, this is often not possible.

132 However, phylogenomics can provide important indirect insights into the potential  
133 contribution of the abovementioned alternative processes that can underlie the development of  
134 phenotypic parallelism: First, the species tree provides first clues on whether speciation events  
135 occurred over short enough time scales for ancestral variation to be passed to descent lineages  
136 and thus remain incompletely sorted in important proportions beyond speciation events. Second,  
137 insights into the extent of ILS and presence of introgression can be gained from levels of gene tree  
138 heterogeneity (Degnan and Rosenberg 2006; Funk and Omland 2003; Jarvis et al. 2014; Nater et  
139 al. 2015; Suh 2016; Suh et al. 2015) and symmetries of gene tree frequencies (Hibbins and Hahn  
140 2022). Gene tree heterogeneity is high under both ILS and introgression, but the two processes  
141 leave different proportions of alternative gene trees, based on which they can be distinguished  
142 (Hibbins and Hahn 2022; Sayyari and Mirarab 2018; Sayyari et al. 2018). In the presence of  
143 extensive ILS or of introgression, a parsimonious approach attributes the source of phenotypic  
144 parallelism to these processes, even though independent mutations cannot be excluded as the  
145 source of phenotypic parallelism (Colosimo et al. 2005; Cresko et al. 2004; Pardo-Diaz et al. 2012;  
146 Stryjewski and Sorenson 2017). The absence of detecting these processes, conversely, would  
147 indirectly suggest novel mutations as a potential source of phenotypic parallelism. Therefore,  
148 surveys of gene tree heterogeneity and symmetries of gene tree proportions represent a  
149 promising avenue to probe the potential of the alternative processes to contribute to phenotypic  
150 parallelism.

151 Here, we reconstructed the species tree and assessed the contribution of ILS and  
152 introgression to gene tree heterogeneity in open-habitat chats (genera *Campicoloides*,  
153 *Emarginata*, *Myrmecocichla*, *Oenanthe*, *Pinarochroa* and *Thamnolaea*), a monophyletic group of  
154 songbirds displaying a high incidence of phenotypic parallelism (Aliabadian et al. 2012; Mayr and  
155 Stresemann 1950; Schweizer et al. 2019a; Schweizer et al. 2019b). The phylogenetic relationships

1  
2  
3 156 among open-habitat chats as inferred from mitochondrial data were entirely unexpected from a  
4 morphological perspective (Aliabadian et al. 2012). Species similar in plumage coloration and  
5 other traits were often spread far apart across the mitochondrial phylogeny, suggesting that their  
6 phenotypic similarity may have evolved in parallel (Aliabadian et al. 2012; Outlaw et al. 2010;  
7 Schweizer and Shirihai 2013; Schweizer et al. 2019a; Schweizer et al. 2019b). For a limited subset  
8 of species studied, genome-wide variation (ddRAD data) confirmed the mitochondrial  
9 relationships and thus the high incidence of phenotypic parallelism (Schweizer et al. 2019b).  
10 Furthermore, hybridization resulted in substantial introgression in the *Oenanthe hispanica*  
11 complex (Schweizer et al. 2019a) and is suspected to have played a role in phenotypic and species  
12 evolution of other groups of wheatears, such as the *O. picata* complex (Panov 2005). In these  
13 instances, introgression between non-sister taxa may well explain phenotypic parallelism.  
14 However, genomic data is essential to refine the species tree and assess the incidence ILS and/or  
15 introgression across open-habitat chats.  
16  
17 169 Based on whole-genome resequencing data from 50 taxa of 44 open-habitat species (**Tab.**  
18 **1**), we aimed to obtain insights into the potential roles of alternative processes in driving  
19 phenotypic parallelism in these songbirds. To this end, we (i) reconstructed the species tree, (ii)  
20 estimated gene tree variation across the genome, and (iii) explored ILS and introgression as  
21 drivers of the underlying high gene tree heterogeneity. Our results reveal a comprehensive picture  
22 of open-habitat chat evolution involving multiple instances of introgression. Footprints of ILS and  
23 introgression as well as considerable divergence times between the main clades of wheatears with  
24 phenotypic parallelism suggest that a combination of ILS, introgression and novel mutations,  
25 depending on the time scales involved most likely explains the phenotypic parallelism observed  
26 in wheatears.  
27  
28 179  
29  
30 180 **Results**  
31  
32  
33 181 **Sampling, Nuclear Data Preparation, and Mitogenome Assembly**  
34  
35 182 To achieve an almost complete taxon sampling, we resequenced the genomes of 50 open-habitat  
36 chat taxa from 44 out of 47 currently recognized species (**Fig. 1**; **Tab. 1**). Data from *Saxicola*  
37 *maurus* was included as an outgroup for open-habitat chats (Sangster et al. 2010; Zuccon and  
38 Ericson 2010). We mapped the sequencing reads to the reference genome assembly of *Oenanthe*  
39 *melanoleuca* (Peona et al. in prep.) and followed GATK best practices for nuclear data preparation.  
40 Mapping efficiency was not correlated to the degree of evolutionary divergence from the reference  
41 genome, but data obtained from DNA extracted off museum skins mapped at a lower percentage  
42 (linear model,  $d_{XY}$ :  $t=-0.41$ ,  $p=0.68$ ;  $tissue_{museum}$ :  $t=-6.56$ ,  $p<0.001$ ;  $R^2=0.53$ ). After mapping,  
43 sequencing coverage ranged from  $4.6 \times$  to  $40.6 \times$ , with an average coverage of  $12.2 \times \pm 6.2 \times$  (**Tab.**  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 191 1). We extracted mitochondrial sequence data for all 13 protein-coding genes and two rRNA  
4 genes using MitoFinder 1.2 from the sequencing data.  
5  
6  
7  
8



51 **Figure 1 | Time-calibrated phylogenetic tree of open-habitat chats and levels of ILS.** All nodes are  
52 supported by bootstrap values of 100. Pie charts depict the gene tree heterogeneity for each internal  
53 branch, with the brown proportion indicating the proportion of concordant gene trees (gCF). Coloured  
54 branches indicate internal branches for which ILS alone is statistically sufficient to explain the observed  
55 gene tree heterogeneity. Stars indicate branches that are in the phylogenetic anomaly zone. The  
56 character states of three selected characters: Sexual dimorphism (SD), monomorphic female-type  
57 (white), monomorphic male-type (pale green), dimorphic (dark green); Migratory behaviour (Mig),  
58 sedentary (white), short-distance migrant (pale green), long-distance migrant (dark green); and throat  
59 coloration (throat), white (white), black (pale green), and polymorphic (white and pale green). Drawing  
60 courtesy of Chris Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc.

**Table 1 | Sampling and sequence data information.**

Taxon	Collection, no. <sup>a</sup>	Tissue	Library Type	Mapping %	Coverage
<i>Campicoloides bifasciatus</i>	MVZ, RSA073,	Muscle	Illumina DNA Prep Kit	99.1	8.5
<i>Emarginata schlegelii</i>	FMNH, 453197,	Muscle	Illumina DNA Prep Kit	99.1	9.5
<i>Emarginata sinuata</i>	UWMB, 95470	Muscle	ThruPLEX DNA-Seq Kit	96.7	9.7
<i>Myrmecocichla aethiops</i>	MVZ, 129113	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	93.3	6.8
<i>Myrmecocichla arnotti</i>	FMNH, 468111	Toepad	Illumina DNA Prep Kit	99.3	12.8
<i>Myrmecocichla formicivora</i>	MVZ, RSA205	Muscle	Illumina DNA Prep Kit	98.6	8.3
<i>Myrmecocichla melaena</i>	A1153	Blood	ThruPLEX DNA-Seq Kit	98.8	13.3
<i>Myrmecocichla monticola</i>	NMBE, 1043860	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	94.3	7.6
<i>Myrmecocichla nigra</i>	NRM, 570041	Toepad	ThruPLEX DNA-Seq Kit	94.5	7.0
<i>Myrmecocichla tholloni</i>	YPM, ORN95640	Toepad	ThruPLEX DNA-Seq Kit	92.1	7.3
<i>Oenanthe (C.) dubia</i>	FMNH, 83201	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	93.0	6.8
<i>Oenanthe (C.) familiaris</i>	MVZ, G0866	Blood	Illumina DNA Prep Kit	99.4	13.3
<i>Oenanthe (C.) fusca</i>	YPM, ORN011707	Toepad	ThruPLEX DNA-Seq Kit	86.4	4.7
<i>Oenanthe (C.) melanura</i>	A1203	Blood	ThruPLEX DNA-Seq Kit	98.8	15.4
<i>Oenanthe (C.) scotocerca</i>	LACM, 61131	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	96.6	8.3
<i>Oenanthe (M.) albifrons clericalis</i>	NRM, 558941	Toepad	ThruPLEX DNA-Seq Kit	95.4	6.3
<i>Oenanthe (M.) albifrons frontalis</i>	KU, 115365	Muscle	ThruPLEX DNA-Seq Kit	98.5	12.7
<i>Oenanthe albonigra</i>	IR-KIL-010	Blood	ThruPLEX DNA-Seq Kit	99.2	12.9
<i>Oenanthe bottae frenata</i>	NRM, 558917	Toepad	ThruPLEX DNA-Seq Kit	97.0	6.0
<i>Oenanthe chrysopygia</i>	IR-FIR-002	Blood	ThruPLEX DNA-Seq Kit	98.9	11.5
<i>Oenanthe cypriaca</i>	19e	Blood	Chromium Genome Library kit	99.8	40.6
<i>Oenanthe deserti</i>	MO-BOULMANE-2013	Blood	ThruPLEX DNA-Seq Kit	99.3	12.6
<i>Oenanthe finschii</i>	IR-ESF-004	Blood	ThruPLEX DNA-Seq Kit	98.9	8.7
<i>Oenanthe halophila</i>	3Y42902	Blood	Illumina DNA PCR-free	99.6	17.4
<i>Oenanthe heuglinii</i>	ZFMK, H.II.16p2.α	Dry skin	ACCEL-NGS 1S DNA Library Prep Kit	93.6	5.7

1						
2						
3	<i>Oenanthe hispanica</i>	E-GUI-013	Blood	Chromium Genome Library kit	99.8	15.5
4	<i>Oenanthe isabellina</i>	GR-LES-001	Blood	ThruPLEX DNA-Seq Kit	99.2	9.9
5	<i>Oenanthe leucopyga leucopyga</i>	A1137	Blood	ThruPLEX DNA-Seq Kit	98.8	10.7
6	<i>Oenanthe leucura leucura</i>	E-MAT-2012	Blood	ThruPLEX DNA-Seq Kit	99.0	9.5
7	<i>Oenanthe lugens lugens</i>	9b	Blood	ThruPLEX DNA-Seq Kit	98.7	22.0
8	<i>Oenanthe lugens persica</i>	ZMUC, 137759	Muscle	ThruPLEX DNA-Seq Kit	98.6	10.4
9	<i>Oenanthe lugentoides lugentoides</i>	NHMUK, 1965.M.12140	Toepad	ThruPLEX DNA-Seq Kit	96.9	4.6
10	<i>Oenanthe lugentoides boscaweni</i>	NHMUK, 1977.M.21.36	Toepad	ThruPLEX DNA-Seq Kit	97.2	6.8
11	<i>Oenanthe lugubris lugubris</i>	A1129	Blood	ThruPLEX DNA-Seq Kit	99.0	12.5
12	<i>Oenanthe lugubris vauriei</i>	AMNH, 461151	Toepad	ThruPLEX DNA-Seq Kit	95.9	5.3
13	<i>Oenanthe melanoleuca</i>	IT-GRA-006	Blood	Chromium Genome Library kit	99.7	12.6
14	<i>Oenanthe moesta</i>	A1109	Blood	ThruPLEX DNA-Seq Kit	98.8	12.0
15	<i>Oenanthe monacha</i>	A1174	Blood	ThruPLEX DNA-Seq Kit	99.3	17.2
16	<i>Oenanthe oenanthe</i>	GEO-VAR-002	Blood	ThruPLEX DNA-Seq Kit	99.1	12.2
17	<i>Oenanthe phillipsi</i>	YPM, ORN035210	Toepad	ThruPLEX DNA-Seq Kit	95.6	7.9
18	<i>Oenanthe picata capistrata</i>	ZMUC, 29495	Toepad	ThruPLEX DNA-Seq Kit	96.7	5.3
19	<i>Oenanthe picata opistholeuca</i>	ZMUC, 29578	Toepad	ThruPLEX DNA-Seq Kit	97.1	5.2
20	<i>Oenanthe picata picata</i>	IR-TAN-005	Blood	ThruPLEX DNA-Seq Kit	98.8	14.2
21	<i>Oenanthe pileata</i>	TCWC, 15606	Muscle	Illumina DNA Prep Kit	98.9	13.4
22	<i>Oenanthe pleschanka</i>	CN-XS-006	Blood	Chromium Genome Library kit	99.8	15.1
23	<i>Oenanthe seebohmi</i>	KA69373	Blood	ThruPLEX DNA-Seq Kit	99.2	11.2
24	<i>Oenanthe warriae</i>	12c	Blood	ThruPLEX DNA-Seq Kit	98.7	13.4
25	<i>Oenanthe xanthoprymna</i>	NHMO, 15188	Blood	ThruPLEX DNA-Seq Kit	99.0	12.9
26	<i>Pinarochroa sordida</i>	YPM, ORN80066	Toepad	ThruPLEX DNA-Seq Kit	95.4	6.5
27	<i>Thamnolaea cinnamomeiventris</i>	NRM, 20086147	Muscle	ThruPLEX DNA-Seq Kit	98.6	15.9

<sup>a</sup> AMNH: American Museum of Natural History; NHMUK: Natural History Museum, Tring; FMNH: Field Museum of Natural History; LACM: Natural History Museum of Los Angeles County; MVZ: Museum of Vertebrate Zoology, UC Berkeley; NHMO: Natural History Museum, University of Oslo; NRM: Naturhistoriska riksmuseet, Stockholm; TCWC: Texas A&M University Biodiversity Research and Teaching Collections; UWBM: University of Washington Burke Museum; YPM: Yale Peabody Museum; ZMUC: Zoological Museum, Natural History Museum of Denmark; ZFMK: Zoologisches Forschungsmuseum König. Samples for which no institution is indicated are part of the research group's collection.

1  
2  
3 200 To ensure that results did not depend on filtering strategy, all analyses were run with four sets of  
4  
5 201 differently filtered data (see Material and Methods).  
6  
7

## 202 Species Tree Reconstruction Based on Nuclear Genomic Data

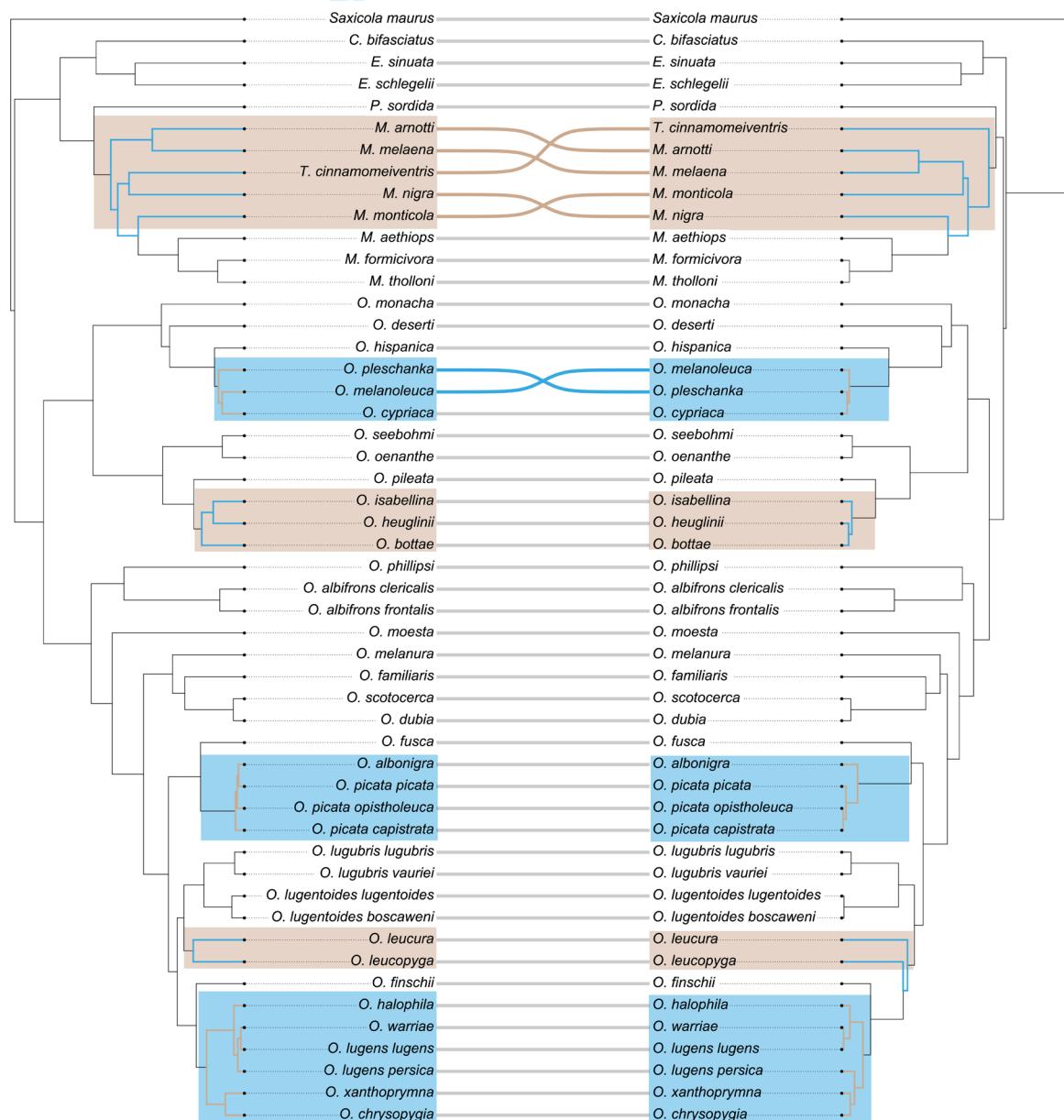
203 We first set out to reconstruct and root the species tree based on regions of the genome least likely  
204 affected by mapping biases. To this end, we extracted data from genomic intervals hosting avian  
205 Benchmarking Universal Single-Copy Orthologs (BUSCOs). This resulted in data from 7,335  
206 BUSCOs, with alignment lengths varying from 89,898 kb to 140,640 kb (depending on filtering  
207 strategy) for ML analyses of concatenated data, respectively 2,091 BUSCOs with alignment lengths  
208 varying from 10,575 kb to 15,290 kb for LD-pruned data free of interlocus recombination for  
209 multispecies coalescent (MSC)-based species tree reconstruction. Results were consistent  
210 between filtering strategies. Hence, we only report results based on the most stringent filtering of  
211 read depth (ii, DP=5, PW=50%, MD=15%). Both, maximum likelihood (ML) analyses in IQtree2  
212 based on concatenated data and MSC analyses in ASTRAL-III (based on BUSCO ML gene trees)  
213 established sub-Saharan species of the genera *Campicoloides*, *Emarginata*, *Myrmecocichla*,  
214 *Pinarochroa*, and *Thamnolaea* as the sister clade to all other open-habitat chats (**Fig. S1a**). For the  
215 subsequent analyses we excluded the *Saxicola* outgroup and rooted the trees on the sub-Saharan  
216 clade.  
217

218 We then moved to reconstruct the species tree based on an as broad representation to the  
219 genome as possible. To this end, we extracted alignments including variant and invariant sites for  
220 non-overlapping 10 kb windows. We henceforth refer to these windowed data as “loci”. Analyses  
221 included only loci that fulfilled filtering criteria for read depth, alignment length, data missingness  
222 (see Material and Methods), and absence of evidence for intra-locus recombination. Furthermore,  
223 we sub-sampled filtered loci to be at least 10 kb apart to ensure free inter-locus recombination.  
224 Depending on filtering strategy, this left us with 5,267-6,791 loci with total alignment lengths of  
225 34,556-52,243 kb (**Tab. S1**). We identified branches in the “anomaly zone” (Degnan and  
226 Rosenberg 2006) in several clades of wheatears: in the *hispanica* and *picata* complexes, and in the  
227 *isabellina* clade (**Fig. 1**). Nevertheless, the polytomy test based on local quartet supports in  
228 ASTRAL-III showed no evidence for polytomies in the species tree (P = 0 for all branches). The ML  
229 tree based on concatenated data and the MSC-based species tree were fully supported and in  
230 agreement both with each other (except the position of *T. cinnamomeiventris* within the sub-  
231 Saharan clade) and with the tree based on BUSCOs (**Fig. S1b**). Finally, a SNP-based species tree  
232 estimated in SVDquartets mostly confirmed the sequence-based results (**Fig. S2**). The only three  
233 disagreements (position of *O. leucura* and *O. leucopygia*, position of *O. bottae* and *O. pileata*, and  
234 position of *T. cinnamomeiventris*) were poorly supported in the SNP-based analysis and are likely  
235 a result of high levels of ILS under which sequence-based approaches are more accurate than  
approaches based on SNP data alone (Chou et al. 2015).

1  
2  
3 236 **Mitogenomic Relationships and Mito-Nuclear Discordances**  
4  
5

6 237 We were interested in whether previously inferred relationships based predominantly on single  
7 238 mitochondrial genes (Alaei Kakhki et al. 2016; Aliabadian et al. 2012; Schweizer and Shirihi  
8 239 2013) were supported by full mitogenomes and in inferring mito-nuclear discordances.  
9  
10

11 Mitogenomic relationships were in remarkable agreement with previously inferred  
12 phylogenetic relationships based predominantly on individual mitochondrial genes (Alaei Kakhki  
13 et al. 2016; Aliabadian et al. 2012; Schweizer and Shirihi 2013), yet showed several discordances  
14 with the species tree recovered from nuclear data (Fig. 2). Mito-nuclear discordances in  
15 wheatears were found in several places across the species tree but were mostly restricted to the  
16 placements of tip taxa: (i) In the *lugens* complex, nuclear data placed *O. l. persica* within the  
17 placements of tip taxa: (i) In the *lugens* complex, nuclear data placed *O. l. persica* within the  
18



59 **Figure 2 | Mito-nuclear discordances.** Shown are the time-calibrated phylogenetic trees based on  
60 nuclear data (left) and full mitogenomes (right).

1  
2  
3 246 complex, whereas the mitogenome placed it with *O. xanthoprymna* and *O. chrysopygia*. (ii) In the  
4 247 *picata* complex, *O. albonigra* that by mitochondrial data was considered a sister taxon to the *picata*  
5 248 complex, was placed within the latter as a sister taxon to the phenotypically almost identical *O. p.*  
6 249 *picata* by nuclear data. (iii) In the *hispanica* complex, *O. cypriaca* was placed as sister to either *O.*  
7 250 *melanoleuca* or *O. pleschanka* in nuclear and mitogenomic data respectively. (iv) In the *isabellina*  
8 251 clade, *O. heuglini* as sister to either *O. isabellina* or *O. bottae* by nuclear or mitogenomic data,  
9 252 respectively. (v) Moreover, *O. leucura* and *O. leucopyga* formed a sister clade to the *O.*  
10 253 *lugubris/lugentoides* clade according to the nuclear species tree, but mitogenomes placed them  
11 254 consecutively at the root of the clade including *O. finschi* and the *lugens* complex.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

255 To understand whether nuclear gene trees were entirely discordant with mitogenomic  
256 relationships or in part reflected the latter, for each of the above discordances we checked for  
257 nuclear gene trees that agreed with the mitogenomic tree. This showed that for most of the  
258 mitonuclear discordances, roughly 15% of the gene trees agreed with the mitogenomic  
259 relationships (*picata* complex: 14.40%, 4,282 of 29,730 gene trees; *hispanica* complex: 13.13%,  
260 3,905 of 29,730 gene trees; *isabellina* clade: 15.71%, 4,671 of 29,730 gene trees; *lugens* complex:  
261 2.77%, 824 of 29,730 gene trees).

## 262 Time trees

263 In addition to the species' relationships, we were interested in understanding the time scales at  
264 which species diverged. Due to the lack of appropriate fossils, we resorted to first estimate a time-  
265 calibrated mitochondrial phylogeny based on the 13 mitochondrial protein-coding genes, for  
266 which substitution rates are available (Lerner et al. 2011). The analysis in BEAST 2.6.6 showed  
267 high convergence of all parameters in the three independent runs after 25% of the trees were  
268 discarded as burn-in (ESS >300). The results were in good agreement with previous results  
269 obtained from single genes (Alaei Kakhki et al. 2016), dating the origin of open-habitat chats to  
270 the Miocene about 5.67 million years ago (mya) (95% highest posterior density (HPD): 5.32–6.06  
271 mya). The diversification of wheatears (genus *Oenanthe*) started about 5.09 mya (95% HPD: 4.75–  
272 5.44 mya) (**Fig. 1**, **Fig. S3**).

273 We then used the diversification time of the open-habitat chats estimated from  
274 mitochondrial data as a time constraint in dating analyses based on nuclear data. For these  
275 analyses, we first provided the topology and branch lengths obtained from ML analyses of  
276 concatenated BUSCO data along with 1.8 Mb high-confidence nuclear data (see Material and  
277 Methods) to generate the time-calibrated tree with RelTime-ML (Kumar et al. 2018). Compared  
278 to the mitochondrial results, the nuclear data mostly estimated similar divergence times between  
279 clades and shorter divergence times within clades (Pearson's r=0.93, p<0.001) (**Fig. 1**, **Fig. S3**).  
280 Second, we performed dating analyses for windowed loci across the genome the same way as for  
281 BUSCO by providing 3.8 Mb high-confidence data. Divergence times based on BUSCO strongly  
12

1  
2  
3 282 correlated with ones estimated from windowed loci (Pearson's  $r=0.99$ ,  $p<0.001$ ) (**Fig. S4**). A test  
4 in which we re-ran the estimation of mitochondrial divergence times in RelTime-ML the same way  
5 as for nuclear data yielded the same divergence times as estimated in BEAST, thus confirming that  
6 differences in divergence times between mitochondrial and nuclear data are not due to the  
7 approach but reflect the different data types.  
8  
9  
10  
11

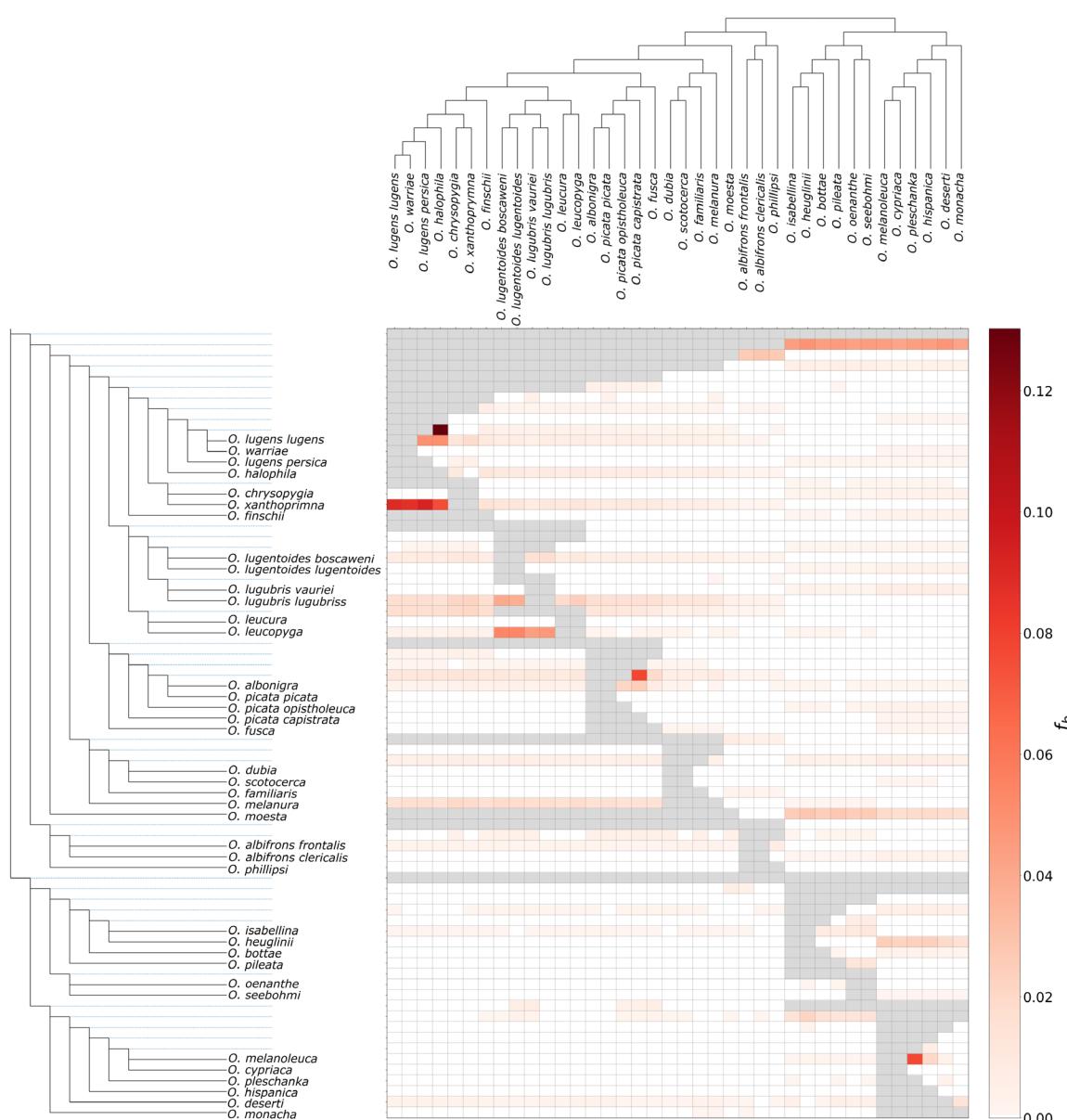
## 12 287 **Extensive Gene Tree Heterogeneity** 13

14 288 Having established the species tree, we aimed to quantify the levels of gene tree heterogeneity in  
15 wheatears to understand whether the processes generating gene tree heterogeneity could  
16 potentially underly phenotypic parallelism in this core group of open-habitat chats that displays  
17 290 the highest incidence of phenotypic parallelism.  
18 291  
19

20 292 Several lines of evidence demonstrate extensive gene tree heterogeneity in wheatears.  
21 Remarkably, not a single gene tree out of 29,730 gene trees matched the species tree.  
22 Furthermore, many branches of the species tree – including ones with local posterior probability  
23 294 1 – showed a high number of conflicting compared to concordant bipartitions, as evidenced by  
24 low Internode Certainty All (ICA) scores (**Fig. S5**), with ICA ranging from 1 to 0.35 and average  
25 296 ICA of  $0.65 \pm 0.19$  (mean  $\pm$  standard deviation). The high gene tree heterogeneities highlighted by  
26 297 ICA were further supported by low percentages of gene trees recovering the topology of the  
27 298 species tree at these internodes, as estimated by the gene concordance factor (gCF) (**Fig. 1**) that  
28 299 ranged from 1 to 0.06 with an average of  $0.52 \pm 0.30$  (mean  $\pm$  standard deviation). ICA and gCF  
29 300 were highly correlated (Pearson's  $r=0.94$ ,  $p<0.001$ ) (**Fig. S5**). As expected, evidence for extensive  
30 301 gene tree heterogeneity was highest in clades with branches classified as within the phylogenetic  
31 302 anomaly zone. These included the *lugens*, *picata*, and *hispanica* complexes, the *isabellina* clade,  
32 303 and the placement of *O. leucopyga* and *O. leucura*.  
33  
34  
35  
36  
37  
38  
39  
40  
41

## 42 305 **Contributions of ILS to Gene Tree Heterogeneity** 43

44 306 Next, we aimed to understand to which extent the levels of gene tree heterogeneity observed in  
45 wheatears can be explained by ILS alone. To this end, we first tested whether the MSC without  
46 hybridization adequately explains the gene tree heterogeneity observed across the entire species  
47 tree. The Tree Incongruence Checking in R (TICR) test (Stenz et al. 2015) showed an excess of  
48 309 outlier quartets ( $p < 0.01$ ), indicating that a model including ILS but not introgression does not  
49 310 adequately explain the observed gene tree heterogeneity. This suggests that introgression  
50 311 occurred during the evolutionary history of wheatears.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



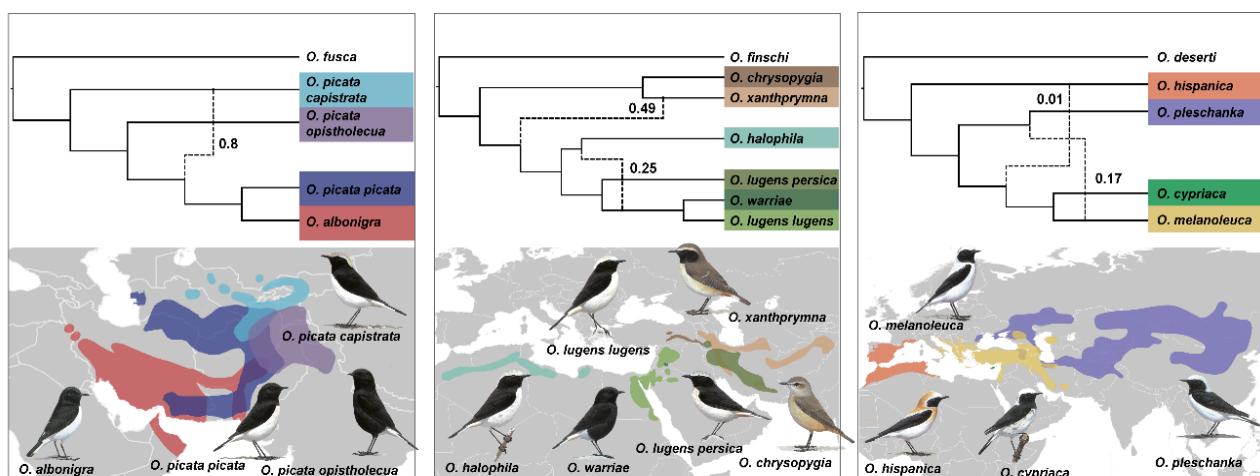
**Figure 3 | Footprints of introgression as estimated by the  $f$ -branch statistic.** The heat map summarizes the  $f$ -branch statistics estimated in Dsuite. Darker colors depict increasing evidence for gene flow between lineages. Dotted lines in the phylogeny represent ancestral lineage.

Therefore, we moved on to infer for each branch in the species tree separately whether ILS alone may explain the level of gene tree heterogeneity. To this end, for each internal branch, we estimated the number of gene trees supporting the first and second alternative topologies, based on the rationale that under ILS the first and second alternative gene tree topologies should be supported by an equal number of gene trees (Sayyari and Mirarab 2018). We identified 11 out of 37 internal branches (30%) for which the number of gene trees supporting the two alternative topologies were not significantly different (colored branches in Fig. 1). At these 11 internal branches, ILS alone can thus explain gene tree heterogeneity, while asymmetries at the other 26 internal branches may need to invoke other processes.

1  
2  
3 322 **Contributions of Introgression to Gene Tree Heterogeneity**  
4  
5

6 323 Given that gene tree heterogeneity at many branches could not be explained by ILS alone, we set  
7 324 out to infer footprints of introgression across wheatears. To this end, we first applied the approach  
8 325 based on D-statistics (Durand et al. 2011) implemented in Dsuite using > 58 million biallelic SNPs.  
9 326 This approach estimates D and f4 statistics across all possible combinations of trios in wheatears  
10 327 and then performs an f-branch test to assign gene flow to specific internal branches. The f-branch  
11 328 test suggested multiple events of introgression (**Fig. 3**):(i) *O. halophila* and the ancestor of *O.*  
12 329 *lugens lugens* and *O. wariae*, (ii) *O. xanthopygma* and the ancestor of the *lugens* complex, (iii) *O.*  
13 330 *leucopyga* and the ancestor of the *O. lugubris/lugentoides* clade, (iv) *O. picata capistrata* and the  
14 331 ancestor of *O. picata picata* and *O. albonigra*, and (v) *O. melanoleuca* and *O. pleschanka*.

15 332 Finally, we corroborated the evidence for introgression in the *hispanica*, *lugens*, and *picata*  
16 333 complexes with MSC network analyses in phyloNet, allowing for 0-5 introgression events.  
17 334 According to the Bayesian Information Criterion (BIC), models involving reticulation events better  
18 335 fit the data than strictly bifurcating trees in all three complexes (**Tab. S2**). In the *lugens* complex,  
19 336 two introgression events were detected: between *O. xanthopygma* and the ancestor of *O. lugens*  
20 337 ( $\gamma=49\%$ ), and between *O. halophila* and the *O. lugens lugens-O. wariae* ancestor ( $\gamma=25\%$ ) (**Fig. 4**).  
21 338 One introgression event was detected in the *picata* complex, between *O. picata capistrata* and the  
22 339 ancestor of *O. picata picata* and *O. albonigra* ( $\gamma=8\%$ ) (**Fig. 4**). In the *hispanica* complex, the highest  
23 340 scoring network involved two introgression edges: one between *O. melanoleuca* and *O. pleschanka*  
24 341 ( $\gamma=17\%$ ), and one between *O. hispanica* and the *O. cypriaca-melanoleuca* ancestor ( $\gamma=1\%$ ) (**Fig. 4**).  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36



57 **Figure 4 | Phylogenomic networks and distribution ranges for the *picata* (left), *lugens* (middle) and**  
58 ***hispanica* (right) complexes.** Phylogenomic networks were estimated under the maximum pseudolikelihood  
59 approach implemented in phyloNet. Numbers on the edges indicate the inheritance probabilities, which  
60 correspond to the proportion of gene trees supporting the reticulate relationship. Drawings courtesy of Chris  
61 Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc. Distribution ranges  
modified from BirdLife International and the Handbook of the Birds of the World (2016).

## 342 Discussion

343 The present study provides the first genomic insights into the speciation history of open-habitat  
344 chats and into the processes involved in shaping gene tree heterogeneity that may also underpin  
345 the high incidence of phenotypic parallelism in this group of songbirds. Our analyses reveal  
346 unambiguous species relationships despite considerable gene tree heterogeneity, including  
347 several mito-nuclear discordances that result from a combination of ILS and introgression. These  
348 relationships reconstructed from genomic data provides the strongest evidence yet for a high  
349 incidence of phenotypic parallelism in open-habitat chats, as exemplified for three phenotypes in  
350 **Fig. 1.**

351 We first discuss how mito-nuclear discordances and incidences of introgression together  
352 with known histories of hybridization and biogeographic history mold into a comprehensive  
353 picture of open-habitat chat evolution. We close by concluding based on the indirect evidence  
354 presented here that phenotypic parallelism in open habitat chats likely evolved through a  
355 combination of ILS, introgression and novel mutations in independent lineages. Together, our  
356 results paint a picture of genomic and phenotypic evolution that is in part marked by the sharing  
357 of ancestral variation and an exchange of genetic variation between species. Our study therefore  
358 contributes to the increasing body of evidence that phenotypic and species evolution not only  
359 proceed from novel mutations but abundantly reuse genetic variation present in ancestral and  
360 related species (Marques et al. 2019b; Meier et al. 2018; Seehausen et al. 2014).

### 361 **Mito-nuclear discordances, patterns of introgression, hybridization history, and** 362 **biography mold into a coherent picture of complex open-habitat chat evolution**

363 The species relationships inferred from nuclear genomic data were in good agreement with  
364 previous phylogenies based predominantly on single mitochondrial markers (Aliabadian et al.  
365 2012; Schweizer and Shrihai 2013) and thereby confirmed the biogeographic history of open-  
366 habitat chats (Alaei Kakhki et al. 2016). Still, we recovered several species relationships  
367 discordant between the nuclear genome and the mitogenome (Toews and Brelsford 2012). In the  
368 light of (i) the histories of introgression also uncovered here, (ii) the known hybridization history,  
369 and (iii) the here confirmed biogeography, most of these mito-nuclear discordances can be well  
370 embedded in a coherent history of open-habitat evolution.

371 The close nuclear relationship of *O. albonigra* with the nominate subspecies *O. p. picata* is in  
372 stark contrast with the mitochondrial divergence of *O. albonigra* with all *O. picata* subspecies  
373 about 0.5 mya (**Fig. 4a**). However, as an exception for wheatears, even from a perspective of  
374 plumage coloration the nuclear species tree implies a more parsimonious history of phenotypic  
375 evolution, as *O. albonigra* and *O. p. picata* display almost identical plumages. The high  
376 mitochondrial similarity of all subspecies currently treated under *O. picata* according to Panov

377 (2005) may be a result of introgressive hybridization. Indeed, the high abundance of admixed  
378 phenotypes in zones of contact between the members of this species complex (Panov 2005)  
379 suggests a high incidence of hybridization. Different from the *hispanica* complex, where taxa meet  
380 in restricted zones, lineages of the *picata* complex all mold together in a relatively large area in  
381 southern Central Asia, and their degree of reproductive isolation is largely unknown. Further  
382 population genomic insights are required from the *picata* complex to obtain detailed insights into  
383 its history of hybridization and phenotypic evolution.

384 The evolution of the *lugens* complex was marked by two incidences of introgression that  
385 likely underpin the mito-nuclear discordance observed in this complex (**Fig. 4b**). Introgression  
386 occurred between *O. xanthopyrymna* and the *O. lugens* ancestor and between north-African *O.*  
387 *halophila* and the middle eastern *O. l. lugens-O. warriae* ancestor. Both incidences of introgression  
388 make sense in the light of biogeography, as they occurred between geographically neighboring  
389 taxa (**Fig. 4b**). Together they can explain the close mitochondrial relationship of *O. l. persica* with  
390 *O. xanthopyrymna* and *O. chrysopygia*: *O. xanthopyrymna* mitochondria were introduced into the *O.*  
391 *lugens* ancestor by introgression and may at first have segregated in the *O. lugens* lineage but then  
392 have been lost in *O. halophila*. Mitochondrial replacement with *O. halophila* variation upon genetic  
393 exchange of the latter taxon with the *O. l. lugens-O. warriae* ancestor would have left *O. l. persica*  
394 the only taxon with a *O. xanthopyrymna*-like mitogenome. Importantly, our results shed first  
395 genomic light on the divergence of Basalt Wheatear (*O. warriae*), a species with a very restricted  
396 range that is interesting from the perspective of phenotypic evolution: this species turns out to be  
397 highly similar to *O. l. lugens* at the genomic level, which contrasts with its marked phenotypic  
398 divergence (**Fig. 4b**). This result is similar to the situation observed, for instance, in Hooded and  
399 Carrion Crows (*Corvus cornix* and *C. corone*, respectively) (Poelstra et al. 2014) and opens  
400 interesting questions on the evolutionary history of this taxon's coloration. Finally, in the  
401 *hispanica* complex, the incomplete sorting of mitochondrial variation was previously well  
402 documented (Alaei Kakhki et al. 2018; Randler et al. 2012), and footprints of introgression came  
403 as no surprise: The complex is characterized by pervasive hybridization of *O. melanoleuca* with *O.*  
404 *pleschanka* in several geographic regions (Haffer 1977; Panov 1992) and population genomic  
405 analyses suggest rates of introgression of up to almost 20% between these species (Schweizer et  
406 al. 2019a). Research is underway to uncover the detailed histories of hybridization in this  
407 Eurasian wheatear complex.

408 The thus far discussed mito-nuclear discordances were all accompanied with high levels of  
409 gene tree heterogeneity (most within the phylogenetic anomaly zone). However, most of these  
410 cases were not explained by ILS alone but went along with footprints of introgression. Still, part  
411 of the observed mito-nuclear discordances might well be a consequence of ILS. In the *picata*  
412 complex, for instance, lineage divergence occurred in rapid succession (**Fig. 1**), and ILS might well

1  
2  
3 413 be an alternative explanation for the mitochondrial divergence of the *O. albonigra* mitogenome. In  
4 414 addition, in the clade including *O. heuglinii* and the very widespread *O. isabellina*, species split in  
5 415 fast succession and the high levels of ILS likely explain the observed mito-nuclear discordance.  
6  
7

8 416 Taken together, our results demonstrate that the speciation history of open-habitat chats is  
9 417 similarly complex as their phenotypic evolution. Multiple events of introgression at both extant  
10 418 and ancestral time scales, along with abundant ILS, contributed to reticulate evolution and thus a  
11 419 mosaic of genomic variation in several clades of wheatears. Our study thus adds to an increasing  
12 420 number of examples (Enciso-Romero et al. 2017; Han et al. 2017; Lamichhaney et al. 2018; Meier  
13 421 et al. 2017) highlighting that species diversification is often complex and rather than by a linear  
14 422 process is at least in part a network of interacting lineages (Marques et al. 2019b)  
15  
16

17 423 **Diverse routes to phenotypic parallelism in open-habitat chats**

18 424 The reconstruction of relationships among open-habitat chats using genomic data has a deep  
19 425 impact on our understanding of phenotypic evolution in these songbirds: the species tree provides  
20 426 firm evidence for an extraordinary incidence of phenotypic parallelism (**Fig. 1**). For numerous  
21 427 traits, including plumage coloration, sexual dimorphism, and migration behavior, not related  
22 428 species display more similar phenotypes than sister species (**Fig. 1**). Almost entirely black  
23 429 plumages, for instance, evolved in five clades (*O. picata opistholeuca*, *O. wariae*, *O. leucura*, female  
24 430 *M. monticola*, and juvenile *O. leucopyga*), and sexually monomorphic female-type plumage is found  
25 431 in another five clades (*O. chrysopygia*, *O. fusca*, the *O. melanura* clade, the *O. isabellina* clade, and  
26 432 in the sub-Saharan clade), to name just two out of many examples.  
27  
28

29 433 Furthermore, our results suggest (directly for introgression and ILS, indirectly for parallel  
30 434 evolution *sensu stricto*) that phenotypic parallelism in open-habitat chats is unlikely explained by  
31 435 a single process but may need to invoke all possible routes (Hedrick 2013; Konečná et al. 2021;  
32 436 Montejo-Kovacevich et al. 2021; Natarajan et al. 2015; Pease et al. 2016), with the most likely  
33 437 processes depending on both demography and the phylogenetic scale.  
34  
35

36 438 For ILS to contribute to phenotypic parallelism, species must diverge in fast succession and  
37 439 maintain critically high effective population sizes to pass on ancestral variation and maintained it  
38 440 in daughter lineages. In open-habitat chats, such fast radiations occurred predominantly at rather  
39 441 recent timescales. The shortest split intervals are observed (in increasing order) in the *picata*,  
40 442 *hispanica*, and *lugens* complexes (**Fig. 1**). However, phenotypic parallelism of species in the *lugens*  
41 443 complex and of the *picata* complex is only found with other clades but not within the complexes.  
42 444 Given that the levels of ILS at the root of the *lugens* complex are restricted, ILS is unlikely to have  
43 445 contributed to phenotypic parallelism with other clades of wheatears sporting, for instance,  
44 446 similar plumages (see for instance the aforementioned example including *O. wariae*). Phenotypic  
45 447 parallelism is, however, observed for back and neck-side coloration in the *hispanica* complex  
46 448 (Schweizer et al. 2019a), and could be explained by ILS.  
47  
48

Likewise, introgression would need to involve taxa with similar phenotype to explain phenotypic parallelism. Our analyses indeed uncovered several instances of in part substantial introgression (**Fig. 3**, **Fig. 4**). However, despite suggesting that introgression provided the opportunity to exchange phenotypes between species, none of the inferred introgression events can be tied to concrete examples of phenotypic parallelism. This raises the question, whether the inferences of introgression applied here are underpowered to infer footprints of introgression relevant to phenotypic evolution of open-habitat chats, or whether, indeed, introgression played a limited role in these songbirds' phenotypic evolution.

Finally, many if not most phenotypic similarities in open-habitat chats are found across major phylogenetic clades that diverged around 5 mya (for instance the examples provided at the entry of the discussion). The split events at these time scales did not occur within short evolutionary time scales and, accordingly, levels of ILS are rather low for at least one of the relevant nodes (**Fig. 1**). Although gene tree heterogeneity was non-negligible for the larger of the two major wheatear clades, gene trees were mostly concordant for the root nodes of the wheatear clade including the *hispanica* complex and the *O. oenanthe* and *O. isabellina* clades (**Fig. 1**). Moreover, the phenotypically similar species occur in geographically well separated ranges and hybridization is thus rather unexpected. In conclusions, unless the approaches used here to detect the ILS and introgression are underpowered, the indirect evidence provided by our results suggests that many incidences of phenotypic parallelism at such timescale may have involved independent mutations to underpin the same phenotype, that is, parallel evolution *sensu stricto*.

## Conclusion

In the present study we set out to probe gene tree variation for footprints of ILS and introgression with the goal of understanding how ILS and introgression may have contributed to phenotypic parallelism in open-habitat chats. Our results reveal a complex speciation history and provide conclusive evidence for abundant phenotypic parallelism in open-habitat chats. While we cannot conclude on the involvement of specific processes in the evolution of specific phenotypic parallelisms, the indirect evidence gained from the structure of the species tree and inferred levels of ILS and introgression suggest that the evolution of phenotypic parallelism in open-habitat chats likely occurred along all diverse trajectories, namely ILS, introgression, and parallel evolution through novel mutations. Thereby, our results contribute to a growing body of evidence that evolution makes use and re-use of all resources it has at hand including both standing (ancestral or heterospecific) as well as novel genetic variation.

481

## 482 Material and Methods

### 483 Taxon sampling, DNA extraction, and whole-genome resequencing

484 Aiming for complete taxon sampling, we sequenced the genomes of 50 open-habitat chat taxa from  
485 a total of 44 species from the genera *Oenanthe*, *Campicoloides*, *Emarginata*, *Myrmecocichla*,  
486 *Pinarochroa*, and *Thamnolaea* (**Fig. 2; Tab. 1**). This sampling included all but three species (*E.*  
487 *tractrac*, *M. collaris*, *T. coronata*) of the 47 currently recognized open-habitat chat species (Gill et  
488 al. 2020). A genome sequence of *Saxicola maurus* (European Nucleotide Archive accession  
489 number: ERR2560200-ERR2560209), a species of open-habitat chats' sister lineage (Sangster et  
490 al. 2010; Zuccon and Ericson 2010), was included as an outgroup to root the open-habitat chat  
491 species tree. We followed the taxonomy of the IOC World Bird List (v12.1) (Gill et al. 2020) except  
492 for the *picata* complex, where we treat subspecies *picata*, *capistrata* and *opistholeuca* separately,  
493 following Panov (2005).

494 We extracted DNA from blood stored in ≥96% ethanol or Queen's Lysis buffer or tissues  
495 stored in 96% ethanol for taxa for which fresh material was available, or from toepads or dried  
496 skin from skin-preparation sutures for taxa for which only museum samples were available (**Tab.**  
497 **1**). From blood and tissue samples DNA was extracted using the DNeasy Blood and Tissue Kit  
498 (Qiagen) or the MagAttract HMW DNA kit (Qiagen) following the manufacturer's protocol with  
499 exception of an adapted digestion of blood samples as reported in Lutgen et al (Lutgen et al. 2020).  
500 DNA from toepads and dried skin was extracted using the QIAamp DNA Micro Kit (Qiagen) with  
501 an adapted digestion protocol that ensures high quantities of DNA  
502 ([dx.doi.org/10.17504/protocols.io.dm6gpwrplzp/v1](https://dx.doi.org/10.17504/protocols.io.dm6gpwrplzp/v1)). DNA concentrations were quantified on a  
503 Qubit fluorometer (dsDNA BR assay, Thermo Fisher Scientific) and DNA integrity was evaluated  
504 on a TapeStation (MANUFACTURER, KIT). We prepared sequencing libraries using the ThruPLEX  
505 DNA-Seq Kit (Takara), the Illumina DNA Prep Kit, the Illumina DNA PCR-free Kit, or the Chromium  
506 Genome Library kit (10X Genomics) for intact DNA, or for fragmented DNA with the ACCEL-NGS  
507 1S DNA Library Prep Kit (Swift Biosciences) (**Tab. 1**). All libraries were sequenced (150 bp paired-  
508 end) on Illumina NovaSeq6000 instruments with a target coverage of ca. 15x.

### 509 Data preparation

#### 510 Adapter trimming and mapping of resequencing data

511 Prior to further analysis, for all but the linked-read sequencing data, we trimmed adapters and  
512 merged overlapping paired-end reads using fastp 0.20.0 (Chen et al. 2018). For linked-read  
513 sequences, we trimmed the first 22 bp on the R1 read to eliminate the 10X indexes. We then  
514 mapped the reads to the reference genome assembly of *Oenanthe melanoleuca* (Peona et al. in  
515 prep.) using BWA 0.7.17 (Li 2013) and marked duplicates with PicardTools 2.9.1  
516 (<http://broadinstitute.github.io/picard>). After excluding duplicates, the average sequencing

1  
2  
3 517 coverage per individual ranged from 4.6x to 40.6x (mean and median 12.2, standard deviation  
4 518 6.20) (**Tab. 1**).  
5  
6

7 519 *Base quality score recalibration (BQSR), SNP calling, and SNP genotyping*  
8 520 Data preparation followed the GATK 4.1.4.1 (McKenna et al. 2010) best practices pipeline. First,  
9 521 to prepare a list of high-confidence SNPs for BQSR, we ran HaplotypeCaller to generate gvcf files  
10 522 for each sample and then merged gvcf files of all samples with CombineGVCFs before genotyping  
11 523 SNPs using GenotypeGVCFs. To retain only high-confidence SNPs in the SNP-exclude set for BQSR,  
12 524 we retained only SNPs that fulfilled the following criteria: mapping quality > 40, Fisher strand (FS)  
13 525 phred-scaled p-value < 60, SNP quality score > 20, mapping quality rank sum value > -12.5, read  
14 526 pos rank-sum test value > -8.0 and quality by depth > 2. We retained only biallelic SNPs with at  
15 527 least one homozygous reference and one homozygous alternative genotype or with at least three  
16 528 observations of reference and alternative alleles. We excluded the resulting set of SNPs from BQSR  
17 529 in GATK. Following BQSR, we ran HaplotypeCaller on base-score-recalibrated bam files. The  
18 530 resulting gvcf files of all samples where merged (CombineGVCFs) and variant and invariant sites  
19 531 genotyped using the ‘include-non-variant-sites’ flag in GenotypeGVCFs. For all subsequent  
20 532 analyses we based genotypes on genotype likelihoods. This resulted in 871,428,254 unfiltered  
21 533 sites when the outgroup was included and 872,152,150 unfiltered sites without the outgroup.  
22  
23

24 534 In phylogenomic data sets, which are based on mapping of resequencing data to a  
25 535 reference genome, data of species more divergent from the reference genome may risk mapping  
26 536 at a lower percentage. To check for such mapping-related biases in our dataset, we estimated the  
27 537 average number of nucleotide differences ( $d_{xy}$ ) between *Oenanthe melanoleuca* (reference  
28 538 genome) and all other species using pixy 0.95.02 (Korunes and Samuk 2021). We then estimated  
29 539 the mapping percentage for all species using SAMtools (Li et al. 2009) and tested whether there  
30 540 was a correlation between  $d_{xy}$  and mapping success.  
31  
32

33 541 *Data filtering*  
34 542 Before data analysis, we removed all repeat regions from the multi-sample VCF file using the  
35 543 repeat mask reported in Peona et al. (in prep.). Then we used BCFtools 1.11 (Li 2011) to remove  
36 544 indels, sites close to indels (up to 10 bp) and all the sites at which exclusively alternative alleles  
37 545 were called. For analyses requiring variant sites only, we removed all SNPs with more than 20%  
38 546 missing data and all invariant sites using BCFtools and retained only SNPs with a minimum read  
39 547 depth of five. To ensure linkage-disequilibrium (LD) among SNPs, we LD-pruned SNPs in VCFtools  
40 548 0.1.16 (Danecek et al. 2011) such as to only retain SNPs with a minimum distance of 1 kb between  
41 549 them. This physical distance is expected to remove most LD between SNPs, as e.g. in flycatchers  
42 550 LD breaks down in most genomic regions after 1 kb (Ellegren et al. 2012). After this filtering, we  
43 551 genotyped based on genotype likelihoods and retained 994,150 multiallelic SNPs. In addition, for  
44  
45

1  
2  
3 552 analyses that require biallelic SNPs exclusively, we removed all multiallelic SNPs from the VCF file  
4 553 after the above filtering, using BCFtools.  
5

6 554 For phylogenomic analyses requiring sequence data including both variant and invariant  
7 sites, we defined 10 kb non-overlapping windows across the genome. Henceforth, we refer to the  
8 windowed data as “loci” and to phylogenetic trees inferred therefrom as “gene trees”. To make  
9 sure that the adopted filtering strategy did not affect our results, we generated four sets of fasta  
10 alignments using different filter settings for minimum read depth (DP), minimum percentage of  
11 the window covered by data (PW), and missing data per site (MD) for both the 10 kb loci and the  
12 BUSCO data set: (i) DP=1, PW=50%, MD=15%, (ii) DP=5, PW=50%, MD=15%, (iii) DP=1,  
13 PW=50%, MD=5%, and (iv) DP=1, PW=80%, MD=10%. These four filtering strategies yielded the  
14 same species tree and concatenated tree for 10 kb loci as well as for BUSCOs. For these analyses,  
15 we therefore exclusively report the results based on the most stringent filtering on read depth (ii,  
16 DP=5, PW=50%, MD=15%). For gene tree heterogeneity analyses, on the other hand, we aimed to  
17 include the broadest representation of the genome and to this end retained all loci (N=29,730)  
18 that fulfilled less stringent filtering criteria (i, DP=1, PW=50%, MD=15%).  
19 562  
20 563  
21 564  
22 565  
23 566  
24 567  
25 568  
26 569  
27 570  
28 571  
29 572  
30 573  
31 574  
32 575  
33 576  
34 577  
35 578  
36 579  
37 580  
38 581  
39 582  
40 583  
41 584  
42 585  
43 586  
44 587

For phylogenomic analyses requiring sequence data including both variant and invariant sites, we defined 10 kb non-overlapping windows across the genome. Henceforth, we refer to the windowed data as “loci” and to phylogenetic trees inferred therefrom as “gene trees”. To make sure that the adopted filtering strategy did not affect our results, we generated four sets of fasta alignments using different filter settings for minimum read depth (DP), minimum percentage of the window covered by data (PW), and missing data per site (MD) for both the 10 kb loci and the BUSCO data set: (i) DP=1, PW=50%, MD=15%, (ii) DP=5, PW=50%, MD=15%, (iii) DP=1, PW=50%, MD=5%, and (iv) DP=1, PW=80%, MD=10%. These four filtering strategies yielded the same species tree and concatenated tree for 10 kb loci as well as for BUSCOs. For these analyses, we therefore exclusively report the results based on the most stringent filtering on read depth (ii, DP=5, PW=50%, MD=15%). For gene tree heterogeneity analyses, on the other hand, we aimed to include the broadest representation of the genome and to this end retained all loci (N=29,730) that fulfilled less stringent filtering criteria (i, DP=1, PW=50%, MD=15%).

Finally, for analyses making assumptions on intra- and inter-locus recombination (such as species tree reconstructions) we made sure to include only loci with no intra-locus but free inter-locus recombination. To this end, we excluded all loci with recombination signals ( $P \leq 0.05$ ) as inferred from the pairwise homoplasy index Phi ( $\Phi_w$ ) estimated in PhiPack 1.1 program (Bruen et al. 2006). The criterion  $P \leq 0.05$  does not account for multiple testing, but we preferred to conservatively exclude loci with evidence for intra-locus recombination. To possibly retain only loci among which free recombination occurs, we ensured a minimum distance of 10 kb by including no two consecutive loci. At this distance, no LD occurs in flycatchers (Ellegren et al. 2012).

#### 576 *Inference of Benchmarking Universal Single-Copy Ortholog (BUSCO) sequences*

Phylogenomic analyses based on the mapping of resequencing data to a reference genome, especially when including species well diverged from the latter, may be affected by several biases. For species more divergent from the reference genome, data from faster evolving genomic regions (i) risks not being mapped, if these regions are too diverged from the reference sequence, or (ii) may map to paralogs, if the species experienced different duplication histories (Chakrabarty et al. 2017; Fitz-Gibbon et al. 2017). These biases are expected to be least important in slowly evolving regions of the genome, especially in BUSCOs, that are conserved and by definition present in single copies in most species. To minimize mapping-related biases in our phylogenomic reconstructions, especially on rooting and placements of the most divergent species, we therefore extracted the intervals in which avian BUSCOs (aves\_odb10) are situated in our reference genome using BUSCO 5.0.0 (Simão et al. 2015).

**588 Phylogenomic reconstructions and multispecies coalescent analyses***589 BUSCO-based rooting of the open-habitat chat species tree*

590 First, to establish the root within open-habitat chats, we applied both concatenation and  
591 multispecies coalescent-based methods on BUSCO sequences, including the outgroup. First, we  
592 used all BUSCOs ( $N=7,335$ ) to estimate the maximum likelihood tree in IQ-TREE 2.1.2 (Minh et al.  
593 2020b) based on the concatenated BUSCOs, using one partition for each BUSCO and a  
594 GTR+I+G substitution model for all partitions (Abadi et al. 2019). One thousand bootstrap  
595 replicates were run using the ultrafast bootstrap approximation (Hoang et al. 2018). Second, we  
596 estimated the species tree under the multispecies coalescent using ASTRAL-III (Zhang et al. 2018)  
597 based on BUSCOs without recombination signals and free inter-locus recombination ( $N=2,091$ ).  
598 To this end, we inferred BUSCOs' gene trees in IQ-TREE 2.1.2 using a GTR+I+G substitution model  
599 and one thousand ultrafast bootstrap approximations. To ensure that species tree inferences were  
600 not affected by inaccurately estimated gene trees (Zhang et al. 2018), we collapsed branches with  
601 bootstrap support inferior to 80% using Newick Utilities 1.6 (Junier and Zdobnov 2010).  
602 Reconstructing the species tree by including all BUSCOs not considering intra- and inter-locus  
603 recombination ( $N=7,335$ ) did not affect the result.

*604 Phylogenomic and multispecies coalescent analyses based on full evidence*

605 To reconstruct the concatenated tree and species tree based on full evidence data, that is, data  
606 from the maximal possible fraction of the genome, and to study gene tree heterogeneity along the  
607 genome, we excluded the *Saxicola* outgroup. Instead, we rooted the trees with the clade that is the  
608 outgroup to all other open-habitat chats (sub-Saharan clade, **Fig. S1**). Excluding *Saxicola* ensured  
609 that analyses were not biased by mapping issues caused by this outgroup's divergence.

610 To estimate the concatenated tree using maximum likelihood in IQ-TREE 2.1.2 we used all  
611 loci with a GTR+I+G substitution model and 1,000 ultrafast bootstrap approximations. To  
612 estimate the species tree under the multispecies coalescent using ASTRAL-III, we at first  
613 estimated maximum likelihood gene trees using IQ-TREE 2.1.2 with a GTR+I+G substitution model  
614 and one thousand ultrafast bootstrap approximations. Based on these gene trees (pruned for  
615 within-locus recombination and assuring free recombination between loci), we inferred the  
616 species tree using ASTRAL-III. Because ASTRAL relies on accurately estimated gene trees, we  
617 collapsed branches with bootstrap support inferior to 80% using Newick Utilities 1.6.

618 To find regions of the species tree that represent "anomaly zones" where the frequency of  
619 one of the alternative quartets is higher than that of the topology in agreement the species tree,  
620 we estimated local quartet supports for the main topology and its two alternatives in ASTRAL-III  
621 (Degnan and Rosenberg 2006). We used the `anomaly_finder.py` script to search for anomaly zones  
622 in our species tree (Linkem et al. 2016). To test if the gene tree discordance could be explained by

1  
2  
3 623 polytomies instead of bifurcating nodes, we carried out a quartet-based polytomy test as  
4 implemented in ASTRAL-III.  
5

6 625 To see whether the SNP-based species tree could confirm the sequence-based species tree,  
7 we used the unlinked multiallelic SNPs to the multispecies coalescent model implemented in  
8 SVDQuartets (Chifman and Kubatko 2014) in PAUP\* 4 (Swofford 2003). We ran this with 1000  
9 bootstrap replicates and summarized the result in a 50% majority-rule consensus tree.  
10  
11

12 629 *Phylogenetic relationships of mitogenomes*  
13  
14

15 630 We were interested in whether previously inferred relationships based predominantly on single  
16 mitochondrial genes (Alaei Kakhki et al. 2016; Aliabadian et al. 2012; Schweizer and Shirihai  
17 2013) were supported by full mitogenomes and in how the mitogenomic relationships compare  
18 to the ones inferred from nuclear loci. To this end, we extracted and assembled mitochondrial  
19 genomes from the genomic data of all open-habitat chats using MitoFinder 1.2 (Allio et al. 2020).  
20  
21 634 We used the published Isabelline Wheatear (*Oenanthe isabellina*) mitochondrial genome as a  
22 reference (Genbank accession number: NC\_040290.1) and annotated the mitochondrial genome  
23 using the annotation pipeline integrated in MitoFinder. Finally, we aligned the 13 mitochondrial  
24 protein coding gene sequences using the automatic alignment strategy in MAFFT 7.471 (Katoh  
25 and Standley 2013). We checked the alignments in AliView 1.26 (Larsson 2014) and removed stop  
26 codons within the coding sequences or indels for downstream analyses. We determined the best  
27 partition scheme using the Akaike information criterion (AIC) implemented in PartitionFinder  
28 2.1.1 (Lanfear et al. 2017) and used the GTR+G+I model for all partitions. Then we constructed  
29 the maximum-likelihood tree from the concatenated supermatrix of all 13 genes in IQ-TREE 2.1.2  
30 using the ultrafast bootstrap approximations with 1,000 replicates.  
31  
32

33 641 **Dating analyses**  
34  
35

36 645 Beside species' relationships we were interested in estimating the divergence time in open-  
37 habitat chats. Because there are no appropriate fossils for calibration, we first ran BEAST 2.6.6  
38 (Bouckaert et al. 2019) for 13 mitochondrial protein coding genes to estimate a time-calibrated  
39 mitochondrial phylogeny. We included the mitochondrial genome sequence of *Saxicola maurus*  
40 (GenBank accession number: MN356403.1) as an outgroup in these analyses. Substitution models  
41 were inferred during the MCMS analyses with bModelTest (Bouckaert and Drummond 2017)  
42 implemented as a package in BEAST 2.6.6. Published substitution rates for each mitochondrial  
43 gene (Lerner et al. 2011) were implemented as means of the clock rates in real space of lognormal  
44 distribution with standard deviations of 0.005. We defined a Yule speciation process for the tree  
45 prior and an uncorrelated lognormal relaxed clock model. Three independent MCMC chains were  
46 run for 50 million generations, each with sampling every 5,000 generations. Effective sample sizes  
47 for all parameters and appropriate numbers of burn-in generations were checked with Tracer 1.5  
48  
49

1  
2  
3 658 (Rambaut and Drummond 2009). The three independent runs were combined using LogCombiner  
4 659 2.6.6 (Bouckaert et al. 2019). We used TreeAnnotator 2.6.6 (Bouckaert et al. 2019) to calculate a  
5 660 maximum clade credibility tree and the 95% highest posterior density (HPD) distributions of each  
6 661 estimated node.  
7  
8

9 662 We then used the divergence time of the sub-Saharan clade from wheatears estimated  
10 663 from mitochondrial data as time constraint in dating analyses based on nuclear data using  
11 664 RelTime-ML implemented in MEGA 11 (Tamura et al. 2021). For this analysis, we provided the  
12 665 topology with branch length estimated in IQtree2 based on concatenated BUSCOs data retained  
13 666 after the most stringent filtering (ii, DP=5, PW=50%, MD=15%), along with high-confidence  
14 667 BUSCO alignments. The latter consisted of BUSCO data filtered for DP=5, MD=5% and length of  
15 668 each BUSCO alignments longer than 1kb. We used the same filtering to get the 10 kb non-  
16 669 overlapping windows across the genome and used the concatenated tree retained after most  
17 670 stringent filtering (ii, DP=5, PW=50%, MD=15%) to repeat the analyses based on loci across the  
18 671 genome. To ensure that the differences in divergence times between mitochondrial and nuclear  
19 672 data were not due to the different dating approaches, we re-estimated the mitochondrial  
20 673 divergence times in RelTime-ML using the same approach as for the nuclear datasets.  
21  
22

### 23 674 **Inference of gene tree variation, ILS, and introgression**

#### 24 675 *Inference of the levels of gene tree variation*

25 676 To investigate gene tree heterogeneity across the genome, we used gene trees inferred from less  
26 677 stringent filtering criteria (i, DP=1, PW=50%, MD=15%) as described above. To infer how many  
27 678 gene trees reflect the species tree topology, we used the script 'findCommonTrees.py' (Edelman  
28 679 et al. 2019). To characterize the levels of gene tree heterogeneity across open-habitat chats, we  
29 680 compared the gene trees to the species tree. Specifically, we estimated "internode certainty all"  
30 681 (ICA) and the "gene concordance factor" (gCF). ICA quantifies the amount of gene tree  
31 682 heterogeneity for each internode of the species tree by calculating the number of all most  
32 683 prevalent conflicting bipartitions. It takes values ranging from -1 to 1, with values around zero  
33 684 indicating strong conflict; values towards 1 indicate robust concordance of gene trees with the  
34 685 species tree in the bipartition of interest; and negative values indicate discordance between the  
35 686 bipartition of interest and one or more bipartitions with a higher frequency (Salichos et al. 2014).  
36 687 While ICA thus represents the degree of conflict on each node of a species tree, gCF better reflects  
37 688 the gene tree heterogeneity around each branch, and is the percentage of gene trees supporting  
38 689 the two alternative topologies for each branch (Minh et al. 2020a). We estimated ICA and gCF with  
39 690 PhyParts 0.0.1 (Smith et al. 2015) and IQ-TREE 2.1.2 respectively.  
40  
41

42 691  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 692 *Tests of an ILS model*

4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
693 Next, we were interested in understanding whether ILS can sufficiently explain the level of gene  
694 tree heterogeneity observed at the level of the whole species tree. To this end, we applied the Tree  
695 Incongruence Checking in R (TICR) test (Stenz et al. 2015) implemented in the Phylolm R package.  
696 This test evaluates whether the multispecies coalescent adequately explains gene tree  
697 heterogeneity across the species tree with no hybridization edges. TICR requires posterior  
698 distributions of gene tree topologies inferred through Bayesian inference of gene trees. Therefore,  
699 we first estimated posterior distributions of individual gene trees with MrBayes 3.2.7 (Ronquist  
700 et al. 2012). MrBayes analyses ran using three independent runs of 20 million generations each,  
701 sampling every 20,000th generation using a GTR+I+G model. We estimated the length of burn-in  
702 using Tracer 1.5 (Rambaut and Drummond 2009) to ensure that our sampling of the posterior  
703 distribution had reached sufficient effective sample sizes (ESS > 200) for parameter estimation.  
704 We then ran BUCKy (Ané et al. 2007; Larget et al. 2010) using the posterior distribution of gene  
705 trees after discarding 25% as burn-in to estimate the concordance factors (CFs) for the three  
706 possible splits of all quartets. The inferred CF values were then tested against those expected  
707 under a coalescent model that takes ILS but not hybridization into account (chi-squared test).

708 We then tested for each branch in the species tree whether the gene tree heterogeneity  
709 reflected in gCF can be sufficiently explained by a model incorporating ILS alone. Under ILS alone  
– assuming sorting of variation occurs by random genetic drift – proportions of alternative gene  
710 trees for a rooted triplet are expected to be approximately equal (Hibbins and Hahn 2022; Sayyari  
711 and Mirarab 2018; Sayyari et al. 2018), and the concordant tree topology (the topology in  
712 agreement with the species tree) should be at least as frequent as the two discordant topologies  
713 (Hibbins and Hahn 2022; Sayyari et al. 2018). In contrast, introgression between non-sister taxa  
714 results in asymmetric proportions of gene trees in the rooted triplet (Durand et al. 2012; Green et  
715 al. 2010). Therefore, we performed a chi-square tests comparing the number of gene trees  
716 supporting the two discordant topologies. Under ILS, these two alternative topologies are  
717 expected to be equally frequent among gene trees (He et al. 2020). For all these analyses we  
718 accounted for uncertainty in gene tree topologies by collapsing branches with bootstrap support  
719 <80%.

50  
51 721 *Inferring footprints of introgression*

52  
53  
54  
55  
56  
57  
58  
59  
60  
722 To infer footprints of introgression across the entire species tree, we estimated Patterson's D  
723 (Durand et al. 2011) and related statistics in Dsuite (Malinsky et al. 2021) based on 58,963,109  
724 biallelic SNPs. D and f4 statistics were estimated across all possible combinations of trios in our  
725 38 wheatear taxa. We used Dtrios to calculate the sums of three different patterns (BABA, BBAA  
726 and, ABBA) and D and f4-ratio statistics for all 8,437 possible trios. Dsuite uses the standard block-  
727 jackknife procedure to assess the significance of the D statistic. Due to the large number of D-  
26

1  
2  
3 728 statistics comparisons and difficulties disentangling false positives that may arise due to ancient  
4 gene flow, we performed the f-branch test (fb) implemented in Dsuite to assign gene flow to  
5 specific internal branches on the species tree. Then we visualized the output using Dsuite's  
6 dttools.py script.  
7  
8 731

9 732 We then aimed to obtain further support for the footprints of introgression that were  
10 suggested in *lugens*, *picata* and *hispanica* complex by the above approach based on the D-statistics.  
11 733 To this end, for these three complexes, we estimated phylogenetic networks from maximum  
12 likelihood trees generated from BUSCOs using the pseudolikelihood (InferNetwork\_MPL) (Yu and  
13 Nakhleh 2015) and likelihood (CalGTPProb) (Yu et al. 2014) approaches implemented in phyloNet  
14 3.6.9 (Than et al. 2008). Due to the high computational demands, analyses were run for each of  
15 the clades containing signals of introgression in earlier analyses separately, namely for the *lugens*,  
16 *picata* and *hispanica* complexes. Furthermore, we only included BUSCO loci that had data  
17 available for all taxa of the respective complex. Outgroup species for each complex were selected  
18 based on the species tree. Analyses included 7,323 rooted gene trees for the *lugens* complex, 7,310  
19 rooted gene trees for the *picata* complex, and 7,335 rooted gene trees for the *hispanica* complex.  
20 738 For each complex, we allowed for one to five reticulation events, with the starting tree  
21 corresponding to the species tree topology (-s), 0.9 bootstrap threshold for gene trees (-b) and  
22 1,000 iterations (-x). To ensure convergence, the network searches were repeated 10 times. Then  
23 741 we estimated the likelihood by fixing the topology of the focal clade for the species tree (without  
24 any reticulation) and for each of the five networks (with different numbers of introgression edges)  
25 744 and calculated their likelihood scores. We determined the optimal network by calculating the  
26 Bayesian Information Criterion (BIC) from the maximum likelihood scores, the number of gene  
27 747 trees, the number of branch length being estimated, plus the number of admixture edges in each  
28 750 model. We used the browser-based tree viewer IcyTree (Vaughan 2017) to visualize the  
29 751 estimated networks (Vaughan 2017).  
30  
31 752  
32  
33 753  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 754 Acknowledgements

755 We thank all the natural history museums and their staff who provided material for this study:  
756 namely the American Museum of Natural History, New York; Natural History Museum, Tring; Field  
757 Museum of Natural History, Chicago; Natural History Museum of Los Angeles County, Los Angeles;  
758 Museum of Vertebrate Zoology, UC Berkeley; Natural History Museum, University of  
759 Oslo; Naturhistoriska riksmuseet, Stockholm; Texas A&M University Biodiversity Research and  
760 Teaching Collections, Texas College Station; University of Washington Burke Museum;, Seattle,  
761 Yale Peabody Museum; Zoological Museum, Natural History Museum of Denmark, Copenhagen;  
762 Zoologisches Forschungsmuseum König, Bonn; and Martin Haase, Vogelwarte Hiddensee,  
763 Universität Greifswald. Further samples were provided by José Luis Copete, and Marc Illa. We are  
764 indebted to the sequencing facilities of NGI Sweden in Solna and to the NGS platform of the  
765 University of Berne and their respective staff for their excellent services and to Marta Burri for  
766 sequencing library preparations. Computations were performed at the High-Performance  
767 Computing Cluster EVE, a joint effort of the Helmholtz Centre for Environmental Research (UFZ)  
768 and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. We thank  
769 the administration and support staff of EVE, Thomas Schnicke and Ben Langenberg (UFZ), and  
770 Christian Krause (iDiv). We thank Chris Rose and Claire Weatherhead from Bloomsbury  
771 Publishing Plc for their permission to use bird drawings in our figures. This research was  
772 supported by a German Research Foundation (DFG) research grant (BU3456/3-1) to RB, the  
773 National Research Fund (FNR), Luxembourg, grant number 14575729 to DL, and a Georg Foster  
774 Research Stipend of the Alexander von Humboldt Foundation and a scholarship for female  
775 researchers from Friedrich-Schiller-University Jena, both to NAK.

## 776 References

- 777 Abadi S, Azouri D, Pupko T, Mayrose I. 2019. Model selection may not be a mandatory step for  
778 phylogeny reconstruction. *Nature communications*. 10(1):1-11.
- 779 Alaei Kakhki N, Aliabadian M, Förschler MI, Ghasempouri SM, Kiabi BH, Verde Arregoitia LD,  
780 Schweizer M. 2018. Phylogeography of the oenanthe hispanica-pleschanka-cypriaca  
781 complex (aves, muscicapidae: Saxicolinae): Diversification history of open-habitat  
782 specialists based on climate niche models, genetic data, and morphometric data. *Journal of Zoological Systematics and Evolutionary Research*. 56(3):408-427.
- 783 Alaei Kakhki N, Aliabadian M, Schweizer M. 2016. Out of africa: Biogeographic history of the open-  
784 habitat chats (aves, muscicapidae: Saxicolinae) across arid areas of the old world.  
785 *Zoologica Scripta*. 45(3):237-251.
- 786 Aliabadian M, Kaboli M, Förschler MI, Nijman V, Chamani A, Tillier A, Prodon R, Pasquet E, Ericson  
787 PG, Zuccon D. 2012. Erratum to: Convergent evolution of morphological and ecological  
788 traits in the open-habitat chat complex (aves, muscicapidae: Saxicolinae). *Molecular  
789 Phylogenetics Evolution*. 65(3):35-45.
- 790

- 1  
2  
3 791 Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020. Mitofinder:  
4 792 Efficient automated large-scale extraction of mitogenomic data in target enrichment  
5 793 phylogenomics. *Molecular ecology resources*. 20(4):892-905.  
6  
7 794 Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2007. Bayesian estimation of concordance among  
8 795 gene trees. *Molecular biology and evolution*. 24(2):412-426.  
9  
10 796 Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: What have we learned  
11 797 about the genetics of adaptation? *Trends in ecology & evolution*. 23(1):26-32.  
12  
13 798 Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in ecology &*  
14 799 *evolution*. 23(1):38-44.  
15  
16 800 Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin P-A. 2009. Phylogenomics of c4  
17 801 photosynthesis in sedges (cyperaceae): Multiple appearances and genetic convergence.  
18 802 *Molecular Biology and Evolution*. 26(8):1909-1919.  
19  
20 803 BirdLife International and Handbook of the Birds of the World (2016). Bird species distribution  
maps of the world. Version 6.0.  
21  
22 805 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones  
23 806 G, Kühnert D, De Maio N. 2019. Beast 2.5: An advanced software platform for bayesian  
24 807 evolutionary analysis. *PLoS computational biology*. 15(4):e1006650.  
25  
26 808 Bouckaert RR, Drummond AJ. 2017. Bmodeltest: Bayesian phylogenetic site model averaging and  
model comparison. *BMC evolutionary biology*. 17(1):1-11.  
27  
28 810 Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the  
29 presence of recombination. *Genetics*. 172(4):2665-2681.  
30  
31 812 Brusatte SL, O'Connor JK, Jarvis ED. 2015. The origin and diversification of birds. *Current Biology*.  
32 813 25(19):R888-R898.  
33  
34 814 Campagna L, Repenning M, Silveira LF, Fontana CS, Tubaro PL, Lovette IJ. 2017. Repeated  
35 815 divergent selection on pigmentation genes in a rapid finch radiation. *Science advances*.  
36 816 3(5):e1602404.  
37  
38 817 Chakrabarty P, Faircloth BC, Alda F, Ludt WB, McMahan CD, Near TJ, Dornburg A, Albert JS,  
39 818 Arroyave J, Stiassny ML. 2017. Phylogenomic systematics of ostariophysan fishes:  
40 819 Ultraconserved elements support the surprising non-monophyly of characiformes.  
Systematic Biology. 66(6):881-895.  
41  
42 821 Chen S, Zhou Y, Chen Y, Gu JJB. 2018. Fastp: An ultra-fast all-in-one fastq preprocessor.  
43 822 34(17):i884-i890.  
44  
45 823 Chifman J, Kubatko L. 2014. Quartet inference from snp data under the coalescent model.  
Bioinformatics. 30(23):3317-3324.  
46  
47 825 Chou J, Gupta A, Yaduvanshi S, Davidson R, Nute M, Mirarab S, Warnow T. 2015. A comparative  
48 826 study of svdquartets and other coalescent-based species tree estimation methods. *BMC*  
49 827 genomics. 16(10):1-11.  
50  
51 828 Christin P-A, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in  
52 829 grasses via parallel adaptive genetic changes. *Current biology*. 17(14):1241-1247.  
53  
54 830 Christin P-A, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic  
convergence. *Trends in Genetics*. 26(9):400-405.  
55  
56 832 Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers  
57 833 RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by  
58 834 repeated fixation of ectodysplasin alleles. *Science*. 307(5717):1928-1933.  
59  
60 835 Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, Bell MA, Kimmel CB, Postlethwait  
836 JH. 2004. Parallel genetic basis for repeated evolution of armor loss in alaskan threespine

- 1  
2  
3 837           stickleback populations. *Proceedings of the National Academy of Sciences.* 101(16):6050-  
4 838           6055.  
5  
6 839 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth  
7           GT, Sherry ST. 2011. The variant call format and vcftools. *Bioinformatics.* 27(15):2156-  
8 841           2158.  
9  
10 842 Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees.  
11           *PLoS genetics.* 2(5):e68.  
12  
13 844 Doyle JJ. 1997. Trees within trees: Genes and species, molecules and morphology. *Systematic  
14           Biology.* 46(3):537-553.  
15  
16 846 Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely  
17           related populations. *Molecular biology and evolution.* 28(8):2239-2252.  
18  
19 848 Durand S, Bouché N, Strand EP, Loudet O, Camilleri C. 2012. Rapid establishment of genetic  
20           incompatibility through natural epigenetic variation. *Current Biology.* 22(4):326-331.  
21  
22 850 Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van  
23           Belleghem SM, Patterson N, Neafsey DE. 2019. Genomic architecture and introgression  
24           shape a butterfly radiation. *Science.* 366(6465):594-599.  
25  
26 853 Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H,  
27           Nadachowska-Brzyska K, Qvarnström A. 2012. The genomic landscape of species  
28           divergence in ficedula flycatchers. *Nature.* 491(7426):756-760.  
29  
30 856 Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: Insights from parallelism  
31           and convergence. *Trends in ecology & evolution.* 26(6):298-306.  
32  
33 858 Enciso-Romero J, Pardo-Díaz C, Martin SH, Arias CF, Linares M, McMillan WO, Jiggins CD, Salazar  
34           C. 2017. Evolution of novel mimicry rings facilitated by adaptive introgression in tropical  
35           butterflies. *Molecular ecology.* 26(19):5160-5172.  
36  
37 861 Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nature  
38           Reviews Genetics.* 8(8):610-618.  
39  
40 863 Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork VL. 2017. Phylogenomic inferences from  
41           reference-mapped and de novo assembled short-read sequence data using radseq  
42           sequencing of california white oaks (quercus section quercus). *Genome.* 60(9):743-755.  
43  
44 866 Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and  
45           consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology,  
46           Evolution, and Systematics.* 34(1):397-423.  
47  
48 869 Genner MJ, Turner GF. 2012. Ancient hybridization and phenotypic novelty within lake malawi's  
49           cichlid fish radiation. *Molecular Biology and Evolution.* 29(1):195-206.  
50  
51 871 Gill F, Donsker D, Rasmussen P. 2020. Ioc world bird list (v10. 1). IOC World Bird List [consultado  
52           el 26 de julio de 2020] DOI: <https://doi.org/10.14344/IOC.ML.10>.  
53  
54 873 Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. *Developmental  
55           biology.* 332(1):36-47.  
56  
57 875 Grant PR, Grant BR, Petren K. 2005. Hybridization in the recent past. *The American Naturalist.*  
58           166(1):56-67.  
59  
60 877 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-  
Y. 2010. A draft sequence of the neandertal genome. *Science.* 328(5979):710-722.  
878  
879 Gross JB, Borowsky R, Tabin CJ. 2009. A novel role for mc1r in the parallel evolution of  
880           depigmentation in independent populations of the cavefish astyanax mexicanus. *PLoS  
881           genetics.* 5(1):e1000326.

- 1  
2  
3 882 Haffer J. 1977. Secondary contact zones of birds in northern iran. Bonner Zoologische  
4 883 Monographien. 10:1-64.  
5  
6 884 Han F, Lamichhaney S, Grant PR, Andersson L, Webster MT. 2017. Gene flow, ancient  
7 885 polymorphism, and ecological adaptation shape the genomic landscape of divergence  
8 886 among darwin's finches. *Genome research.* 27(6):1004-1015.  
9  
10 887 He C, Liang D, Zhang P. 2020. Asymmetric distribution of gene trees can arise under purifying  
11 888 selection if differences in population size exist. *Molecular biology and evolution.*  
12 889 37(3):881-892.  
13  
14 890 Hedrick PW. 2013. Adaptive introgression in animals: Examples and comparison to new mutation  
15 891 and standing variation as sources of adaptive variation. *Molecular ecology.* 22(18):4606-  
16 892 4618.  
17  
18 893 Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of  
19 894 mimicry adaptations among species. *Nature.* 487:94-98.  
20  
21 895 Hermission J, Pennings PS. 2005. Soft sweeps: Molecular population genetics of adaptation from  
22 896 standing genetic variation. *Genetics.* 169(4):2335-2352.  
23  
24 897 Hibbins MS, Hahn MW. 2022. Phylogenomic approaches to detecting and characterizing  
25 898 introgression. *Genetics.* 220(2):iyab173.  
26  
27 899 Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. Ufboot2: Improving the  
28 900 ultrafast bootstrap approximation. *Molecular biology and evolution.* 35(2):518-522.  
29  
30 901 Hoballah ME, Gübitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo A, Arnold M,  
31 902 Kuhlemeier C. 2007. Single gene-mediated shift in pollinator attraction in petunia. *The  
32 903 Plant Cell.* 19(3):779-790.  
33  
34 904 Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns  
35 905 among human, chimpanzee, and orangutan suggest recent orangutan speciation and  
36 906 widespread selection. *Genome research.* 21(3):349-356.  
37  
38 907 Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT.  
39 908 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds.  
40 909 *Science.* 346(6215):1320-1331.  
41  
42 910 Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC,  
43 911 White S. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature.*  
44 912 484(7392):55-61.  
45  
46 913 Junier T, Zdobnov EM. 2010. The newick utilities: High-throughput phylogenetic tree processing  
47 914 in the unix shell. *Bioinformatics.* 26(13):1669-1670.  
48  
49 915 Katoh K, Standley DM. 2013. Mafft multiple sequence alignment software version 7:  
50 916 Improvements in performance and usability. *Molecular biology and evolution.* 30(4):772-  
51 917 780.  
52  
53 918 Konečná V, Bray S, Vlček J, Bohutínská M, Požárová D, Choudhury RR, Bollmann-Giolai A, Flis P,  
54 919 Salt DE, Parisod C. 2021. Parallel adaptation in autopolyploid *arabidopsis arenosa* is  
55 920 dominated by repeated recruitment of shared alleles. *Nature communications.* 12(1):1-  
56 921 13.  
57  
58 922 Korunes KL, Samuk K. 2021. Pixy: Unbiased estimation of nucleotide diversity and divergence in  
59 923 the presence of missing data. *Molecular ecology resources.* 21(4):1359-1368.  
60  
61 924 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. Mega x: Molecular evolutionary genetics  
62 925 analysis across computing platforms. *Molecular biology and evolution.* 35(6):1547.

- 1  
2  
3 926 Lamers RP, Muthukrishnan G, Castoe TA, Tafur S, Cole AM, Parkinson CL. 2012. Phylogenetic  
4 927 relationships among staphylococcus species and refinement of cluster groups based on  
5 928 multilocus data. *BMC evolutionary biology*. 12(1):1-15.  
6  
7 929 Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid  
8 930 speciation in darwin's finches. *Science*. 359(6372):224-228.  
9  
10 931 Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. Partitionfinder 2: New methods for  
11 932 selecting partitioned models of evolution for molecular and morphological phylogenetic  
12 933 analyses. *Molecular biology and evolution*. 34(3):772-773.  
13  
14 934 Larget BR, Kotha SK, Dewey CN, Ané C. 2010. Bucky: Gene tree/species tree reconciliation with  
15 935 bayesian concordance analysis. *Bioinformatics*. 26(22):2910-2911.  
16  
17 936 Larsson A. 2014. Aliview: A fast and lightweight alignment viewer and editor for large datasets.  
18 937 *Bioinformatics*. 30(22):3276-3278.  
19  
20 938 Lerner HR, Meyer M, James HF, Hofreiter M, Fleischer RC. 2011. Multilocus resolution of  
21 939 phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers.  
940 *Current Biology*. 21(21):1838-1844.  
22  
23 941 Li H. 2011. A statistical framework for snp calling, mutation discovery, association mapping and  
24 942 population genetical parameter estimation from sequencing data. *Bioinformatics*.  
25 943 27(21):2987-2993.  
26  
27 944 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arXiv  
28 945 preprint arXiv:13033997.  
29  
30 946 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009.  
31 947 The sequence alignment/map format and samtools. *Bioinformatics*. 25(16):2078-2079.  
32  
33 948 Linkem CW, Minin VN, Leaché AD. 2016. Detecting the anomaly zone in species trees and evidence  
34 949 for a misleading signal in higher-level skink phylogeny (squamata: Scincidae). *Systematic  
950 biology*. 65(3):465-477.  
35  
36 951 Lutgen D, Ritter R, Olsen RA, Schielzeth H, Gruselius J, Ewels P, García JT, Shirihai H, Schweizer M,  
37 952 Suh A. 2020. Linked-read sequencing enables haplotype-resolved resequencing at  
38 953 population scale. *Molecular Ecology Resources*. 20(5):1311-1322.  
39  
40 954 Maddison WP. 1997. Gene trees in species trees. *Systematic biology*. 46(3):523-536.  
41  
42 955 Malinsky M, Matschiner M, Svardal H. 2021. Dsuite-fast d-statistics and related admixture  
43 956 evidence from vcf files. *Molecular Ecology Resources*. 21(2):584-595.  
44  
45 957 Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome  
46 958 sequences of malawi cichlids reveal multiple radiations interconnected by gene flow.  
47 959 *Nature ecology & evolution*. 2(12):1940-1955.  
48  
49 960 Marques DA, Lucek K, Sousa VC, Excoffier L, Seehausen O. 2019a. Admixture between old lineages  
50 961 facilitated contemporary ecological speciation in lake constance stickleback. *Nature  
51 962 communications*. 10(1):1-14.  
52  
53 963 Marques DA, Meier JI, Seehausen O. 2019b. A combinatorial view on speciation and adaptive  
54 964 radiation. *Trends in ecology & evolution*. 34(6):531-544.  
55  
56 965 Martin A, Orgogozo V. 2013. The loci of repeated evolution: A catalog of genetic hotspots of  
57 966 phenotypic variation. *Evolution*. 67(5):1235-1250.  
58  
59 967 Mayr E, Stresemann E. 1950. Polymorphism in the chat genus oenanthe (aves). *Evolution*. 4:291-  
60 968 300.  
61  
62 969 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,  
63 970 Gabriel S, Daly M. 2010. The genome analysis toolkit: A mapreduce framework for  
64 971 analyzing next-generation DNA sequencing data. *Genome research*. 20(9):1297-1303.

- 1  
2  
3 972 Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization  
4 973 fuels rapid cichlid fish adaptive radiations. *Nature communications*. 8(1):1-11.  
5  
6 974 Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of parallel ecological  
7 975 speciation in lake victoria cichlids. *Molecular biology and evolution*. 35(6):1489-1506.  
8  
9 976 Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for  
10 977 phylogenomic datasets. *Molecular biology and evolution*. 37(9):2727-2733.  
11  
12 978 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R.  
13 979 2020b. Iq-tree 2: New models and efficient methods for phylogenetic inference in the  
14 980 genomic era. *Molecular biology and evolution*. 37(5):1530-1534.  
15  
16 981 Montejo-Kovacevich G, Meier JI, Bacquet CN, Warren IA, Chan YF, Kucka M, Salazar C, Rueda N,  
17 982 Montgomery SH, McMillan WO. 2021. Repeated genetic adaptation to high altitude in two  
18 983 tropical butterflies. *bioRxiv*.  
19  
20 984 Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Muñoz-Fuentes V, Green AJ, Kopuchian C,  
21 985 Tubaro PL, Alza L, Bulgarella M. 2015. Convergent evolution of hemoglobin function in  
22 986 high-altitude andean waterfowl involves limited parallelism at the molecular sequence  
23 987 level. *PLoS genetics*. 11(12):e1005681.  
24  
25 988 Nater A, Burri R, Kawakami T, Smeds L, Ellegren J. 2015. Resolving evolutionary relationships in  
26 989 closely related species with whole-genome sequencing data. *Systematic Biology*.  
27 990 64(6):1000-1017.  
28  
29 991 Outlaw RK, Voelker G, Bowie RC. 2010. Shall we chat? Evolutionary relationships in the genus  
30 992 cercomela (muscicapidae) and its relation to oenanthe reveals extensive polyphyly among  
31 993 chats distributed in africa, india and the palearctic. *Molecular Phylogenetics and*  
32 994 *Evolution*. 55(1):284-292.  
33  
34 995 Panov E. 1992. Emergence of hybridogenous polymorphism in the oenanthe picata complex.  
35 996 *Bulletin of the British Ornithologists' Club Centenary Volume*. 112:237-249.  
36  
37 997 Panov EN. 2005. Wheatears of palearctic. Pensoft.  
38  
39 998 Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins  
40 999 CD. 2012. Adaptive introgression across species boundaries in heliconius butterflies. *PLoS*  
41 1000 *Genet*. 8(6):e1002752.  
42  
43 1001 Paterson RS, Rybczynski N, Kohno N, Maddin HC. 2020. A total evidence phylogenetic analysis of  
44 1002 pinniped phylogeny and the possibility of parallel evolution within a monophyletic  
45 1003 framework. *Frontiers in Ecology and Evolution*. 7:457.  
46  
47 1004 Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive  
48 1005 variation during a rapid radiation. *PLoS biology*. 14(2):e1002379.  
49  
50 1006 Peona V, Palacios-Gimenes O, Lutgen D, Olsen R, Alaei Kakhki N, Andriopoulos P, Bontzorlos V,  
51 1007 Schweizer M, Suh A, Burri R. in prep. A chromosome-scale reference genome for eastern  
52 1008 black-eared wheatear (oenanthe melanoleuca).  
53  
54 1009 Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M,  
55 1010 Grabherr MG. 2014. The genomic landscape underlying phenotypic integrity in the face of  
56 1011 gene flow in crows. *Science*. 344(6190):1410-1414.  
57  
58 1012 Preston JC, Hileman LC. 2009. Developmental genetics of floral symmetry evolution. *Trends in*  
59 1013 *plant science*. 14(3):147-154.  
60  
61 1014 Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ.  
62 1015 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of  
63 1016 albinism. *Nature genetics*. 38(1):107-111.  
64  
65 1017 Rambaut A, Drummond A. 2009. Tracer v1. 5.0. Available at: [Http://beast.Bio.Ed.Ac.Uk/tracer](http://beast.Bio.Ed.Ac.Uk/tracer).

- 1  
2  
3 1018 Randler C, Förschler MI, Gonzalez J, Aliabadian M, Bairlein F, Wink M. 2012. Phylogeography, pre-  
4 1019 zygotic isolation and taxonomic status in the endemic cyprus wheatear *oenanthe cypriaca*.  
5 1020 Journal of Ornithology. 153(2):303-312.  
6  
7 1021 Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard  
8 1022 MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and  
9 1023 model choice across a large model space. Systematic biology. 61(3):539-542.  
10  
11 1024 Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying  
12 1025 incongruence among phylogenetic trees. Molecular biology and evolution. 31(5):1261-  
13 1026 1271.  
14  
15 1027 Sangster G, Alström P, Forsmark E, Olsson U. 2010. Multi-locus phylogenetic analysis of old world  
16 1028 chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level  
17 1029 (aves: Muscicapidae). Molecular Phylogenetics and Evolution. 57(1):380-392.  
18  
19 1030 Sayyari E, Mirarab S. 2018. Testing for polytomies in phylogenetic species trees using quartet  
20 1031 frequencies. Genes. 9(3):132.  
21  
22 1032 Sayyari E, Whitfield JB, Mirarab S. 2018. Discovista: Interpretable visualizations of gene tree  
23 1033 discordance. Molecular Phylogenetics and Evolution. 122:110-115.  
24  
25 1034 Schweizer M, Shirihai H. 2013. Phylogeny of the oenanthe lugens complex (aves, muscipapidae:  
26 1035 Saxicolinae): Paraphyly of a morphologically cohesive group within a recent radiation of  
1036 open-habitat chats. Molecular Phylogenetics Evolution. 69(3):450-461.  
27  
28 1037 Schweizer M, Warmuth V, Alaei Kakhki N, Aliabadian M, Förschler M, Shirihai H, Suh A, Burri R.  
29 1038 2019a. Parallel plumage colour evolution and introgressive hybridization in wheatears.  
30 1039 Journal of evolutionary biology. 32(1):100-110.  
31  
32 1040 Schweizer M, Warmuth VM, Kakhki NA, Aliabadian M, Förschler M, Shirihai H, Ewels P, Gruselius  
33 1041 J, Olsen R-A, Schielzeth H et al. 2019b. Genome-wide evidence supports mitochondrial  
34 1042 relationships and pervasive parallel phenotypic evolution in open-habitat chats.  
35 1043 Molecular phylogenetics evolution. 139:106568.  
36  
37 1044 Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P,  
38 1045 Bank C, Bränström Å. 2014. Genomics and the origin of species. Nature Reviews Genetics.  
1046 15(3):176-192.  
39  
40 1047 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. Busco: Assessing  
41 1048 genome assembly and annotation completeness with single-copy orthologs.  
42 1049 Bioinformatics. 31(19):3210-3212.  
43  
44 1050 Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict,  
45 1051 concordance, and gene duplications with examples from animals and plants. BMC  
46 1052 evolutionary biology. 15(1):1-15.  
47  
48 1053 Song Y, Endepols S, Klemann N, Richter D, Matuschka F-R, Shih C-H, Nachman MW, Kohn MH.  
49 1054 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization  
1055 between old world mice. Current Biology. 21(15):1296-1301.  
50  
51 1056 Stahl BA, Gross JB. 2015. Alterations in *mc1r* gene expression are associated with regressive  
52 1057 pigmentation in *astyanax* cavefish. Development genes and evolution. 225(6):367-375.  
53  
54 1058 Stenz NW, Larget B, Baum DA, Ané C. 2015. Exploring tree-like and non-tree-like patterns using  
55 1059 genome sequences: An example using the inbreeding plant species *arabidopsis thaliana*  
56 1060 (L.) heynei. Systematic Biology. 64(5):809-823.  
57  
58 1061 Stern DL. 2013. The genetic causes of convergent evolution. Nature Reviews Genetics. 14(11):751-  
59 1062 764.  
60  
61 1063 Stryjewski KF, Sorenson MD. 2017. Mosaic genome evolution in a recent and rapid avian radiation.  
1064 Nature Ecology & Evolution. 1(12):1912-1922.

- 1  
2  
3 1065 Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of  
4 1066 neoaves. *Zoologica Scripta*. 45:50-62.  
5  
6 1067 Suh A, Smeds L, Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient  
7 1068 adaptive radiation of neoavian birds. *PLoS Biol*. 13(8):e1002224.  
8  
9 1069 Swofford D. 2003. Paup\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4  
10 1070 Sinauer Associates, Sunderland, Massachusetts.  
11  
12 1071 Tamura K, Stecher G, Kumar S. 2021. Mega11: Molecular evolutionary genetics analysis version  
13 1072 11. *Molecular biology and evolution*. 38(7):3022-3027.  
14  
15 1073 Than C, Ruths D, Nakhleh L. 2008. Phylonet: A software package for analyzing and reconstructing  
16 1074 reticulate evolutionary relationships. *BMC bioinformatics*. 9(1):1-16.  
17  
18 1075 Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in  
19 1076 animals. *Molecular ecology*. 21(16):3907-3930.  
20  
21 1077 Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Möst M, Rastas P, De Meester L, Hendrickx  
22 F. 2018. Evolution at two time frames: Polymorphisms from an ancient singular  
23 1078 divergence event fuel contemporary parallel evolution. *PLoS genetics*. 14(11):e1007796.  
24  
25 1080 Wallbank RW, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, Dasmahapatra KK, Salazar C,  
26 1081 Joron M, Nadeau N. 2016. Evolutionary novelty in a butterfly wing pattern through  
27 1082 enhancer shuffling. *PLoS biology*. 14(1):e1002353.  
28  
29 1083 Whittall JB, Voelckel C, Kliebenstein DJ, Hodges SA. 2006. Convergence, constraint and the role of  
30 1084 gene expression during adaptive radiation: Floral anthocyanins in aquilegia. *Molecular  
31 1085 Ecology*. 15(14):4645-4657.  
32  
33 1086 Wu M, Kostyun JL, Hahn MW, Moyle LC. 2018. Dissecting the basis of novel trait evolution in a  
34 1087 radiation with widespread phylogenetic discordance. *Molecular Ecology*. 27(16):3301-  
35 1088 3316.  
36  
37 1089 Yu Y, Dong J, Liu KJ, Nakhleh L. 2014. Maximum likelihood inference of reticulate evolutionary  
38 1090 histories. *Proceedings of the National Academy of Sciences*. 111(46):16448-16453.  
39  
40 1091 Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC  
41 1092 genomics*. 16(10):1-10.  
42  
43 1093 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. Astral-iii: Polynomial time species tree  
44 1094 reconstruction from partially resolved gene trees. *BMC bioinformatics*. 19(6):15-30.  
45  
46 1095 Zuccon D, Ericson PG. 2010. A multi-gene phylogeny disentangles the chat-flycatcher complex  
47 (aves: Muscicapidae). *Zoologica Scripta*. 39(3):213-224.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1099 **Table 1 | Sampling and sequence data information.**  
4 1100

Taxon	Collection, no. <sup>a</sup>	Tissue	Library Type	Mapping %	Coverage
<i>Campicoloides bifasciatus</i>	MVZ, RSA073,	Muscle	Illumina DNA Prep Kit	99.1	8.5
<i>Emarginata schlegelii</i>	FMNH, 453197,	Muscle	Illumina DNA Prep Kit	99.1	9.5
<i>Emarginata sinuata</i>	UWMB, 95470	Muscle	ThruPLEX DNA-Seq Kit	96.7	9.7
<i>Myrmecocichla aethiops</i>	MVZ, 129113	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	93.3	6.8
<i>Myrmecocichla arnotti</i>	FMNH, 468111	Toepad	Illumina DNA Prep Kit	99.3	12.8
<i>Myrmecocichla formicivora</i>	MVZ, RSA205	Muscle	Illumina DNA Prep Kit	98.6	8.3
<i>Myrmecocichla melaena</i>	A1153	Blood	ThruPLEX DNA-Seq Kit	98.8	13.3
<i>Myrmecocichla monticola</i>	NMBE, 1043860	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	94.3	7.6
<i>Myrmecocichla nigra</i>	NRM, 570041	Toepad	ThruPLEX DNA-Seq Kit	94.5	7.0
<i>Myrmecocichla tholloni</i>	YPM, ORN95640	Toepad	ThruPLEX DNA-Seq Kit	92.1	7.3
<i>Oenanthe (C.) dubia</i>	FMNH, 83201	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	93.0	6.8
<i>Oenanthe (C.) familiaris</i>	MVZ, G0866	Blood	Illumina DNA Prep Kit	99.4	13.3
<i>Oenanthe (C.) fusca</i>	YPM, ORN011707	Toepad	ThruPLEX DNA-Seq Kit	86.4	4.7
<i>Oenanthe (C.) melanura</i>	A1203	Blood	ThruPLEX DNA-Seq Kit	98.8	15.4
<i>Oenanthe (C.) scotocerca</i>	LACM, 61131	Toepad	ACCEL-NGS 1S DNA Library Prep Kit	96.6	8.3
<i>Oenanthe (M.) albifrons clericalis</i>	NRM, 558941	Toepad	ThruPLEX DNA-Seq Kit	95.4	6.3
<i>Oenanthe (M.) albifrons frontalis</i>	KU, 115365	Muscle	ThruPLEX DNA-Seq Kit	98.5	12.7
<i>Oenanthe albonigra</i>	IR-KIL-010	Blood	ThruPLEX DNA-Seq Kit	99.2	12.9
<i>Oenanthe bottae frenata</i>	NRM, 558917	Toepad	ThruPLEX DNA-Seq Kit	97.0	6.0
<i>Oenanthe chrysopygia</i>	IR-FIR-002	Blood	ThruPLEX DNA-Seq Kit	98.9	11.5
<i>Oenanthe cypriaca</i>	19e	Blood	Chromium Genome Library kit	99.8	40.6
<i>Oenanthe deserti</i>	MO-BOULMANE-2013	Blood	ThruPLEX DNA-Seq Kit	99.3	12.6
<i>Oenanthe finschii</i>	IR-ESF-004	Blood	ThruPLEX DNA-Seq Kit	98.9	8.7
<i>Oenanthe halophila</i>	3Y42902	Blood	Illumina DNA PCR-free	99.6	17.4
<i>Oenanthe heuglinii</i>	ZFMK, H.II.16p2.α	Dry skin	ACCEL-NGS 1S DNA Library Prep Kit	93.6	5.7

3	<i>Oenanthe hispanica</i>	E-GUI-013	Blood	Chromium Genome Library kit	99.8	15.5
4	<i>Oenanthe isabellina</i>	GR-LES-001	Blood	ThruPLEX DNA-Seq Kit	99.2	9.9
5	<i>Oenanthe leucopyga leucopyga</i>	A1137	Blood	ThruPLEX DNA-Seq Kit	98.8	10.7
6	<i>Oenanthe leucura leucura</i>	E-MAT-2012	Blood	ThruPLEX DNA-Seq Kit	99.0	9.5
7	<i>Oenanthe lugens lugens</i>	9b	Blood	ThruPLEX DNA-Seq Kit	98.7	22.0
8	<i>Oenanthe lugens persica</i>	ZMUC, 137759	Muscle	ThruPLEX DNA-Seq Kit	98.6	10.4
9	<i>Oenanthe lugentoides lugentoides</i>	NHMUK, 1965.M.12140	Toepad	ThruPLEX DNA-Seq Kit	96.9	4.6
10	<i>Oenanthe lugentoides boscaweni</i>	NHMUK, 1977.M.21.36	Toepad	ThruPLEX DNA-Seq Kit	97.2	6.8
11	<i>Oenanthe lugubris lugubris</i>	A1129	Blood	ThruPLEX DNA-Seq Kit	99.0	12.5
12	<i>Oenanthe lugubris vaurieei</i>	AMNH, 461151	Toepad	ThruPLEX DNA-Seq Kit	95.9	5.3
13	<i>Oenanthe melanoleuca</i>	IT-GRA-006	Blood	Chromium Genome Library kit	99.7	12.6
14	<i>Oenanthe moesta</i>	A1109	Blood	ThruPLEX DNA-Seq Kit	98.8	12.0
15	<i>Oenanthe monacha</i>	A1174	Blood	ThruPLEX DNA-Seq Kit	99.3	17.2
16	<i>Oenanthe oenanthe</i>	GEO-VAR-002	Blood	ThruPLEX DNA-Seq Kit	99.1	12.2
17	<i>Oenanthe phillipsi</i>	YPM, ORN035210	Toepad	ThruPLEX DNA-Seq Kit	95.6	7.9
18	<i>Oenanthe picata capistrata</i>	ZMUC, 29495	Toepad	ThruPLEX DNA-Seq Kit	96.7	5.3
19	<i>Oenanthe picata opistholeuca</i>	ZMUC, 29578	Toepad	ThruPLEX DNA-Seq Kit	97.1	5.2
20	<i>Oenanthe picata picata</i>	IR-TAN-005	Blood	ThruPLEX DNA-Seq Kit	98.8	14.2
21	<i>Oenanthe pileata</i>	TCWC, 15606	Muscle	Illumina DNA Prep Kit	98.9	13.4
22	<i>Oenanthe pleschanka</i>	CN-XS-006	Blood	Chromium Genome Library kit	99.8	15.1
23	<i>Oenanthe seebohmi</i>	KA69373	Blood	ThruPLEX DNA-Seq Kit	99.2	11.2
24	<i>Oenanthe warriae</i>	12c	Blood	ThruPLEX DNA-Seq Kit	98.7	13.4
25	<i>Oenanthe xanthoprymna</i>	NHMO, 15188	Blood	ThruPLEX DNA-Seq Kit	99.0	12.9
26	<i>Pinarochroa sordida</i>	YPM, ORN80066	Toepad	ThruPLEX DNA-Seq Kit	95.4	6.5
27	<i>Thamnolaea cinnamomeiventris</i>	NRM, 20086147	Muscle	ThruPLEX DNA-Seq Kit	98.6	15.9

<sup>a</sup> AMNH: American Museum of Natural History; NHMUK: Natural History Museum, Tring; FMNH: Field Museum of Natural History; LACM: Natural History Museum of Los Angeles County; MVZ: Museum of Vertebrate Zoology, UC Berkeley; NHMO: Natural History Museum, University of Oslo; NRM: Naturhistoriska riksmuseet, Stockholm; TCWC: Texas A&M University Biodiversity Research and Teaching Collections; UWBM: University of Washington Burke Museum; YPM: Yale Peabody Museum; ZMUC: Zoological Museum, Natural History Museum of Denmark; ZFMK: Zoologisches Forschungsmuseum König. Samples for which no institution is indicated are part of the research group's collection.

**Figure captions**

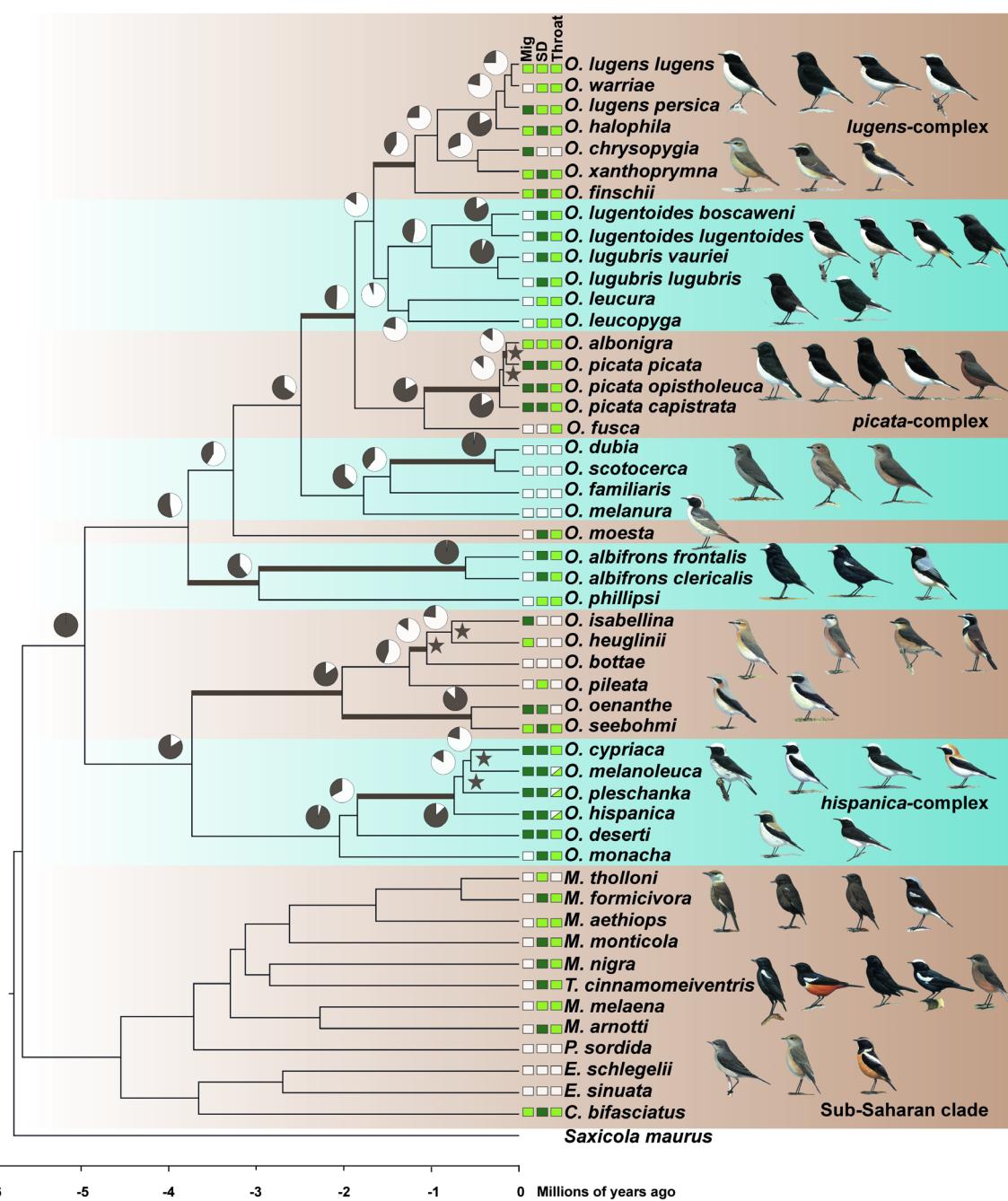
**Figure 1 | Time-calibrated phylogenetic tree of open-habitat chats and levels of ILS.** All nodes are supported by bootstrap values of 100. Pie charts depict the gene tree heterogeneity for each internal branch, with the brown proportion indicating the proportion of concordant gene trees (gCF). Coloured branches indicate internal branches for which ILS alone is statistically sufficient to explain the observed gene tree heterogeneity. Stars indicate branches that are in the phylogenetic anomaly zone. The character states of three selected characters: Sexual dimorphism (SD), monomorphic female-type (white), monomorphic male-type (pale green), dimorphic (dark green); Migratory behaviour (Mig), sedentary (white), short-distance migrant (pale green), long-distance migrant (dark green); and throat coloration (throat), white (white), black (pale green), and polymorphic (white and pale green). Drawing courtesy of Chris Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc.

**Figure 2 | Mito-nuclear discordances.** Shown are the time-calibrated phylogenetic trees based on nuclear data (left) and full mitogenomes (right).

**Figure 3 | Footprints of introgression as estimated by the f-branch statistic.** The heat map summarizes the f-branch statistics estimated in Dsuite. Darker colors depict increasing evidence for gene flow between lineages. Dotted lines in the phylogeny represent ancestral lineage.

**Figure 4 | Phylogenomic networks and distribution ranges for the *picata* (left), *lugens* (middle) and *hispanica* (right) complexes.** Phylogenomic networks were estimated under the maximum pseudolikelihood approach implemented in phyloNet. Numbers on the edges indicate the inheritance probabilities, which correspond to the proportion of gene trees supporting the reticulate relationship. Drawings courtesy of Chris Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc. Distribution ranges modified from BirdLife International and the Handbook of the Birds of the World (2016).

1132



**Figure 1 | Time-calibrated phylogenetic tree of open-habitat chats and levels of ILS.** All nodes are supported by bootstrap values of 100. Pie charts depict the gene tree heterogeneity for each internal branch, with the brown proportion indicating the proportion of concordant gene trees (gCF). Coloured branches indicate internal branches for which ILS alone is statistically sufficient to explain the observed gene tree heterogeneity. Stars indicate branches that are in the phylogenetic anomaly zone. The character states of three selected characters: Sexual dimorphism (SD), monomorphic female-type (white), monomorphic male-type (pale green), dimorphic (dark green); Migratory behaviour (Mig), sedentary (white), short-distance migrant (pale green), long-distance migrant (dark green); and throat coloration (throat), white (white), black (pale green), and polymorphic (white and pale green). Drawing courtesy of Chris Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc.

1133

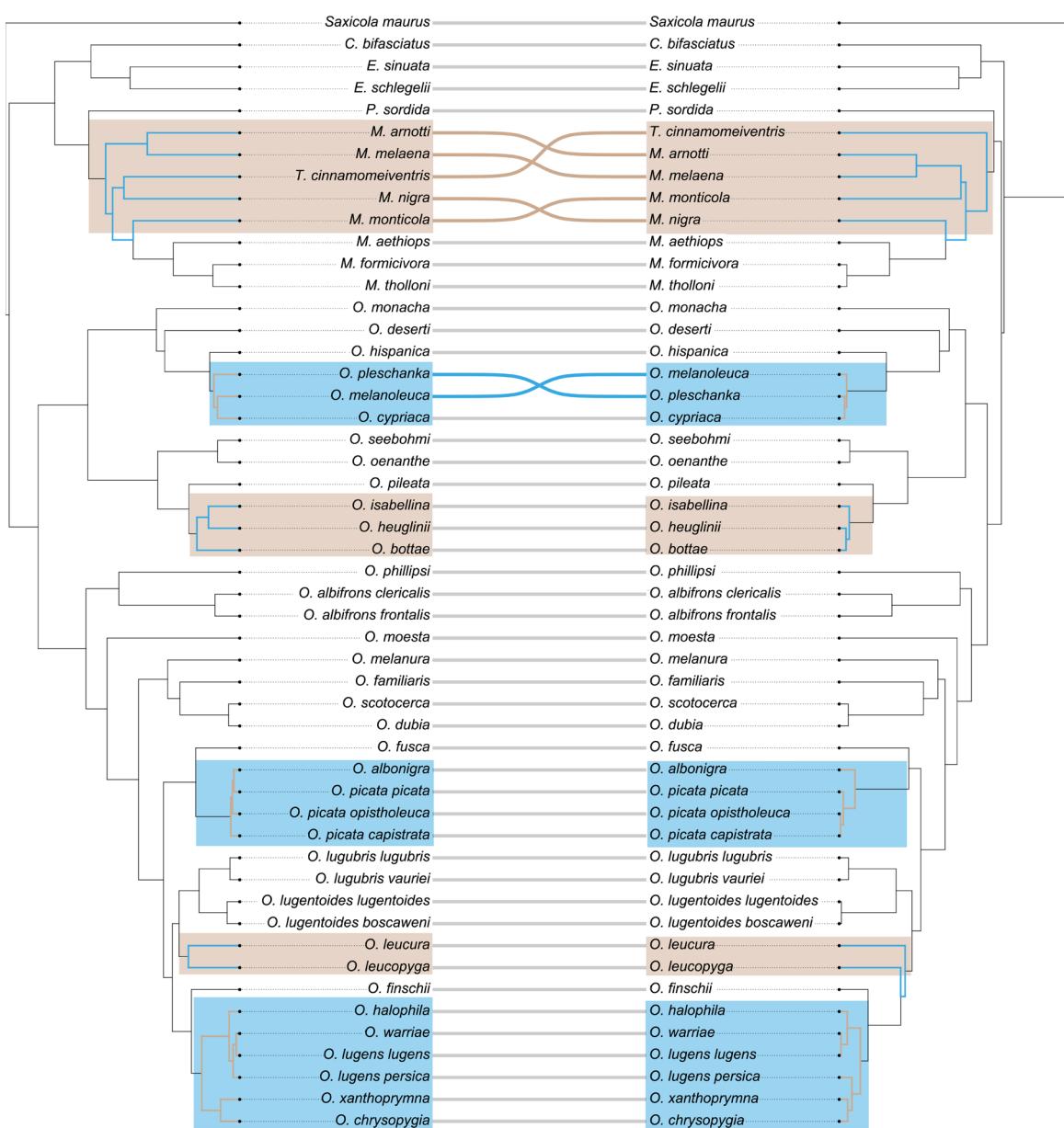
1  
2  
3 1134  
4  
5

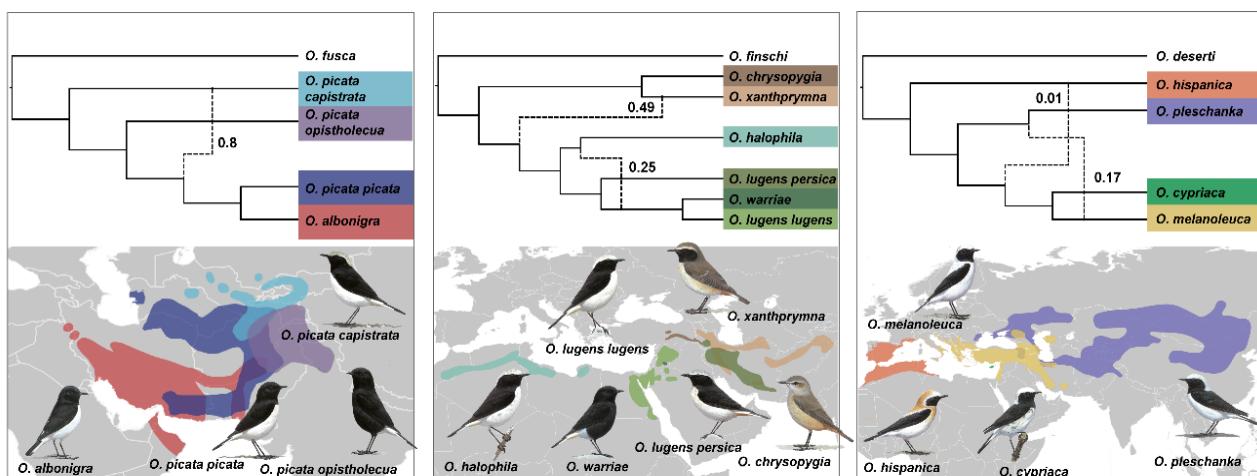
Figure 2 | Mito-nuclear discordances. Shown are the time-calibrated phylogenetic trees based on nuclear data (left) and full mitogenomes (right).

40

1135



**Figure 3 | Footprints of introgression as estimated by the f-branch statistic.** The heat map summarizes the f-branch statistics estimated in Dsuite. Darker colors depict increasing evidence for gene flow between lineages. Dotted lines in the phylogeny represent ancestral lineage.

1  
2  
3  
4  
5  
1136

**Figure 4 | Phylogenomic networks and distribution ranges for the *picata* (left), *lugens* (middle) and *hispanica* (right) complexes.** Phylogenomic networks were estimated under the maximum pseudolikelihood approach implemented in phyloNet. Numbers on the edges indicate the inheritance probabilities, which correspond to the proportion of gene trees supporting the reticulate relationship. Drawings courtesy of Chris Rose ([www.chrisrose-artist.co.uk](http://www.chrisrose-artist.co.uk)) with permission from Bloomsbury Publishing Plc. Distribution ranges modified from BirdLife International and the Handbook of the Birds of the World (2016).