

# Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species

PAOLO FRANCHINI,\*  IKER IRISARRI,\* ADAM FUDICKAR,†<sup>1</sup> ANDREAS SCHMIDT,† AXEL MEYER,\* MARTIN WIKELSKI\*† and JESKO PARTECKE\*†

\*Department of Biology, University of Konstanz, 78457 Konstanz, Germany, †Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, 78315 Radolfzell, Germany

## Abstract

Seasonal migration is a widespread phenomenon, which is found in many different lineages of animals. This spectacular behaviour allows animals to avoid seasonally adverse environmental conditions to exploit more favourable habitats. Migration has been intensively studied in birds, which display astonishing variation in migration strategies, thus providing a powerful system for studying the ecological and evolutionary processes that shape migratory behaviour. Despite intensive research, the genetic basis of migration remains largely unknown. Here, we used state-of-the-art radio-tracking technology to characterize the migratory behaviour of a partially migratory population of European blackbirds (*Turdus merula*) in southern Germany. We compared gene expression of resident and migrant individuals using high-throughput transcriptomics in blood samples. Analyses of sequence variation revealed a nonsignificant genetic structure between blackbirds differing by their migratory phenotype. We detected only four differentially expressed genes between migrants and residents, which might be associated with hyperphagia, moulting and enhanced DNA replication and transcription. The most pronounced changes in gene expression occurred between migratory birds depending on when, in relation to their date of departure, blood was collected. Overall, the differentially expressed genes detected in this analysis may play crucial roles in determining the decision to migrate, or in controlling the physiological processes required for the onset of migration. These results provide new insights into, and testable hypotheses for, the molecular mechanisms controlling the migratory phenotype and its underlying physiological mechanisms in blackbirds and other migratory bird species.

**Keywords:** blackbird, differential gene expression, migration, radio-tracking, transcriptomics, *Turdus merula*

Received 2 August 2016; revision received 10 March 2017; accepted 10 March 2017

## Introduction

Seasonal migration is one of the most spectacular phenomena in the animal kingdom. While the patterns of variation in migratory behaviour and its ecological and evolutionary causes have been intensively studied, its

genetic and physiological control still remains poorly understood (Newton 2008). This lack of knowledge is limiting our insight into not only the proximate control of migration, but also into its evolutionary origins. Information on these issues is especially relevant in the era of rapid global environmental change.

Migration requires a complex suite of morphological, physiological, sensory and behavioural adaptations that allow animals to return annually to the same breeding area (Berthold 1991; Dingle 2006). In some taxa, where young and adults migrate together in families or

Correspondence: Jesko Partecke, Fax: +49 (0) 7732-150165;

E-mail: partecke@orn.mpg.de

<sup>1</sup>Present address: Department of Biology, Indiana University, Bloomington, IN, 47405, USA

groups, such as in geese or cranes, the young appear to learn from older birds when and where to migrate (Mueller *et al.* 2013; Kolzsch *et al.* 2015). In most other species, however, individuals may migrate alone, unaided by conspecifics. These individuals may have to rely instead – particularly during their first migratory journey – on an inherited spatiotemporal migration program encoding when, where and how far to migrate (Akesson & Hedenstrom 2007). In birds, an excellent example of this latter strategy is the common cuckoo (*Cuculus canorus*) where young grow up with an unrelated host species, but still successfully migrate to sub-Saharan Africa (Willemoes *et al.* 2015). The textbook example of multigenerational migration is Monarch butterflies (*Danaus plexippus*) (Brower 1996), that have at least two nonmigratory generations during the migration cycle. This clearly demonstrates that the spatiotemporal migration program is endogenous, just as in some species of birds (Liedvogel *et al.* 2011). In this case, all individuals must carry the genetic migratory program, but it is only expressed in migratory individuals. The expression of the migratory phenotype most likely includes complex interactions between environmental cues and variation in underlying genetics that in turn may vary between populations or even among individuals of the same population (Liedvogel & Lundberg 2014).

One of the major challenges to a better understanding of the molecular basis of migration in birds is to accurately characterize migratory behaviour in the wild. Because of this limitation, much of our currently limited understanding on the molecular basis of migration comes from laboratory studies using traditional genetic analyses on very few study organisms (e.g. the blackcap, Berthold & Querner 1981). These studies generally rely on proxies for migration such as migratory orientation and nocturnal restlessness in captivity (Newton 2008). Seminal work on the genetic control of migration has been based on partially migratory species, that is, when migratory and nonmigratory individuals coexist in the same location during the breeding season (Lack 1944; Newton 2008; Dingle 2014). One of the strengths of partial migration is the existence of the clearly defined dichotomy between migrant and resident phenotypes that share a common environment until the departure of migratory individuals in the autumn (Liedvogel *et al.* 2011). In large-scale cross-breeding and selection experiments of partially migratory blackcap populations (*Sylvia atricapilla*), Berthold & Querner (1981) showed that many behaviours associated with migration, such as the amount and timing of nocturnal restlessness, are under strong genetic control and can be rapidly modified within a few generations when exposed to strong artificial selection. Furthermore, high

genetic correlation among incidence and intensity of restlessness in blackcaps suggests that these components of migratory behaviour are influenced by shared genetic mechanisms (Pulido *et al.* 1996, 2001).

The demonstration of the existence of nocturnal restlessness also in birds that do not migrate in the wild has renewed questions whether such restlessness is truly a proxy of migration behaviour (Helm 2006; Helm & Gwinner 2006). Moreover, under laboratory conditions, a clear distinction between migrant and resident phenotypes is often blurred because most of the captive individuals exhibit nocturnal restlessness to some extent (Partecke & Gwinner 2007), making it difficult to objectively quantify the resident and migrant phenotypes on the basis of the intensity of nocturnal restlessness.

Recent technological advances and miniaturization of tracking devices enable the continuous monitoring of even small animals in the field throughout the year, including the identification of the time and place of migration (Kays *et al.* 2015). Such precise characterization of individual phenotypes also serves to enhance our understanding of environmental effects on migration and subsequently also genotype vs. environment interactions. Radio-tracking technology has significantly advanced our understanding of the ecology and evolution of migratory strategies (Mitchell *et al.* 2012; Brown & Taylor 2015). For instance, new insights about the causes and consequences of the existence of partial migration have been gained for skylarks (*Alauda arvensis*), where wintering strategies are more related to structural body size and immune functions than to the more classical control mechanisms such as genetic dimorphism or age- and sex-dependent conditions (Hegemann *et al.* 2015).

A second major challenge to better understand the genetic basis of migration stems from the necessary type and quality of genetic data itself. Studies using classical molecular genetic approaches such as mitochondrial DNA or microsatellites explored potential differences in the genetic structure of different populations displaying different migratory strategies (Bensch *et al.* 1999; Pérez-Tris *et al.* 2004). Genomewide AFLP scans were carried out in an effort to identify genetic differences in a migratory species exhibiting a migratory divide (Bensch *et al.* 2002, 2009). Most recently, candidate gene approaches were conducted with the aim to identify genes which are correlated with migratory traits (Mueller *et al.* 2011). Despite these studies, our knowledge about how many and which genes are involved in controlling migration is still limited (Liedvogel *et al.* 2011). Thanks to the recent advances in high-throughput sequencing technology and powerful bioinformatics analyses, it is now possible to accurately quantify gene expression patterns in both model and

nonmodel species, providing an unprecedented opportunity to investigate the genomic basis of migration (Wolf 2013; Kraus & Wink 2015). With the advent of high-throughput sequencing, some recent studies have begun to investigate sequence and gene expression differences between migratory and resident individuals targeting a better understanding of the genetic underpinnings of bird migration. These studies focused on molecular correlates of migratory strategies in captive white crowned sparrows (*Zonotrichia leucophrys*), dark-eyed juncos (*Junco hyemalis*), Swainson's thrush (*Catharus ustulatus*) or using differences in migratory directions of two subspecies of the willow warbler (*Phylloscopus trochilus*) and of the barn swallow (*Hirundo rustica*) (Boss *et al.* 2016; Delmore *et al.* 2015, 2016; Fudickar *et al.* 2016; Jones *et al.* 2008; Lundberg *et al.* 2013; von Ronn *et al.* 2016).

In this study, we took advantage of massive transcriptome sequencing and the latest radio-tracking technology to investigate the genomic basis of migration. To this end, we employed specially engineered radio transmitter (Bowlin *et al.* 2005) and geolocator technology to directly phenotype overwintering strategies (resident vs. migrant) in a free-living partially migratory European blackbird (*Turdus merula*) population (Fudickar *et al.* 2013). Like many other European bird species, blackbirds exhibit substantial variation in overwintering strategies across their breeding range, from year-round residents (in Southern Europe and Africa) to partially migratory (in Central Europe) and fully migratory (in Northern and Eastern Europe and Eurasia) (Glutz von Blotzheim 1988). To identify genes potentially associated with migrant and resident behaviours, we sequenced transcriptomes from blood cells of both migrant and resident individuals of the same breeding population during the fall migratory season.

## Materials and methods

### *Phenotyping overwintering strategies*

To determine overwintering strategies of individual European blackbirds (*Turdus merula*), we caught blackbirds in a forest area near Radolfzell Germany (N 47° 47', E 9° 2') during spring and summer preceding the following fall migration in 2010 and deployed backpacks consisting of light-level geolocators (Mk 10S ≤ 1.2 g; British Antarctic Survey, Cambridge, UK) and long-lasting radio transmitters (≤ 2.6 g delivered by Sparrow Systems, Fisher, IL, USA) that functioned more than a year on the studied birds (Fig. S1, Supporting information). We combined light-level geolocators with radio transmitters using heat shrink tubing (≤ 0.4 g) – see Fudickar *et al.* (2013) for details. Three to five stationary automated receivers (ARU) (Sparrow Systems)

were deployed at the study site to monitor the presence of individuals and eventual departure (Crofoot *et al.* 2008; Kays *et al.* 2011). Each automated receiver searched for 16 frequencies every 60 s. Automated receivers were connected to H antennas (ATS, Isanti, MN, USA), mounted 3–6 m high. In addition, birds were also tracked using manual radiotelemetry. Birds were manually tracked twice per week after capture until recapture the next spring except from 20 December to 10 January when birds were monitored from automated receivers exclusively. For ground tracking, we used a combination of either a hand-held three element Yagi antenna (AF Antronics, Inc., Urbana, IL, USA) and AR 8200 MKIII hand-held receiver (AOR U.S.A., Inc., Torrance, CA, USA) or a hand-held H antenna (Andreas Wagener Telemetry Systems, Köln, DE, USA) and a Yaesu VR 500 hand-held receiver (Vertex Standard USA, Cypress, CA, USA). If an individual could not be located by ground tracking, aerial searches encompassing a 20 km radius (minimum) of the study site were performed using a Cessna 172 airplane equipped with two H antennas, one per wing, and two Biotrack receivers, one per antenna (Lotek, Newmarket, ON, Can). Individuals were classified as migrants after at least two searches from the air without signal.

In September and October 2010 and 2011, we recaptured tagged individuals and collected blood samples for transcriptomic analysis shortly before migratory birds departed. Blood (100 µL) was obtained by puncturing a brachial wing vein with a 25-gauge needle. Blood was immediately frozen on dry ice upon collection and preserved at –80 °C until RNA extraction. We collected blood samples of six resident (three first year and two adult males and one adult female) and six migrant (one first year and three adult males and two adult females) individuals (see Table S1, Supporting information). There was a tendency for a 10-day earlier capture date for resident birds (median Julian date: 272; quartile 1: 264; quartile 3: 279; September 1 = 244) than for migrant birds (median Julian date: 282; quartile 1: 279; quartile 3: 288) (Wilcoxon rank-sum test:  $W = 30$ ,  $P = 0.06$ ). Median departure date for migrants was 296 (quartile 1: 288; quartile 3: 299) and migrants departed 11.5 days (quartile 1: 8.25 days; quartile 3: 14.75 days) after recapture and blood sampling date. In order to test for potential differences in gene expression according to the time lag relative to the departure date, we specifically selected blood samples of migrants to create two subgroups. One group ( $N = 3$ , two females and one male) departed on average 8 days (8, 8 and 9 days) after blood sampling, defined as 'short-term premigration' (SPM). The other group ( $N = 3$ , three males) departed on average 16 days (14, 15, 18 days) after blood sampling, hereafter defined as 'long-term premigration' (LPM) (see Fig. 1).

### RNA extraction, library preparation and next-generation sequencing

Total RNA from each sample was isolated with a customized protocol using both TRI Reagent BD (MP Biomedicals, Solon, USA) and an AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA). Briefly, 50  $\mu$ L of peripheral blood cells (pretreated with anticoagulants) was mixed with 375  $\mu$ L of TRI Reagent BD and 50  $\mu$ L of RNase-free ddH<sub>2</sub>O and a homogenized in a Lysing Matrix A Tube using a FastPrep-24 homogenizer (MP Biomedicals) for 40 s at 5.0 m/s. The upper aqueous phase was mixed with an equal volume of 70% ethanol and processed using a Qiagen AllPrep DNA/RNA Mini Kit according to manufacturer's instructions. RNA quality and quantity were assessed using, respectively, a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, USA) and a QUBIT v2.0 fluorometer (Life Technologies, Darmstadt, Germany), respectively.

For each sample, 500 ng of high-quality RNA (RIN value > 8.5) was used to construct 12 barcoded sequencing libraries with the Illumina TruSeq RNA sample preparation kit v2 (Low-Throughput protocol) according to the manufacturer's instructions (Illumina, San Diego, CA, USA). To increase the average library insert size, input RNA was chemical fragmented at 94 °C for 1 min. The 12 barcoded samples were equimolar-pooled, and clustered template cDNA was sequenced paired-end in three lanes of an Illumina HiSeq2000 platform at the TUCF Genomics Center of the University of Tufts (Boston, USA) with 209 cycles (101 cycles for each paired-read and seven cycles for the barcode sequences).

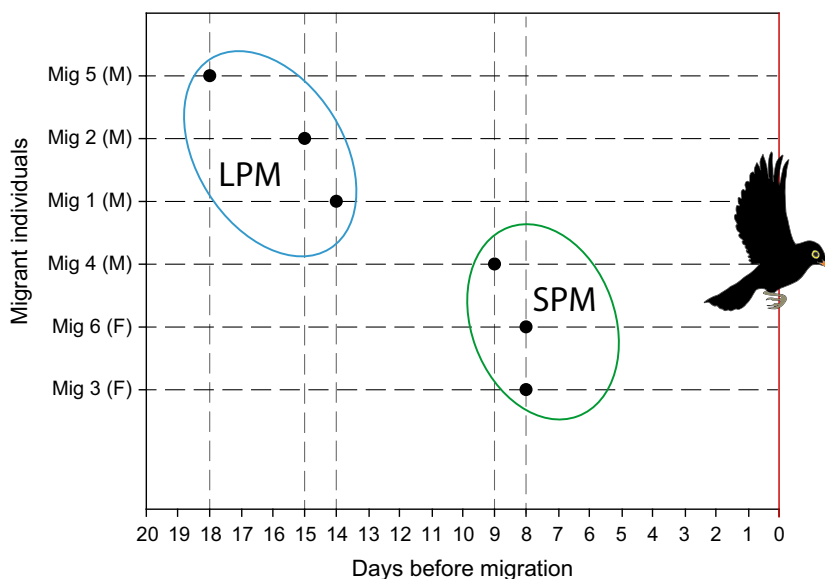
### Transcriptome assembly

We obtained 919 million (M) raw reads (59–101 M reads per sample), each 101 bp in length. Raw reads were quality-trimmed and potential contaminants removed (see Appendix S1, Supporting information for details). A final data set of 642 M reads (377 M paired reads, mean length 99 bp and 265 M single reads obtained by combining the merged overlapping paired reads and the 'orphan' broken paired reads, mean length 158 bp) ranging from 42 to 68 M reads per sample was used to build the transcriptome assembly (see Table S2, Supporting information for details).

The filtered reads of the twelve samples were combined and assembled *de novo* using TRINITY v20140717 (Grabherr *et al.* 2011) with the default k-mer size of 25 and a minimum contig length of 200 bp. To identify transcripts corresponding to coding genes, the obtained assembly was subjected to sequence similarity searches against a custom database containing the available bird proteins (chicken, duck, flycatcher, turkey and zebra finch; source: Ensembl release 77) and the well-annotated human and mouse protein data sets (Ensembl release 77). The BLASTx algorithm (Altschul *et al.* 1990) was used for similarity searches with a cut-off *e*-value of  $1e^{-6}$ . Protein-coding genes identified with the above reference databases were retained for downstream analyses.

### Differential expression analysis

The obtained final set of sequences (118 813 transcripts) was further clustered using the program CORSET v1.04



**Fig. 1** Graph showing the time elapsed between blood collection and departure for the six migratory individuals (sex specified in the *y*-axis label). According to this time, migrant birds were divided into two subgroups: short premigration (SPM) and long premigration (LPM). The experimental design is further shown in Fig. 2.



(Davidson & Oshlack 2014). Using the information from multiple-mapped reads, Corset allows the identification of a minimum number of clusters in the transcript data set (each cluster resembling a gene to which one or more assembled transcripts belong to) and assigns the raw read count from the 12 samples for each cluster. Read mapping for each sample was performed using BOWTIE v2.2.3 (Langmead & Salzberg 2012) with default parameters, but allowing infinite number of alignments (-a), as required by Corset. SAMTOOLS v1.2.1 (Li *et al.* 2009) was used to convert the Bowtie output SAM into BAM format. Corset was run setting the log-likelihood ratio (-D) to 10 000 and retaining only the transcripts with a minimum of 20 aligned reads. Differential expression (DE) analyses were performed on the clustered raw reads count table using the Bioconductor DESEQ2 v1.8.1 (Love *et al.* 2014) program in the R package v3.2.0 (R-Core-Team 2015). To validate the results obtained with DESEQ2, we used the BAYESIAN algorithm implemented in the R package EBSEQ v1.7.1 (Leng *et al.* 2013). A false discovery rate (FDR) was applied to correct the *P*-values generated by DESEQ2's exact test, whereas EBSEQ's posterior probability of 95% was used as threshold to consider a gene as DE (the EBSEQ algorithm does not need multiple test correction because it uses a single model to test all genes simultaneously).

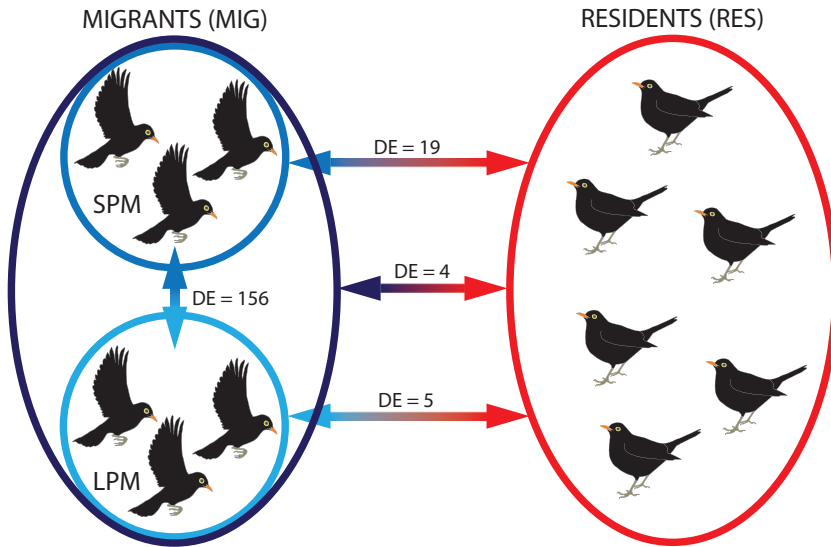
Differential expression analyses were run at two levels. First, we compared the migrant (MIG) and resident (RES) birds (six individuals per group). Second, we performed DE analyses using the two subgroups SPM and LPM with comparisons between those migrant birds (SPM vs. LPM) and each migrant subgroups against the resident birds (SPM vs. RES; LPM vs. RES) (see Fig. 2 for a schematic representation of the experimental design). Because of the lower number of biological replicates ( $N = 3$  in each migrant subgroup) resulting in reduced statistical power, we performed DE analyses only with DESEQ2 because of its greater ability to minimize type I errors caused by outliers offering consistent and robust performances even when groups with a few replicates are analysed (Love *et al.* 2014). Furthermore, to check the effect of the lower sample size and the sex bias in the SPM and LPM groups, we performed DE analyses with DESEQ2 between artificial groups after permuting migratory individuals in all possible combinations of two groups of three individuals. To independently examine the robustness of any identified clusters and to explore the grouping of individuals (Haas *et al.* 2013), a hierarchical cluster analysis was performed. We used a Spearman rank correlation-based matrix for individual clustering and Euclidean distances for gene expression clustering in R.

Functional annotation of the DE clusters/genes was carried out by BLAST2GO v2.8 (Conesa *et al.* 2005), setting

the NCBI's NR as reference database and using default parameters. For the annotation, we used a sequence data set that was obtained by extracting the longest transcript among those assigned to the same Corset cluster. Following this procedure, we obtained 18 219 transcripts ( $N_{50} = 4658$ ) that rendered 10 143 unique BLAST hits to the zebra finch protein data set. An enrichment analysis was performed to identify significantly overrepresented gene ontology (GO) terms in the DE genes (test sets) when compared to the whole assembly (baseline set), using a Fisher's exact test implemented in BLAST2GO.

### Sequence variation analyses

To assess the degree of genetic differentiation between migrant and resident birds, and to identify genomic regions linked to migration behaviour, we used a genome-scan approach on the inferred individual genotypes of the 12 samples. The Bayesian genetic variant detector FREEBAYES v1.0.2 (Garrison and Marth), run with default parameters, was used to infer the individuals' genotypes that were stored in a VCF file and subjected to quality-filtering (see Appendix S1, Supporting information for details). The final filtered VCF file was used to calculate, at each site, the  $F_{ST}$  values between the migrant and resident groups using a customized version of VCFTOOLS v0.1.15 (Danecek *et al.* 2011) that allows to report the partitioning of variance components reflecting differences within or among individuals and among the two overwintering strategies (migrant vs. resident). These values were used to calculate a multilocus  $F_{ST}$  estimate for the whole data set and for each gene independently as the 'ratio of means' (Weir & Cockerham 1984). Finally, to test whether the difference between the two overwintering strategies (migrant vs. resident) is significant, we used a permutation procedure (genomewide level) and a likelihood-ratio test (single-site level). For the genomewide test, we used ARLEQUIN v3.5.2.2 (Excoffier & Lischer 2010) to generate the null distribution of pairwise  $F_{ST}$  values under the hypothesis of no difference between migrant and resident birds by permuting individuals' genotypes between the strategies (10 000 permutations), where the *P*-value is the proportion of permutations leading to a  $F_{ST}$  value larger or equal to the observed one (*P*-value threshold = 0.05). To test the significance of the allele frequency differences at each site, we used a likelihood-ratio test implemented in the *pFst* script in the vcfliib package. Finally, to further investigate the possible presence of genetic clustering of migrant and resident blackbirds, we calculated the likely allelic admixture from the 12 reference individuals using the program ADMIXTURE v1.23 (Alexander *et al.* 2009). The VCF file was converted to PED format using PLINK v1.90 (Purcell *et al.*



**Fig. 2** Scheme of the experimental design showing the different layers in which the differential expression (DE) analyses were performed on the migrant and resident birds' data set (migrants are here represented by flying birds and residents by grounded ones). The number of DE genes in each pairwise comparison is reported. MIG, migrants; RES, residents; SPM, short premigration; LPM, long premigration.

2007), and the ADMIXTURE analysis was run with default parameters and allowing the number of genetic clusters (K) to range from 1 to 5.

To check whether the highly divergent genes (those falling in the upper 1% percentile distribution of the per-gene  $F_{ST}$ ) clustered in 'islands' of genomic differentiation or they were uniformly distributed across the genome, we used the zebra finch genome as reference to check their location (we used the BioMart tool in Ensembl to extract the mapping coordinates of the genes). Then, we applied a permutation procedure implemented in the R package *regioneR* based on the number of overlaps (see Appendix S1, Supporting information for details).

*Candidate genes for differential expression and sequence divergence*

To identify genes putatively linked to migration in birds, we compared the genes that were differentially expressed and highly divergent in our study against two different lists of genes that have been related to migratory behaviour in previous studies targeting both differential expression and sequence variation analyses.

In the first list, we included genes found to be DE in two recent studies: (i) 181 genes identified in blood tissue of two recently diverged migrant and resident dark-eyed Junco (*Junco hyemalis*) subspecies that live in seasonal sympatry during winter and early spring (Fudickar *et al.* 2016); (ii) 1639 genes (selected out of the 3045 microarray probes with reported orthology with zebra finch genes) that showed differential expression in two willow warbler subspecies (*Phylloscopus trochilus trochilus* and *P. t. acredula*) with divergent migratory phenotypes (Boss *et al.* 2016). In our study, we found a total of 177 DE genes in four pairwise comparisons

(MIG vs. RES, SMP vs. LPM, RES vs. SPM and RES vs. LPM – see Results section), and these were used as queries in a similarity search using BLASTx ( $e$ -value  $< 1e^{-6}$ ) against the list of candidate genes (extracted from the Ensembl zebra finch protein data set using the Ensembl IDs, when available). To ensure the correct homology of genes, a second BLASTx search of our 177 DE genes was performed against the whole zebra finch peptide data set to exclude the presence of better BLAST hits (lower  $e$ -value).

In the second list, we included 34 candidate genes that showed sequence divergence in previous studies (Muel-ler *et al.* 2011; Lundberg *et al.* 2013; Peterson *et al.* 2013; Ruegg *et al.* 2014) (Table S3, Supporting information) plus 100 candidate genes that have been found in outlier regions of divergence between coastal and inland sub-species groups of the Swainson's thrush (*Catharus ustulatus ustulatus* and *C. u. swainsoni*) differing, among other traits, in their migratory behaviour (see Table S2, Sup-porting information in Delmore *et al.* 2015). These genes were used as reference for BLASTx searches (as described above) using as queries our set of 65 blackbird genes with high  $F_{ST}$  values identified by a genome-scan approach (see Results section). Finally, in order to investigate whether genes that have previously been shown to differ at the sequence level between groups with differ-ent migration strategy also differ in their expression levels, we further searched for the 177 DE genes found in our study in the second list of candidate genes.

**Results**

*Reference blood transcriptome*

A final data set of 642 206 263 filtered sequences (both paired- and single-end reads – see Table S2, Supporting

information) obtained combining all the reads from the twelve samples were used to assemble *de novo* the blood reference blackbird transcriptome. The first raw assembly yielded a total of 344 389 transcripts ( $N_{50} = 4134$ ), a number that was reduced to 118 813 likely coding sequences (34.5 % of the total –  $N_{50} = 5172$ ; see Materials and Methods for details). The clustering procedure based on multimapped reads implemented in Corset identified 18 219 clusters (sharing a minimum of 20 aligned reads) that contain from one to 174 Trinity transcripts (Mean 6.29; SD 14.02). Considering all the inferred clusters, a comparable number of reads were assigned to each of the two groups (MIG: 197 990 078; RES: 173 653 114).

### Differential gene expression analysis

When applied to the whole data set, the differential expression analysis performed by DESeq2 identified four differentially expressed genes between resident and migrant individuals ( $P < 0.05$  after FDR correction) (RES vs. MIG). The differential expression of these four genes was confirmed by the BAYESIAN algorithm (EBSeq), which identified 13 additional DE genes (Table S4, Supporting information). Nevertheless, we preferred a conservative focus on only the four genes that were consistently estimated by both algorithms. Three of these four DE genes are overexpressed in the migrant group: a motilin receptor (*mlnr*), a DNA topoisomerase (*top2b*) and a sialyltransferase (*st6galnac2*); and one gene, a  $\text{tgf-}\beta$  receptor (*tgfbr1*), is overexpressed in the resident group (Table 1).

When the group of migratory individuals was split into two subgroups, short-term (SPM) and long-term (LPM) premigration, according to the time elapsed between blood collection and departure, a total of 156 DE genes were found. The permutation procedure performed by shuffling all migrant individuals (for the whole gene set) to form all the possible artificial groups of three individuals each rendered from 3 to 104 DE genes compared to the 156 DE genes when real temporal groups (SPM vs. LPM) were compared (see Fig. S2, Supporting information).

We then identified 19 DE genes in the analyses of SPM vs. resident birds (RES) and five DE genes in the analysis of LPM vs. RES (Fig. 2 and Table S4, Supporting information). The DE genes in the three comparisons showed little overlap (Table S4, Supporting information): two genes were DE in the RES vs. MIG and RES vs. SPM comparisons: motilin receptor (*mlnr*) and the sialyltransferase (*st6galnac2*); five DE genes were found in the RES vs. SPM and SPM vs. LPM comparisons: an ubiquitin–protein ligase (*herc4*), a DNA repair protein (*rad51d*), an histone cluster (*hist1H2a4 14*),

a translocator protein (*tspo2*) and ferrochelatase (*fech*). The ubiquitin–protein ligase (*herc4*) was DE between residents and both migrant subgroups (RES vs. SPM, RES vs. LPM, but not RES vs. MIG) as well as between SPM and LPM migrants (see Fig. 2 for a graphical representation of the experimental design and a summary of the DE genes found in each comparison). The 177 DE genes found in all four comparisons had no significant overrepresentation of functional categories when compared against the reference whole blood transcriptomic data set, as revealed by the enrichment analyses (Fisher's exact test;  $P > 0.05$ ).

To visualize the expression patterns of the DE genes observed between SPM and LPM migrants, we performed a hierarchical cluster analysis on the 156 DE genes (SPM vs. LPM). The hierarchical clustering was only performed on this set of 156 DE genes, an adequate number for this analysis compared to the other comparisons that rendered too few genes (DESeq2-normalized expression values for each of the DE genes are reported in Table S4, Supporting information). The hierarchical clustering analysis showed high level of coexpression and clearly separated SPM and LPM groups (Fig. 3).

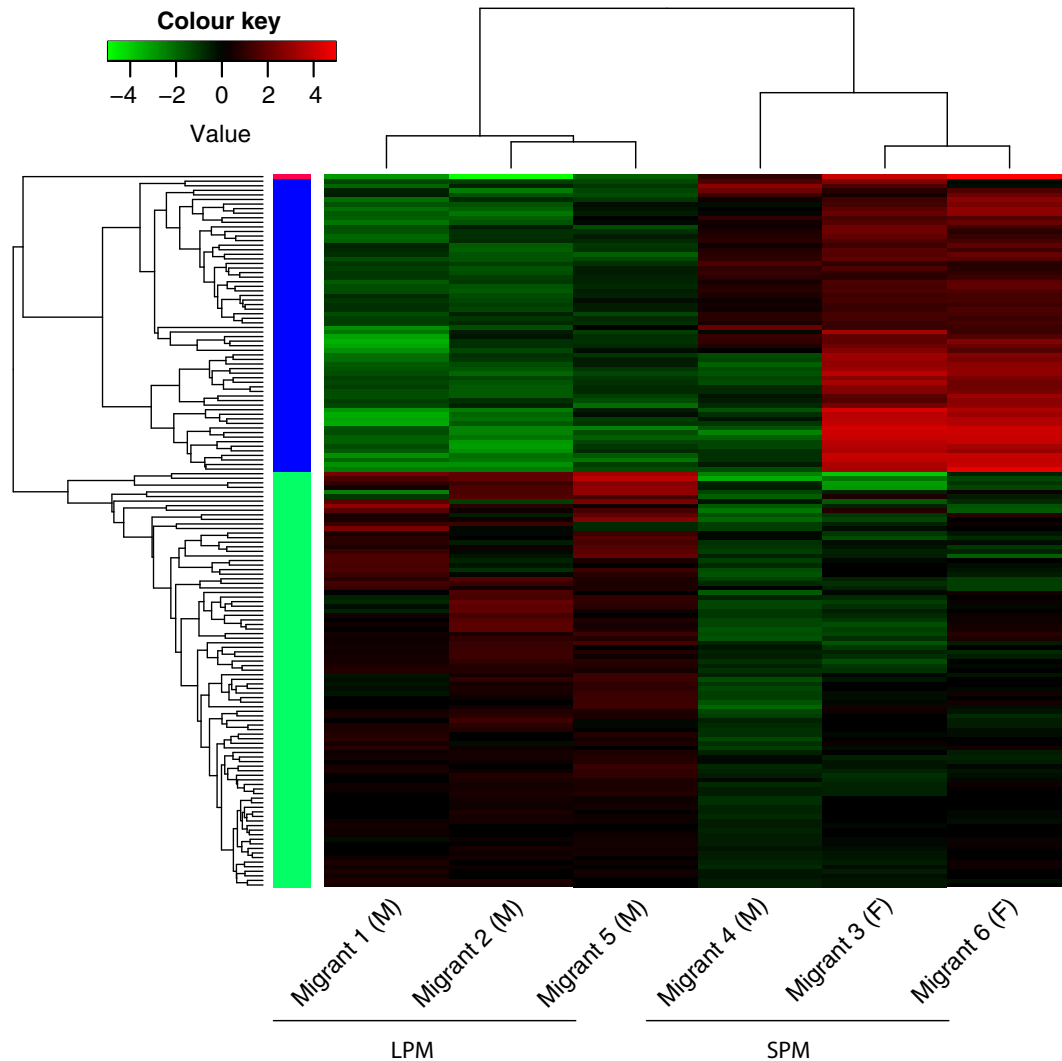
The similarity search using BLASTx found one of the genes that is DE between LPM and SPM also to be DE between two dark-eyed Junco (*Junco hyemalis*) subspecies that differ in their migratory behaviour (Fudickar *et al.* 2016): the ubiquitin-associated protein *uabp1*. Moreover, 18 genes were found DE in both our study and the survey conducted by Boss *et al.* (2016), in which two willow warbler subspecies with divergent migratory phenotypes were analysed. Notably, one of these matches was the motilin receptor (*mlnr*), DE gene in our MIG vs. RES comparison (see Table S4, Supporting information).

### Genetic divergence analysis

Using a genome-scan approach, we estimated  $F_{ST}$  values at 425 578 sites (after filtering reduced to 141 462 sites) that covered 6544 genes. The genomewide average  $F_{ST}$  value obtained by a multisites estimate was 0.041. The permutation test carried out at the genomewide level failed to reject the null hypothesis of no difference between the MIG and RES birds ( $P = 0.35$ ) and the likelihood-ratio test performed at each site on the allele frequencies of the two groups did not identify any significant difference (FDR correction applied, all adjusted  $P$ -values  $> 0.05$ ).  $F_{ST}$  values were then calculated for all the 6544 genes in which at least one polymorphic site was found. The  $F_{ST}$  values of these genes ranged between 0.000 and 0.576 (the distribution of the  $F_{ST}$  values per gene is shown in Fig. S3, Supporting information) where

**Table 1** Annotation of the four differentially expressed genes between resident and migrant blackbirds. Normalized expression values (DESeq2 normalization algorithm), fold change and FDR corrected *P*-values are shown

Transcript ID	Gene name	Sequence description	Migrants (NormExpress)	Residents (NormExpress)	log <sub>2</sub> FoldChange	<i>P</i> -value (FDR)
comp94876_c8_seq3	<i>mlnr</i>	motilin receptor	70.75	9.08	-1.5889	0.00189
comp101476_c2_seq2	<i>tgfbr1</i>	tgf-beta receptor type-1	73.58	710.50	1.5828	0.00189
comp91622_c0_seq1	<i>top2b</i>	dna topoisomerase 2-beta	2983.82	1967.36	-0.5745	0.02219
comp99860_c0_seq3	<i>st6galnac2</i>	alpha-n-acetylgalactosaminide alpha-2,6-sialyltransferase 2	169.31	23.05	-1.3757	0.02584

**Fig. 3** Heat map showing the clustering of the 156 genes found to be differentially expressed (DE) between two migrant subgroups differing in time elapsed between blood collection and migration: short premigration (SPM) and long premigration (LPM). The colours of the heat map reflect expression values normalized with the 'size factors' function implemented in the statistical package DESeq2.

the upper 1% percentile included 65 genes showing  $F_{ST} > 0.233$ , here defined as highly divergent genes (Table S5, Supporting information). The weak but non-significant genetic differentiation showed by the  $F_{ST}$

analysis was confirmed by the cross-validation error estimates of the ADMIXTURE analysis. ADMIXTURE identified  $K = 1$  as the best number of genetic clusters fitting the data set (Fig. S4, Supporting information). The



permutation test implemented in the *regionR* package suggested that the 65 highly divergent genes are not randomly distributed across the genome (51 overlaps,  $P = 0.04$ ; Table S6, Supporting Information), and the enrichment test showed that they are not enriched for any functional categories (Fisher's exact test  $P > 0.05$ ).

The 65 highly divergent genes resulted from our genome-scan did not find any match in the list of genes of other species that showed differences at the nucleotide level. However, three genes that we found DE in the comparison between SPM and LPM migratory individuals were also present in the list of 34 selected candidate genes (see Methods section) and identified by Delmore *et al.* (2015) in the outlier regions of differentiation of the Swainson's thrush subspecies: a periodic circadian clock gene (*per2*), a nuclear factor interleukin 3 regulated gene (*nfil3*) and an arylalkylamine *N*-acetyltransferase gene (*aanat*) (Table S3 and S4, Supporting information).

## Discussion

By linking latest radio-tracking technology with novel RNA sequencing (RNA-seq), we compared gene expression profiles of free-living migrant and resident individuals of a partially migratory European blackbird population. Resident and migrant European blackbirds differentially expressed genes in blood collected during the fall migratory season that are likely associated with migrant or resident processes such as hyperphagia resulting in increased fat deposition, hypertrophy of muscle and liver, changes in aerobic capacity (e.g. erythropoiesis), changes in metabolic and anabolic enzyme expression, moulting and DNA replication and transcription (Ramenofsky & Wingfield 2007; Guglielmo 2010; Hegemann *et al.* 2015). Moreover, genome sequence variation analyses show a nonsignificant genetic differentiation between migrant and resident blackbirds suggesting the absence of reproductive isolation between the two phenotypes. In agreement with the presence of a single panmictic population, we did not find any difference in morphology between migrants and residents either (Fudickar & Partecke 2012).

Notably, of the more than 10 000 genes that were expressed in the blood cells, only four (DESeq2) or 17 genes (EBSeq) were DE between migrant and resident blackbirds. However, when dividing the migrants into two subgroups differing in the time lag of on average 8 (SPM) vs. 16 days (LPM) until their departure, the number of DE genes changed markedly with 156 DE genes between LPM and SPM migrants, 19 between SPM and RES and five between LPM and RES (Fig. 2). It is conceivable that the higher number of DE genes in the comparison between SPM and LPM migrants has

been inflated by false positives due to low sample size of the two migrant subgroups. However, the permutation procedure that we used to investigate effects of low sample size and sex bias indicates the biological significance of the DE genes between SPM and LPM migrants. The higher number of DE genes obtained in the RES vs. SPM compared to the RES. vs. LPM analyses might indicate that the closer the departure date approaches the more genes linked to migratory processes are DE. The drastic temporal variation in DE genes over a week may emphasize the high plasticity of gene regulation over short time periods (Hodgins-Davis & Townsend 2009) and indicates that shortly before migration commences many physiological processes are either up- or downregulated. However, the confounding effect of sex and low sample size could have played a role also in these two latter comparisons.

In this study, we focussed on blood transcriptome, because this noninvasive method allowed us to phenotype the individuals from several days until several months after blood sampling. Even though we could not investigate gene expression in other relevant tissues (e.g. brain or muscle), blood has been recently shown to contain important signals of migration-related gene expression (Liedvogel *et al.* 2011; Fudickar *et al.* 2016). Following guidelines and directions from statistical (Love *et al.* 2014) and empirical studies (Dunham *et al.* 2012; Sims *et al.* 2014), we carefully considered the trade-off between replicated design and sequencing depth. To this end, we sequenced the poly(A)-selected transcriptome obtaining approximately 920 million, averaging 77 million sequences per individual. This substantial sequencing depth was necessary to recover a comprehensive set of genes, even considering that haemoglobin transcripts are dominant in bird blood tissue and these could hinder the detection of lowly expressed genes. Therefore, the high-sequencing depth together with a statistically sound replication design (six migrant and six resident individuals) increased the confidence to the results obtained. Nevertheless, our migrant and resident groups consisted of both sexes and first year and older birds, and either factor could potentially affect gene expression profiles which are not related to migration. Blood sampling of resident and migrant birds inadvertently differed on average by 1 week. Thus, there is the possibility that the DE genes between resident and migrant birds could rather reflect seasonal differences than migration specific differences. To resolve this question, an essential follow-up study would be to sample residents and migrants during the entire migratory season which lasts more than 7 weeks (Fudickar *et al.* 2013) and test whether differences in gene expression between migratory and resident individuals change or persist over this season.

### Differential gene expression

Two of the four DE genes of the MIG vs. RES level of analysis, the  $\text{tgf-}\beta$  receptor (*tgfbr1*) and sialyltransferase (*st6galnac2*), have been recently identified in feather moult in chicken (Ng *et al.* 2015) suggesting differences in the timing of moulting in migrant and resident individuals during the premigratory phase (de la Hera *et al.* 2010). The sialyltransferase catalyses the incorporation of sialic acid into glycoconjugates and has several biological functions including oligosaccharide metabolism or cell adhesion for the recruitment of leucocytes (Samyn-Petit *et al.* 2000). Both sialyltransferase and a further gene, the motilin receptor (*mlnr*), are both significantly upregulated in MIG and LPM migrants as compared to RES. Motilin is an intestinal peptide that stimulates smooth muscle contraction in the intestine, and the upregulation of its receptor is likely associated with hyperphagia in migrants including increased fat uptake or glucose levels in blood (Talley *et al.* 2001). The fourth DE gene between MIG and RES is the DNA topoisomerase (*top2b*), which has the function of relieving torsional stress of DNA during replication and transcription. Its upregulation in migrants is likely associated with a higher activity of the DNA machinery. Interestingly, one out of four DE genes identified by our study (*mlnr*) was also found as DE in two willow warbler subspecies with divergent migratory phenotypes (Boss *et al.* 2016).

When comparing migrants by their time to departure (SPM vs. LPM), we observed upregulation in migrants that were closer to the migration date of several genes associated with the circadian clock (*per2*, *nfil3*, *gfpt1*), metabolic genes for cholesterol transport and steroid and bile acid synthesis (*tspo2*), lipid metabolism (*smpd2*, *st8sia1*, long-chain fatty acid ligases) or haem synthesis (*fech*), cell trafficking and protein degradation (*herc4*), protection from redox stress (thioredoxin, peroxiredoxin, *nudt1*, *osgin1* and several receptors including *zd-hhc21*) and proteins associated with DNA replication and repair (*hist1H2a4 14*, *rad51d*, *smc2*, *ercc8*, *herc4*). Upregulation of certain genes in SPM (*tpx2* and *nup43*) suggests increased cell division. In contrast, we found evidence for decreased transcription in SPM (upregulation of the *tob1* transcription repressor and downregulation of the RNA polymerase *rpc82* and *nup43*).

The candidate gene *adacyp1*, which has previously been associated as a gene controlling the expression of migratory behaviour in blackcaps (Mueller *et al.* 2011), was not expressed in the blood of neither resident nor migrant blackbirds. We found a single gene (an ubiquitin-associated protein) overlapping between our DE genes and those identified by Fudickar *et al.* (2016) when comparing blood from migrant and resident

subspecies of dark-eyed juncos. Differences in the timing and design of the studies are likely to have contributed to differences in gene expression. Indeed, in the current study, we collected blood from free-living migrants in the autumn, after the breeding season while Fudickar *et al.* (2016) held birds in captivity during the winter and sampled blood in captivity during spring migration, just prior to reproduction. Although autumn and spring migrations are superficially considered to be similar behaviours, the seasonal behavioural and physiological transitions are likely very different, which in turn may explain the lack of concordance in differentially expressed gene sets between the two studies. On the other hand, a larger number of DE genes overlapped between our study and that of Boss *et al.* (2016). One of these was the motilin receptor (*mlnr*), which was found to be DE in MIG vs. RES, the most relevant level of our analysis. The fact that these genes have been linked to migratory behaviour in two different systems (partially migratory blackbird populations vs. willow warbler subspecies differing in their wintering migratory direction) provides further evidence that these common genes might have a key role in the onset of migration in birds.

### Genetic divergence

The genomewide  $F_{ST}$  analysis found a nonsignificant genetic differentiation between the MIG and RES individuals ( $F_{ST}$  value: 0.041), and the admixture analysis was unable to identify clear genetic clusters that could be associated with migrant or resident phenotypes. However, the per-gene  $F_{ST}$  analysis does not allow us to completely rule out the presence of weak assortative mating within migrant and resident groups. In a background of very low genetic differentiation, several genes were found to be highly divergent between migrant and resident birds, suggesting that localized genomic regions of divergence might be the effect of an incipient genetic differentiation (Feder *et al.* 2012; Nosil & Feder 2012; Seehausen *et al.* 2014; Fruciano *et al.* 2016; Kautt *et al.* 2016). Remarkably, our permutation analysis found these outlier regions of differentiation to be clustered, thus supporting a 'genomic islands' model of genetic divergence. Nonetheless, it should be noted that we investigated the gene clustering using as reference the zebra finch genome, thus assuming conserved chromosomal organization and gene order between blackbird and zebra finch. Our assumption is supported by the conserved synteny in avian genomes (Ellegren 2010), but our results should be confirmed using the blackbird genome when it will become available. On the whole, our results and conclusions on the sequence divergence are all based on limited sampling of

individuals and the use of transcriptomic data which could have led to an overestimation of the fixed allelic variants.

## Conclusions

By integrating state-of-the-art radiotelemetry and high-throughput RNA sequencing, we investigated what genes may be involved in the expression of migrant versus resident behaviour in a free-living partially migratory population. Our analyses suggest that especially prior to the departure date many genes rapidly change their expression in migrants. These potential rapid changes highlight the need of perfect timing when it comes to the sampling scheme for the transcriptomic analyses. By identifying a set of possible candidate genes that might play a crucial role in determining overwintering strategies (resident vs. migrant), we provide testable hypotheses for future studies exploring the transcriptional basis in other migratory bird species. This may allow to gain new insights into the molecular mechanisms controlling this fascinating behaviour.

Future studies would benefit from comparing not only individuals of one partially migratory population, but also including both migratory and resident populations from different geographical locations, as well as from more extensive sampling of individuals per population genotyped at a dense set of markers (e.g. using RAD-Seq or whole-genome resequencing). In this regard, the sequencing of a high-quality reference genome would definitively help the study of the genetics of migration. Such improvements would allow to overcome the limitation of our transcriptomic study and further testing the generality of our conclusions on a larger scale.

## Acknowledgments

We thank D. Monné for laboratory assistance, C. Fruciano for statistical support and E. Malecore for graphical support. The University of Konstanz and the GeCKo (University of Konstanz Genomic Center) are thanked for their support. PF was supported by a Deutsche Forschungsgemeinschaft Research Grant (FR 3399/1-1). It was supported by postdoctoral fellowships from the Alexander von Humboldt Foundation (1150725) and the European Molecular Biology Organization (EMBO ALTF 440-2013). The study was supported by the Max-Planck-Society.

## References

- Akesson S, Hedenstrom A (2007) How migrants get there: migratory performance and orientation. *BioScience*, **57**, 123–133.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Bensch S, Andersson T, Akesson S (1999) Morphological and molecular variation across a migratory divide in willow warblers, *Phylloscopus trochilus*. *Evolution*, **53**, 1925–1935.
- Bensch S, Akesson S, Irwin D (2002) The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology*, **11**, 2359–2366.
- Bensch S, Grahm M, Müller N, Gay L, Akesson S (2009) Genetic, morphological and feather isotope variation of migratory willow warblers show gradual divergence in a ring. *Molecular Ecology*, **18**, 3087–3096.
- Berthold P (1991) *Orientation in Birds*. Birkhäuser, Basel.
- Berthold P, Querner U (1981) Genetic basis of migratory behaviour in European warblers. *Science*, **212**, 77–79.
- Boss J, Liedvogel M, Lundberg M *et al.* (2016) Gene expression in the brain of a migratory songbird during breeding and migration. *Movement Ecology*, **4**, 1–11.
- Bowlin MS, Cochran WW, Wikelski MC (2005) Biotelemetry of New World thrushes during migration: physiology, energetics and orientation in the wild. *Integrative and Comparative Biology*, **45**, 295–304.
- Brower LP (1996) Monarch butterfly orientation: missing pieces of a magnificent puzzle. *Journal of Experimental Biology*, **199**, 93–103.
- Brown JM, Taylor PD (2015) Adult and hatch-year blackpoll warblers exhibit radically different regional-scale movements during post-fledging dispersal. *Biology Letters*, **11**, 20150593.
- Conesa A, Gotz S, Garcia-Gomez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
- Crofoot MC, Gilby IC, Wikelski MC, Kays RW (2008) Interaction location outweighs the competitive advantage of numerical superiority in *Cebus capucinus* intergroup contests. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 577–581.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Davidson NM, Oshlack A (2014) Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology*, **15**, 410.
- Delmore KE, Hubner S, Kane NC *et al.* (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, **24**, 1873–1888.
- Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE (2016) The genetics of seasonal migration and plumage color. *Current Biology*, **26**, 2167–2173.
- Dingle H (2006) Animal migration: is there a common migratory syndrome? *Journal of Ornithology*, **147**, 212–220.
- Dingle H (2014) *Migration: The Biology of Life on the Move*, 2nd edn. Oxford University Press, New York, New York.
- Dunham I, Kundaje A, Aldred SF *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*, **25**, 283–291.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses



- under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.
- Fruciano C, Franchini P, Kovakova V *et al.* (2016) Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nature Communications*, **7**, 12736.
- Fudickar AM, Partecke J (2012) The flight apparatus of migratory and sedentary individuals of a partially migratory songbird species. *PLoS ONE*, **7**, e51920.
- Fudickar AM, Schmidt A, Hau M, Quetting M, Partecke J (2013) Female-biased obligate strategies in a partially migratory population. *Journal of Animal Ecology*, **82**, 863–871.
- Fudickar AM, Peterson MP, Greives TJ *et al.* (2016) Differential gene expression in seasonal sympatry: mechanisms involved in diverging life histories. *Biology Letters*, **12**, 20160069.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. Preprint at arXiv:1207.3907v2 [q-bio.GN].
- Glutz von Blotzheim UN (1988) Handbuch der Vögel Mitteleuropas. Band 11/1. Aula, Wiesbaden.
- Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.
- Guglielmo C (2010) Move that fatty acid: fuel selection and transport in migratory birds and bats. *Integrative and Comparative Biology*, **50**, 336–345.
- Haas BJ, Papanicolaou A, Yassour M *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494–1512.
- Hegemann A, Marra PP, Tieleman BI (2015) Causes and consequences of partial migration in a passerine bird. *The American Naturalist*, **186**, 531–546.
- Helm B (2006) Zugunruhe of migratory and non-migratory birds in a circannual context. *Journal of Avian Biology*, **37**, 533–540.
- Helm B, Gwinner E (2006) Migratory restlessness in an equatorial nonmigratory bird. *PLoS Biology*, **4**, e110.
- de la Hera I, Hedenstrom A, Perez-Tris J, Telleria JL (2010) Variation in the mechanical properties of flight feathers of the blackcap *Sylvia atricapilla* in relation to migration. *Journal of Avian Biology*, **41**, 342–347.
- Hodgins-Davis A, Townsend JP (2009) Evolving gene expression: from G to E to G × E. *Trends in Ecology & Evolution*, **24**, 649–658.
- Jones S, Pfister-Genskow M, Cirelli C, Benca RM (2008) Changes in brain gene expression during migration in the white-crowned sparrow. *Brain Research Bulletin*, **76**, 536–544.
- Kautt AF, Machado-Schiaffino G, Meyer A (2016) Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genetics*, **12**, e1006157.
- Kays R, Tilak S, Crofoot M *et al.* (2011) Tracking animal location and activity with an automated radio telemetry system in a tropical rainforest. *Computer Journal*, **54**, 1931–1948.
- Kays R, Crofoot MC, Jetz W, Wikelski M (2015) Terrestrial animal tracking as an eye on life and planet. *Science*, **348**, aaa2478.
- Kolzsch A, Bauer S, de Boer R *et al.* (2015) Forecasting spring from afar? Timing of migration and predictability of phenology along different migration routes of an avian herbivore. *Journal of Animal Ecology*, **84**, 272–283.
- Kraus RHS, Wink M (2015) Avian genomics: fledging into the wild!. *Journal of Ornithology*, **156**, 851–865.
- Lack D (1944) The problem of partial migration. *British Birds*, **37**(122–130), 143–150.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.
- Leng N, Dawson JA, Thomson JA *et al.* (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, **29**, 1035–1043.
- Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Liedvogel M, Lundberg M (2014) The genetics of migration. In: *Animal Movement Across Scales* (eds Hansson L, Akesson S), pp. 219–231. Oxford University Press, New York, New York, USA.
- Liedvogel M, Akesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution*, **26**, 561–569.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
- Lundberg M, Boss J, Canbäck B *et al.* (2013) Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler *Phylloscopus trochilus*. *BMC Genomics*, **14**, 330.
- Mitchell GW, Wheelwright NT, Guglielmo CG, Norris DR (2012) Short- and long-term costs of reproduction in a migratory songbird. *Ibis*, **154**, 325–337.
- Mueller J, Pulido F, Kempenaers B (2011) Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society of London Series B*, **278**, 2848–2856.
- Mueller T, O'Hara RB, Converse SJ, Urbanek RP, Fagan WF (2013) Social learning of migratory performance. *Science*, **341**, 999–1002.
- Newton I (2008) *The Migration Ecology of Birds*. Elsevier-Academic Press, Amsterdam, London.
- Ng CS, Chen CK, Fan WL *et al.* (2015) Transcriptomic analyses of regenerating adult feathers in chicken. *Bmc Genomics*, **16**, 756.
- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences Introduction. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **367**, 332–342.
- Partecke J, Gwinner E (2007) Increased sedentariness in European Blackbirds following urbanization: a consequence of local adaptation? *Ecology*, **88**, 882–890.
- Pérez-Tris J, Bensch S, Carbonell R, Helbig A, Telleria J (2004) Historical diversification of migration patterns in a passerine bird. *Evolution*, **59**, 1819–1832.
- Peterson M, Abolins-Abols M, Atwell J *et al.* (2013) Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *1000Research*, **2**, 115.
- Pulido F, Berthold P, van Noordwijk AJ (1996) Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 14642–14647.

- Pulido F, Berthold P, Mohr G, Querner U (2001) Heritability of the timing of autumn migration in a natural bird population. *Proceedings of the Royal Society of London Series B*, **268**, 953–959.
- Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.
- Ramenofsky M, Wingfield JC (2007) Regulation of migration. *BioScience*, **57**, 135–143.
- R-Core-Team (2015) *A Language and Environment for Statistical Computing*. Vienna, Austria, R-Core-Team. URL <http://www.R-project.org/>.
- Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, **23**, 4757–4769.
- von Rönn JAC, Shafer ABA, Wolf JBW (2016) Disruptive selection without genome-wide evolution across a migratory divide. *Molecular Ecology*, **25**, 2529–2541.
- Samyn-Petit B, Krzewinski-Recchi MA, Steelant WFA, Delanoy P, Harduin-Lepers A (2000) Molecular cloning and functional expression of human ST6GalNAc II. Molecular expression in various human cultured cells. *Biochimica Et Biophysica Acta-General Subjects*, **1474**, 201–211.
- Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.
- Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, **15**, 121–132.
- Talley NJ, Verlinden M, Geenen DJ *et al.* (2001) Effects of a motilin receptor agonist (ABT-229) on upper gastrointestinal symptoms in type 1 diabetes mellitus: a randomised, double blind, placebo controlled trial. *Gut*, **49**, 395–401.
- Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.
- Willemoes M, Blas J, Wikelski M, Thorup K (2015) Flexible navigation response in common cuckoos *Cuculus canorus* displaced experimentally during migration. *Scientific Reports*, **5**, 16402.
- Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources*, **13**, 559–572.

---

P.F. performed laboratory work, conducted molecular data analyses and wrote the manuscript. II carried out the molecular data analyses and wrote the manuscript. A.F., M.W., A.S., A.M. and J.P. conceived the study. A.F. and A.S. collected ecological data and J.P. wrote the manuscript. P.F., A.M., M.W. and J.P. participated in the experimental design. All the authors edited the manuscript.

---

## Data accessibility

Raw Illumina sequences were deposited into the NCBI's Sequence Read Archive (SRA) database with Accession no. SRA495362. The final blood transcriptome of the blackbird and the VCF file that stores the genotypes of the 12 individuals was submitted to Dryad, DOI: doi:10.5061/dryad.tc722.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Radio-tagged male European blackbird (*Turdus merula*).

**Fig. S2** Bar plot showing the number of differential expressed (DE) genes found in each round of the 10 DE analyses carried out between groups of three individuals each.

**Fig. S3** Distribution of the  $F_{ST}$  values of the 6544 genes in which at least one polymorphic site was found.

**Fig. S4** Admixture plot showing the estimated ancestry of the 12 blackbird specimens used in this study allowing the number of genetic clusters (K) to range from 1 to 5.

**Table S1** Information of the 12 blackbirds used for this study. Individual code, sex, capture and departure dates, and age class are shown.

**Table S2** Summary statistics of the sequencing. Reads for each individual before and after quality control are shown. See Material and Methods for the filtering criteria applied.

**Table S3** Annotation of 34 candidate genes that have been related to migratory behaviour in sequence variation studies based on a literature review.

**Table S4** (a, b, c, d) Annotation of the differential expressed genes found in the four pairwise comparisons (the genes of each comparison are shown in different excel sheets).

**Table S5** Annotation of the 65 highly divergent genes (falling in the upper 1% percentile distribution of the  $F_{ST}$  values).

**Table S6** Location of the 65 highly divergent genes (falling in the upper 1% percentile distribution of the  $F_{ST}$  values) in the zebra finch genome.

**Appendix S1** Detailed description of Methods.