De novo annotation of the transcriptome of the Northern Wheatear (*Oenanthe oenanthe*)

Roberto Carlos Frias-Soler[1,2], Lilian Villarín Pildaín[1], Agnes Hotz-Wagenblatt[3], Jonas Kolibius[1], Franz Bairlein[2] and Michael Wink[1]

[1] Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Heidelberg, Baden Württemberg, Germany

[2] Institute of Avian Research, Wilhelmshaven, Germany

[3] Bioinformatics Group, Core Facility Genomics and Proteomics, German Cancer Research Center, Heidelberg University, Heidelberg, Baden Württemberg, Germany

Corresponding authors
Roberto Carlos Frias-Soler, roberto.friassoler@ifv-vogelwarte.de
Michael Wink, wink@uni-heidelberg.de, wink@uni-hd.de

ABSTRACT

We have sequenced a partial transcriptome of the Northern Wheatear (*Oenanthe oenanthe*), a species with one of the longest migrations on Earth. The transcriptome was constructed *de novo* using RNA-Seq sequence data from the pooled mRNA of six different tissues: brain, muscle, intestine, liver, adipose tissue and skin. The samples came from nine captive-bred wheatears collected at three different stages of the endogenous autumn migratory period: (1) lean birds prior the onset of migration, (2) during the fattening stage and (3) individuals at their migratory body mass plateau, when they have almost doubled their lean body mass. The sample structure used to build up the transcriptome of the Northern Wheatears concerning tissue composition and time guarantees the future survey of the regulatory genes involved in the development of the migratory phenotype. Through the pre-migratory period, birds accomplish outstanding physical and behavioural changes that involve all organ systems. Nevertheless, the molecular mechanisms through which birds synchronize and control hyperphagia, fattening, restlessness increase, immunity boosting and tuning the muscles for such endurance flight are still largely unknown. The use of RNA-Seq has emerged as a powerful tool to analyse complex traits on a broad scale, and we believe it can help to characterize the migratory phenotype of wheatears at an unprecedented level. The primary challenge to conduct quantitative transcriptomic studies in non-model species is the availability of a reference transcriptome, which we have constructed and described in this paper. The cDNA was sequenced by pyrosequencing using the Genome Sequencer Roche GS FLX System; with single paired-end reads of about 400 bp. We estimate the total number of genes at 15,640, of which 67% could be annotated using Turkey and Zebra Finch genomes, or protein sequence information from SwissProt and NCBI databases. With our study, we have made a first step towards understanding the migratory phenotype regarding gene expression of a species that has become a model to study birds long-distance migrations.

INTRODUCTION

Each year billions of birds move twice a year between nesting areas and non-breeding regions. Some birds even cover distances of up to 15,000 km one way each season (*Bairlein et al., 2012*). To cope with the vast distances between breeding and non-breeding areas, birds typically engage in long endurance flights, which impose extraordinary morphological, physiological and behavioural challenges for the animals (*Bicudo et al., 2010*; *Bairlein, Eikenaar & Schmaljohann, 2015*).

The development of the "migratory phenotype" involves almost all organ systems (*Piersma et al., 2005*). Increased migratory restlessness, hyperphagia, fattening, the switch in their circadian rhythm and the ability to navigate over long distances are common characteristics of the migratory birds (*Bairlein & Gwinner, 1994*; *Berthold, 1996*; *Gwinner, 1996*). Such a complex phenomenon needs a holistic approach for its comprehension; which became possible only recently, with the availability of Next Generation DNA Sequencing techniques (hereafter NGS). A small number of genes which appear to be associated with migration have already been identified by transcriptome and genome analyses of several migratory birds; e.g., Willow Warbler *Phylloscopus trochilus*—a study that showed genes associated with neuronal signalling and calcium flux were regulated (*Boss et al., 2016*). In the same species, *Lundberg et al. (2017)* found genetic variants in the chromosomes 1 and 5 that match perfectly with the geographic distribution of populations with different migratory directions. In these genomic regions can be found cluster of genes related to fatty acid synthesis, a key metabolic process limiting of physical performance. Another research about subspecies of Swainson's Thrush, *Catharus ustulatus* with different migratory behaviour, found gene islands probably related to speciation (*Ruegg et al., 2014*). Nevertheless, other works have failed in the search for specific genetic signals related to migration using candidate genes among multiple species (*Lugo Ramos, Delmore & Liedvogel, 2017*). It seems like the study of the adaptations for long-distance migrations needs more careful experimental design based in functional studies.

Even though the phenomenology of avian long-distance migrations is well-described, still, deep molecular studies on that respect are missing. Some of these adaptations might have medical relevance (*Totzke, Hübinger & Bairlein, 1998*). Many "normal" phenotypic changes in birds like autumnal and spring fattening or glucose metabolism, would be life-threatening for mammals. The genetic characterization of these complex traits is necessary for a reliable interspecific comparison of metabolic and physiological processes. To disclose the pathways beneath the birds' migratory phenotype may guide researches to better understand and treat human diseases like obesity, diabetes and metabolic syndrome.

The Northern Wheatear as a model to study migration

The Northern Wheatear (*Oenanthe oenanthe*, Linnæus, 1758) breeds across most of the Holarctic with all the populations wintering in sub-Saharan Africa (*Del Hoyo, Elliott & Sargatal, 2005*; *Bairlein et al., 2012*).

O. oenanthe is a solitary, nocturnal long-distance songbird migrant which develops the migratory phenotype with consistency in captivity since it has mostly genetic basis (*Maggini & Bairlein, 2010*; *Maggini, Bulte & Bairlein, 2017*). At the same time, the birds

can be studied relatively easy in the field, which makes the wheatear an excellent model to study the endogenous mechanism that commands their preparation for migration (*Bairlein et al., 2013*; *Bairlein et al., 2015*).

The assembly and annotation of a reference transcriptome for wheatear is the primary necessary step in identifying differentially expressed genes related to the migration phenotype in different organs. The goal of this article is to report the first transcriptome of the Northern Wheatear, based on RNA-Seq data. It was annotated *de novo*, using the Turkey and Zebra Finch genomes and Uniprot and NCBI databases. As a source of mRNA, we took samples from captive-bred birds in different stages of their endogenous migration cycle: (1) before the onset of migratory fuelling, (2) during fuelling, and (3) when fat stores reached their maximum, i.e., when they have nearly doubled their pre-migratory body mass (*Maggini & Bairlein, 2010*). The partial transcriptome was constructed by pooling mRNA from the brain, muscle, intestine, liver, adipose tissue and skin from three birds at each stage. The sample was sequenced by pyrosequencing using the Genome Sequencer Roche GS FLX System.

We believe that our data will be a useful tool in the way to dissect the molecular basis of one of the most intriguing phenomenon in nature, birds' long-distance migration.

MATERIALS AND METHODS

Sample collection

We sampled four female and five male F1-hybrid offspring of parental populations of *Oenanthe oenanthe* originating from Norway and Morocco; bred in outdoor aviaries in the natural breeding period on the grounds of the Institute of Avian Research (Wilhelmshaven, Germany) and taken indoors as soon as they were independent of their parents. From mid-August onwards until euthanization in early December, the period when fuelling and migratory activity occur (*Maggini & Bairlein, 2010*), the birds were kept in individual cages (size $50 \times 40 \times 40$ cm) under controlled conditions at a light-dark cycle of 12 h light:12 h dark, 20 ± 1 °C room temperature. The birds were fed with a standardized insectivore diet (*Bairlein, 1986*) supplemented with a few mealworms and *ad libitum* fresh water. The body mass was recorded twice a week to the nearest 0.1 g, early in the morning before the birds were offered food. The birds were euthanized following *Guglielmo, Haunerland & Williams (1998)* in middle August, middle October and early December. Before exsanguination at the jugular vein, the animals were anesthetized with an intramuscular injection of ketamine hydrochloride (40 μl: 25 g). All procedures were in accordance with the German Animal Welfare Act and approved by the Zweckverband Veterinäramt JadeWeser; the number of the permitting authority is 42508-Te. The samples are stored at the Institute of Pharmacy and Molecular Biotechnology (IPMB), Heidelberg University, collection numbers IPMB 66613-66666.

RNA purification

Organs were dissected from freshly killed birds and stored in liquid nitrogen until RNA isolation. For RNA purification, we used the "GeneMATRIX Universal RNA Purification Kit" (Roboklon GmbH, Berlin, Germany), and for a final "clean-up", the "RNeasy

Fibrous Tissue Mini Kit" (Qiagen GmbH, Hilden, Germany). We followed the protocols recommended by the suppliers; moreover, we did not load more than 12 mg of tissue per column, DNAse treatments lasted 1 h at 37 °C, and all washing steps were repeated three times. The general recommendations for RNA isolation by *Sambrook & Russell (2012)* were considered.

To check the absence of genomic DNA, a PCR targeting the mitochondrial gene cytochrome b was carried out. The amplifications were undertaken in a 30 µl reaction volume containing 1.5 mM MgCl$_2$, 100 µM dNTPs, one unit of Taq DNA polymerase (Pharmacia Biotech, Munich, Germany), 4 µL from the RNA isolation and 5 pmol of each oligonucleotide (*Wink et al., 2002*), mta1: CCCCCTACCAACATCTCAGCATGAT-GAAACTTG and mtfr: CTAAGAAGGGTGGAGTCTTCAGTTTTTGGTTTACAAGAC. The PCR program consisted of three steps: (1) denaturation at 94 °C for 5 min; (2) 35 cycles, including denaturation at 94 °C for 25 s, annealing at 50 °C for 2 min and extension at 72 °C for 2 min, followed by (3) final extension at 72 °C for 10 min. PCR products were isolated and visualized by agarose gel electrophoresis.

The RNA concentration, cleanliness and integrity were evaluated using the standard spectrophotometric measurement and agarose gel electrophoresis.

Transcriptome sequencing and annotation

With the aim of sequencing a partial transcriptome of the Northern Wheatear, 54 RNA isolations [~713 ng each] were pooled. The samples were derived from nine wheatears, three per time-point (middle August, October, and December 2014) and six different organs each: brain, liver, intestine, muscle, skin and adipose tissue.

The cDNA library preparation, sequencing, and transcriptome assembly were conducted by GATC-Biotech AG (Konstanz, Germany) using 38.5 µg [350.4 ng/µl] of total high-quality RNA (28s/18s ratio ≥ 1.7; RIN ≥ 8; A260/A280 ≥ 1.8). From the total RNA sample, poly(A) + mRNA was isolated, and the first-strand cDNA synthesis was primed with random hexamer primers (N6). Then the "454 adapters" A and B were ligated to the 5′ and 3′ ends of the cDNA. The resulting cDNA (N0-cDNA) was finally amplified with PCR (number of cycles = 11) using a proof-reading polymerase.

Normalization was carried out by one cycle of denaturation and reassociation of the N0-cDNA, resulting in N1-cDNA. Reassociated ds-cDNA was separated from the remaining ss-cDNA (normalized cDNA) by passing the mixture over a hydroxylapatite column. After the chromatography, the sscDNA was amplified by PCR (number of cycles = 6) (*Shcheglov et al., 2007*). For sequencing, cDNA with a size range of 500–800 bp was selected from N1-cDNA using a preparative agarose gel electrophoresis.

The cDNA was sequenced using the "Genome Sequencer Roche GS FLX System" (GATC-Biotech AG, Konstanz, Germany), which produced reads of up to 400 bp length. The reference transcriptome was assembled using the "Roche GS De Novo Assembler": Newbler (version 2.8). In principle, the assembler starts with a read, looks at its sequence, and searches the reading space for another read that contains an overlapping sequence. The overlap is specified by its length and the number or percentage of matching bases (*Margulies et al., 2005*; *Miller, Koren & Sutton, 2010*; *Schliesky et al., 2012*).
Data pre-processing

Newbler was run using default parameters for reads cleaning and transcriptome assembly. Reads that contained homopolymers (60% over the entire length of the read represented by one nucleotide) and low-quality sequences (quality scores <20) were discarded. The primers, adapters and ambiguous residues (Ns) from both sides of the sequence were trimmed and the read length was filtered at 20 bp.

Assembly parameters: minimum overlap length = 40 bp and minimum overlap identity = 90%. Maximum number of contigs in an isogroup, default = 500, maximum number of isotigs in an isogroup = 100, and maximum number of contigs in one isotig = 100. The assembly was not checked for chimeras. The parameters Large or Complex Genome, Heterozygotic Mode, or Expected Depth were not used.

If the assembly of the cDNA raw data indicated 2 or more possible high-quality assemblies of the same sequence, it was presumed that these differences are due to alternative splicing and the contigs were grouped in the same isogroup as isotigs.

Once the assembly had been carried out, the reads that survived the quality control steps were remapped to the reference transcriptome using bwa-0.7.12-mem (*Li, 2013*). The sam/bam files conversion was conducted using samtools-1.3.1 (*Li et al., 2009*) with default parameters.

Assessment of the completeness of the Oenanthe oenanthe transcriptome assembly

The transcriptome was assessed using the Benchmarking Universal Single-Copy Orthologs, BUSCO, version 3.0.2 (*Kriventseva et al., 2014*; *Simão et al., 2015*). BUSCO sets are groups of orthologous near-universally-distributed single-copy genes in each species, selected from OrthoDB (http://www.orthodb.org; *Zdobnov et al., 2016*) across eukaryotes, fungi, arthropods, vertebrates, and metazoans. Their common presence means that any BUSCO gene can be expected to be found as a single-copy ortholog in any newly-sequenced genome or transcriptome from the appropriate phylogenetic clade. The ortholog genes are classified as "Complete", "Fragmented", "Duplicated" or "Missing" according to the "expected-score" and "protein length distribution" values (*Simão et al., 2015*). The approach was conducted using the lineage dataset: aves_odb9 (Creation date: 2016-02-13, number of species: 40, number of BUSCOs: 4,915) and metazoa_odb9 (Creation date: 2016-02-13, number of species: 65, number of BUSCOs: 978). The analysis was conducted using the following parameters: Blast E-value = 0.001, limit = 3 and run mode: transcriptome. Other standard parameters were considered to evaluate the quality of the assembly, such as the number of unique genes, N50, average transcript length, transcript length distribution and overlapping with the CDS used as reference for the annotation and number of annotated genes.

Transcriptome annotation

The annotation of the transcriptome was conducted at the German Cancer Research Center (DKFZ, Heidelberg, Germany) using the published genomes from the Turkey (*Meleagris gallopavo*) and Zebra Finch (*Taeniopygia guttata*), Ensembl release_75, cDNA FASTA file,

February 2014. The alignment was conducted at the nucleotide level using the E-value = 0.001 as a threshold, and no additional filter was used. The reference transcriptome (contigs/isotigs) was further annotated using BlastX version 223 (*Altschul et al., 1997*) against the Swissprot database from UniProt Knowledgebase, (release 2014_05; 545,388 entries). For the alignment, an E-value = 0.001 was set. The annotation was limited to those transcripts that showed a protein-to-protein identity higher than 50%.

We will refer to a "contig" and "isotig" as analogous to an individual transcript. Isogroups are clusters of isotigs that might be inferred as splicing variants. Some isogroups cluster transcripts with different annotation status (either like a "characterized gene", "uncharacterized", "characterized with biological function" or classified as a "novel gene"); we call them "ambiguous".

The non-annotated genes from the first analyses in 2014 were re-annotated in March 2016 using the non-redundant proteins from NCBI database, http://www.ncbi.nlm.nih.gov (*O'Leary Nuala et al., 2016*). Then, 497 new isotigs could be annotated. The criteria used for the annotation were: protein-to-protein identity >50% and alignment length ≥80 bp. All annotations were combined, and the hits that matched a "known protein" with the highest percentage of identity or alignment score were preferred to name the isotigs. If there was no alternative matching to an "uncharacterized protein", this label was kept.

The gene ontology terms from UniProt database, http://www.Uniprot.org (*Bastian et al., 2015*) were used to identify relevant biological functions in the sequence data. A list of the unique UniProt IDs from the annotation was loaded into the "Retrieve/ID mapping" webpage's menu and the "Biological function" and the associated GO ids were harvested.

The R programming language platform, version 3.2.2, was used for basics statistical analysis and plotting (*R Core Team, 2017*).

RESULTS

Sequencing outputs

A total of 1,358,490 sequence reads were obtained from the Genome Sequencer Roche GS FLX System, corresponding to 597,735,000 nucleotides. After filtering the data, 81.18% of the reads remained and 75.69% were assembled in contigs/isotigs (Table 1). Furthermore, ~89% of the reads were successfully re-mapped to the reference transcriptome. The transcriptome assembly yielded 17,539 isogroups, clustering 21,982 contigs/isotigs with an average size of 1,622 bp and N50 value of 2,165 bp; the largest isotig is 14,979 bp long.

Transcriptome assembly and annotation

When the maximum number of isotigs in an isogroup (default 100) was exceeded, these transcripts were grouped independently and identified as "contigs". 44 of these contigs were clustered in 18 isogroups; among them, 12 transcripts have a biological function, and 32 are novel genes.

Table 1 *O. oenanthe* **transcriptome assembly report from Newbler 2.8.**

	amount	%
Total of reads	1,358,490	
Aligned reads	1,102,879	81.18%
Assembled reads	1,028,291	75.69%
Partial reads	74,000	5.45%
Singleton reads	139,611	10.28%
Repeat reads	68,301	5.03%
Outlier reads	18,759	1.38%
Too short reads	29,262	2.15%
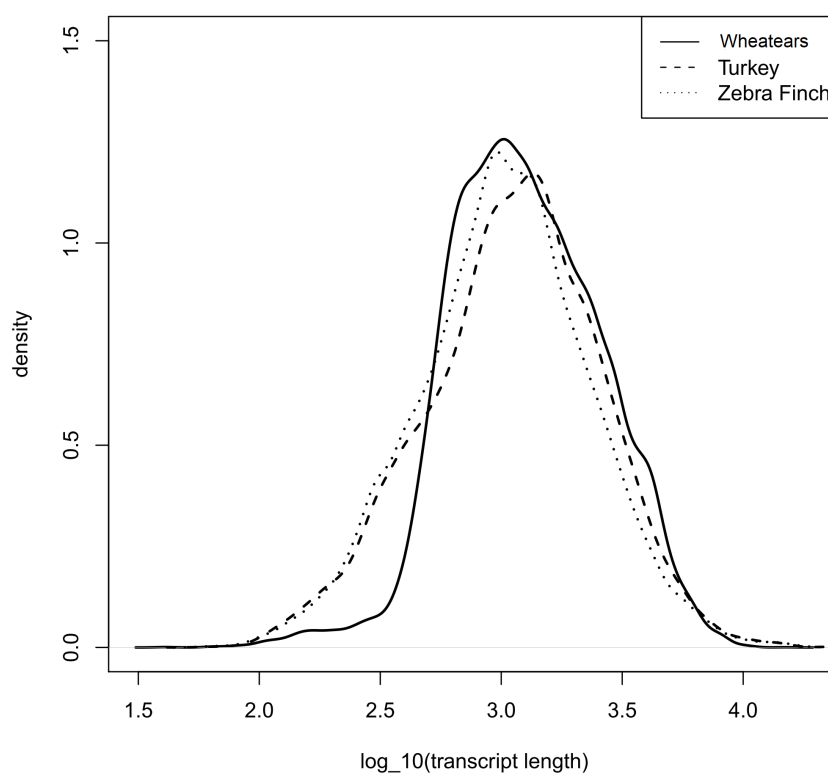
Figure 1 **Transcript length distribution of *O. oenanthe*, *M. gallopavo* and *T. gutatta*, references used for the wheatear's transcriptome annotation.**

The mean length of the transcripts from *O. oenanthe* and the ones from the reference genomes, Turkey and Zebra Finch are similar, (Fig. 1).

Up to 396 isogroups are ambiguous, meaning they cluster isotigs with a different annotation status; we treat them separately to avoid confusion. Most of these isogroups (357) have a single protein name associated with non-annotated or uncharacterized isotigs.

Table 2 Summary of the _O. oenanthe_ transcriptome annotation. Around 6.0% of the genes are "un-characterized", ~49% have biological functions associated (GO) and ~37% could not be annotated. The 396 ambiguous isogroups are not considered here.

Transcriptome	Characterized	Char. with GO	Uncharacterized	Novel	Total
Isogroups	1,442	8,355 (49%)	1,055	6,291 (37%)	17,143
Isotigs	1,757	10,449	1,198	7,016	20,420

For the rest of the isogroups (17,143), all the isotigs have a coherent annotation. Moreover, some isogroups could share the same Uniprot ID and different Uniprot IDs could share the same "gene name" as well. Approximately 67% of the total number of transcripts have orthologous genes in the protein and gene libraries used for the annotation and the residual 33% remain as non-annotated genes. The general information about the transcriptome annotation is summarized in Table 2 and the full annotation can be found in the supplements (Table S1). This Transcriptome Shotgun Assembly project (transcripts longer than 199 bp) has been deposited in the DDBJ/EMBL/GenBank database (http://www.ncbi.nlm.nih.gov/genbank), under the access code GFYT00000000. The version described in this paper is the first one, GFYT00000000. The raw data can be accessed in the "Sequence Read Archive" (SRA) database (http://www.ncbi.nlm.nih.gov/sra), BioProject: PRJNA408074.

The number of genes identified in _O. oenanthe_ might not exceed 15,640, considering that all characterized isogroups with the same description belong to the same gene (7,898). Furthermore, it is assumed that all the novel (6,291), ambiguous (396) and un-characterized (1,055) isogroups are unique genes.

The smallest transcript found is the isotig 21,938 of unknown function (40 bp in length); the largest is the main apolipoprotein of chylomicrons and low-density lipoproteins: "Apolipoprotein B-100" (length 14,979 bp).

Within the ambiguous isogroups, the novel isotigs are significantly shorter than their annotated partners, with a mean length of 1,024 bp (Fig. 2), Student's t-test p-value $= 2.2$ e^{-16}. The characterized transcripts are, on average, 1,913 bp long, the ones with biological functions are a mean 1,924 bp long, the un-characterized 1,626 bp, and 1,621 bp is the mean for the whole transcriptome.

About 10,896 out of 12,206 characterized transcripts have a recognized putative biological function. Among the ambiguous isogroups, 746 out of 1,562 isotigs have an ontology term associated; all these 396 isogroups have at least one characterized isotig with a biological function.

Details regarding the validation of the assembly quality based on orthologous completeness are shown in Table 3.

In total, 8,374 gene ontology terms were compiled (Table S2). Figure 3 shows the GO subcategories shared by more than 100 unique Uniprot.IDs.
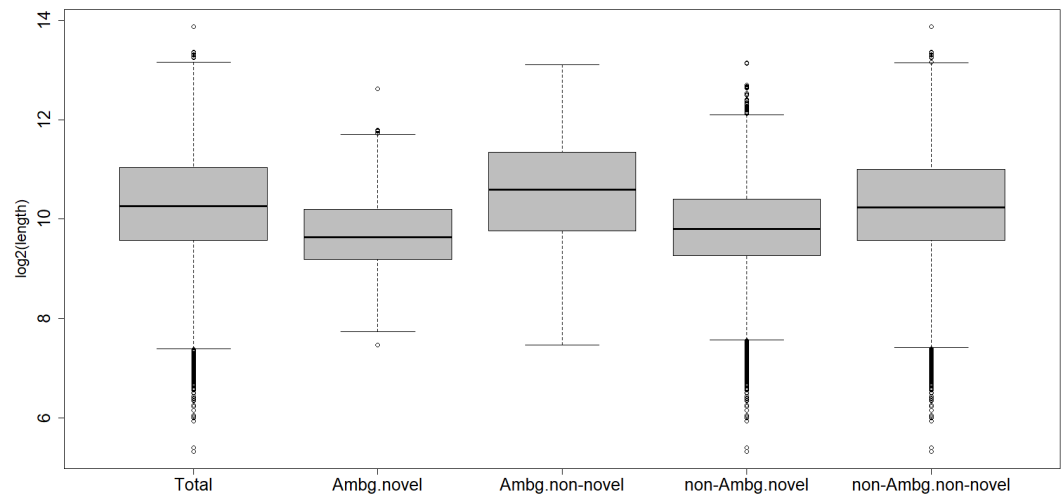
Figure 2 **Length distribution of different subcategories of transcripts.** Total, all transcripts; Ambg.novel, ambiguous and novel isogroups; Ambg.non-novel, ambiguous but annotated isogroups; non.Ambg.novel, novel unambiguous isogroups; non.Ambg.non-novel, annotated and unambiguous isogroups.

Table 3 **Orthologous completeness of the wheatear transcriptome.**

	Total BUSCOs	Complete BUSCOs	Complete and single-copy	Complete and duplicated	Fragmented	Missing
Aves	4,915	2,364 (48.1%)	2,000 (40.7%)	364 (7.4%)	616 (12.5%)	1,935 (39.4%)
Metazoans	978	668 (68.3%)	606 (62.0%)	62 (6.3%)	107 (10.9%)	203 (20.8%)

DISCUSSION

The lack of a reference genome or transcriptome is one of the first challenges at the time to use NGS in studies involving non-model organisms. The assembly of the transcriptome of *O oenanthe*, by using the principal organs involved in, e.g., fattening (intestine, liver and adipose tissue), flight (muscle) and behavior (brain) along the pre-migratory season is not casual. It is the initial step for the understanding of how those processes are regulated at the level of gene expression. The transcriptome we are presenting here is partial, but we believe will be useful to illuminate the genetic architecture behind the development of the migratory phenotype.

The transcriptome of the Northern Wheatear was assembled *de novo*, creating isogroups with the closest related isotigs as putative splicing variants. Later, the transcripts were annotated individually; therefore, 396 isogroups cluster together annotated transcripts and novel ones; we called them ambiguous. Probably because these novel transcripts are smaller, the alignment outputs differ and consequently so do the results of the annotation (Fig. 2).
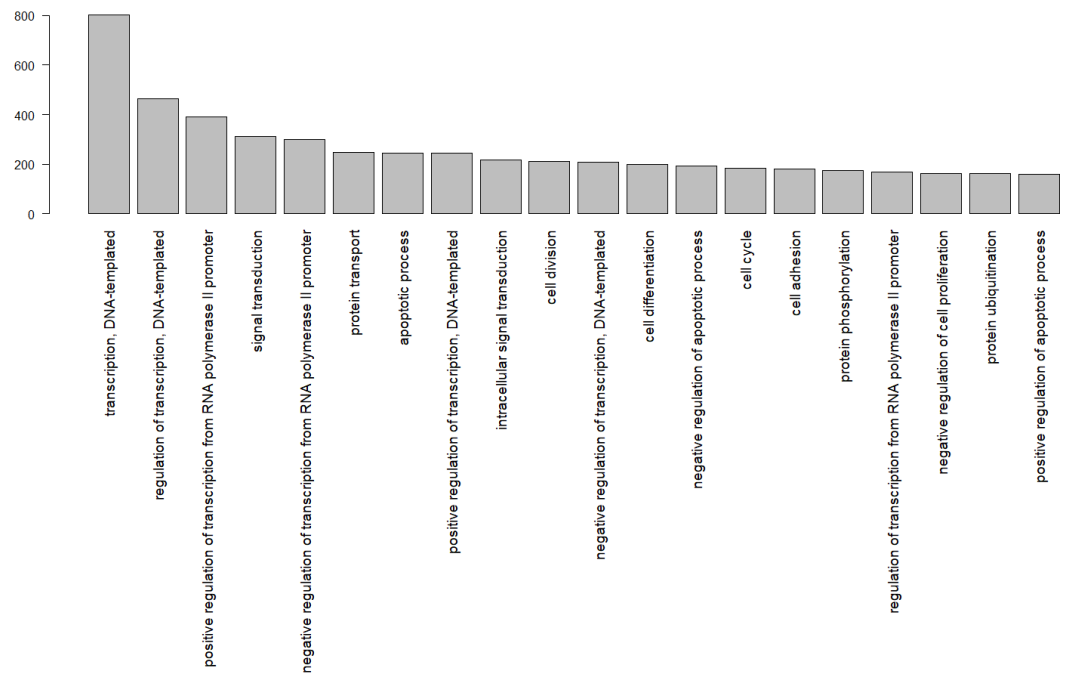
Figure 3 Gene ontology terms shared by at least 100 genes, identified by their Uniprot IDs in the wheatear transcriptome.

Assessment of the quality of the transcriptome assembly

To build up a reliable transcriptome, *de novo*, for non-model organisms coming from natural populations is especially challenging. The indicators of quality should, ideally, yield information about accuracy, completeness, contiguity, chimerism and variant resolution. *Honaas et al. (2016)* have recommended a combination of metrics that best assess the quality of *de novo* transcriptome assemblies: the proportion of reads mapping to an assembly, the recovery of conserved, widely expressed genes, the N50 length, and the total number of unigenes. In practice, it is tough to find one assembly which excels in all aspects of quality (*Bradnam et al., 2013*; *Wang & Gribskov, 2017*).

The transcriptome of *O. oenanthe* was constructed using 75.7% of the original sequenced reads. Approximately 10% of the reads are singletons, and 1.4% are outliers. Analogous results were obtained using the same technology and a similar assembly strategy in other bird species (*Peterson et al., 2012*). These unassembled reads likely correspond to lowly-expressed transcripts with insufficient coverage to enable assembly; they are of low quality or aberrant reads. It has been found that, for assembling non-uniformly expressed transcripts, Newbler can report reads from highly expressed genes as singletons (*O'Neil & Emrich, 2013*), although the wheatears' cDNA library was normalized previous sequencing. Around 89% of the reads that passed the quality control filters mapped back to the contigs/isotigs. We consider that number a good indicator of assembly quality since the final transcripts are the consensus sequences from non-genetically identical individuals, hybrids of two subspecies.

The transcripts have an average size of 1,621 bp and an N50 value of 2,165 bp. These parameters, taken as quality indicators, are controversial and should be used carefully when comparing different studies. They could change depending on the threshold used for transcript lengths and the propensity of the assemblers to create chimeras. N50 should reflect the real values of the mRNAs since, in *de novo* assemblies, it is unknown *a priori* (*O'Neil & Emrich, 2013*). The wheatears' N50 and isotigs' length values compare favourably with other *de novo* transcriptome assemblies (*Singhal, 2013*; *Balakrishnan et al., 2013*; *Theissinger et al., 2016*). Moreover, it lies between the value range of three sparrow transcriptomes: 1,942–4,072 bp (*Balakrishnan et al., 2014*). In that study, the differences in lengths were argued to be strongly influenced by the quality of the sample source, sequencing depth and reading types: single or paired-end. The last two aspects could be considered to improve our work in the future.

Our data was not filtered by sequence length, it will be a decision for future users as to which parameter thresholds are to be used for the analysis of the wheatears' transcriptome. The shortest transcript with a biological function is a splicing regulator, the isotig 21,903: YTH domain-containing protein 1 (*Xiao et al., 2016*) with 106 bp and 90.91% of protein identity to its human orthologous gene. Just 35 transcripts are shorter than ytdc1 (40–105 bp), and all of them are classified as novel genes. We can not discard the presence of chimeras in our data; the most used tools to check for it require, for a reliable output, to have paired-end reads as primary data (*Smith-Unna et al., 2016*).

Incomplete transcript coverage can be reflected in the proportion of genes split into different isogroups, in our data approximately 19% (80% of them with just two isogroups). As an example, titin (TTN) the larger gene found in vertebrates (*Opitz et al., 2003*) is not the largest found in the wheatears' transcriptome. Instead, TTN has 20 isogroups associated, most likely due to the lack of reads in some portion of the mRNA, preventing its full reconstruction.

Completeness of the transcriptome of *O. oenanthe*

The comparison of *O. oenanthe* transcripts with the bird's sets of Benchmarking Universal Single- Copy Orthologs, using BUSCO software, yielded ∼40% of missing genes. Just 2,000 "complete" orthologs out of 4,915 were identified in our dataset. A better recovery, approximately 70%, of common genes occurred when comparing our results with metazoans. Even though these results could be interpreted as a non-optimum sequencing coverage or assembly issues; there is no agreement about the extent to which well-conserved orthologous genes can serve as a proxy for the whole transcriptome (*Peterson et al., 2012*; *Honaas et al., 2016*). That the BUSCO sets of genes are based on information derived from genomes must be taken into account. We should not expect all the referenced genes to be found in the partial transcriptome of wheatears. Usually, the proportion of complete genes is higher when genome completeness is tested (*Simão et al., 2015*).

The transcriptome assembly of the wheatears yielded 21,982 contigs/isotigs, grouped into 17,539 isogroups. The putative number of coding genes identified in this communication, 15,640, is similar to other well-annotated birds' transcriptomes, around 18,000 (see: *Künstner et al., 2010*; *Balakrishnan et al., 2014*; *Finseth & Harrison, 2014*; *Koglin et al.,*
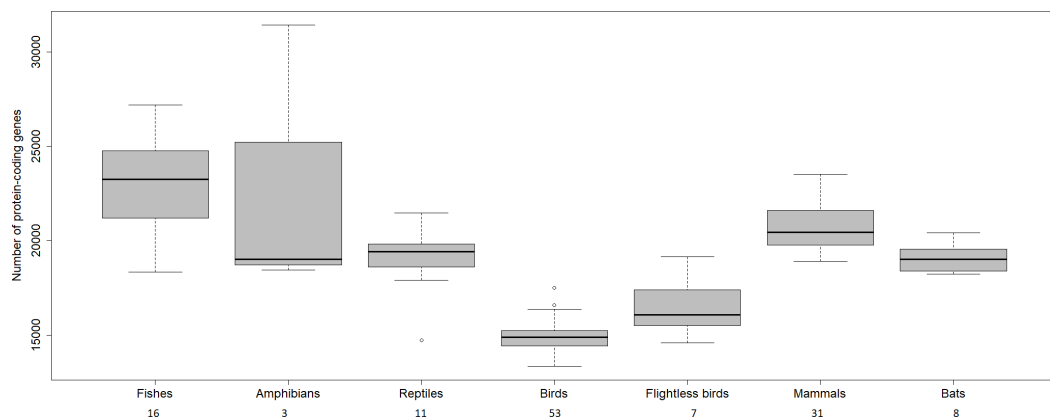
Figure 4   Number of protein-coding genes estimated for different vertebrates. The numbers below refer to the number of species used to construct the figure. Data source: www.ncbi.nlm.nih.gov/genome (*O'Leary Nuala et al., 2016*); www.ensembl.org/info/about/species.html (*Zerbino et al., 2018*).

2017). We should mention that birds have suffered massive deletions of coding and non-coding regions in respect to reptilian sister clades and the rest of the vertebrates (*Huang et al., 2013*; *Lovell et al., 2014*; *Zhang et al., 2014*) (Fig. 4). We consider the finding of a reliable number of genes in wheatears a point in favour of the assembly usefulness. That is especially the case, as the main purpose of our immediate work will be the analysis of the regulated genes related to migration.

Among the biological functions associated with the annotated transcripts of *O. oenanthe*, there is a preponderance of terms related to transcription, signal transduction, and cell division. The analysis of the ontology of the genes related to migration in different tissues will be a topic for upcoming articles.

CONCLUSIONS

We have established the first transcriptome of *O. oenanthe*, which is not complete but contains about 16,000 genes compared to ~18,000 in other birds. More than 66% of the transcripts could be annotated. It will be an important base for further studies based on RNA-Seq of multiple organs and different migratory stages. The information coming from many other newly sequenced genomes and transcriptomes will help to further annotate this transcriptome.

ACKNOWLEDGEMENTS

We would like to thank Hedwig Sauer-Gürth and Heidi Staudter, Alisa Samoylova, and Thore Burgel for their support in the laboratory work. Thanks are due to Prof. Dr. Martin Vingron from the Max-Planck-Institut für Molekulare Genetik for his advice on bioinformatics procedures. We are indebted to Markus Santhosh Braun, Liviu G. Pârâu and Cas Eikenaar for their useful comments on the content of this manuscript. We acknowledge Alisa Samoylova and Douglas Fear for the language revision.

Funding

This work was supported by the Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Germany and the Institute of Avian Research, Wilhelmshaven, Germany. Additional financial support was provided by Deutsche Forschungsgemeinschaft within the funding programme Open Access Publishing, by the Baden-Württemberg Ministry of Science, Research and the Arts and by Ruprecht-Karls-Universität Heidelberg. RCFS was supported by a fellowship from the German Academic Exchange Service (DAAD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Germany.
Institute of Avian Research, Wilhelmshaven, Germany.
Deutsche Forschungsgemeinschaft.
Baden-Württemberg Ministry of Science, Research.
Ruprecht-Karls-Universität Heidelberg.
German Academic Exchange Service (DAAD).

Competing Interests

Michael Wink is an Academic and Subject Editor for PeerJ.

Author Contributions

- Roberto Carlos Frias-Soler conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Lilian Villarín Pildaín and Jonas Kolibius analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Agnes Hotz-Wagenblatt analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Franz Bairlein conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Michael Wink conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

All procedures were in accordance with the "German Animal Welfare Act" and approved by "Zweckverband Veterinäramt JadeWeser" (42508-Te).

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

This Transcriptome Shotgun Assembly project in available at DDBJ/EMBL/GenBank under the accession GFYT00000000.

Data Availability

The raw data have been provided as Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5860#supplemental-information.

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389–3402 DOI 10.1093/nar/25.17.3389.

Bairlein F. 1986. Ein standardisiertes Futter für Ernährungsuntersuchungen an omnivoren Kleinvögeln. *Journal of Ornithology* **127**:338–340 DOI 10.1007/BF01640417.

Bairlein F, Dierschke V, Delingat J, Eikenaar C, Maggini I, Bulte M, Schmaljohann H. 2013. Revealing the control of migratory fueling: an integrated approach combining laboratory and field studies in northern wheatears *Oenanthe oenanthe*. *Current Zoology* **59**:381–392 DOI 10.1093/czoolo/59.3.381.

Bairlein F, Eikenaar C, Schmaljohann H. 2015. Routes to genes: unravelling the control of avian migration—an integrated approach using Northern Wheatear *Oenanthe oenanthe* as model organism. *Journal of Ornithology* **156**:3–14 DOI 10.1007/s10336-015-1224-3.

Bairlein F, Fritz J, Scope A, Schwendenwein I, Stanclova G, Van Dijk G, Meijer HAJ, Verhulst S, Dittami J. 2015. In-flight energy expenditure and metabolic processes in a free-ranging migratory bird. *PLOS ONE* **10(9)**:e0134433 DOI 10.1371/journal.pone.0134433.

Bairlein F, Gwinner E. 1994. Nutritional mechanisms and temporal control of migratory energy accumulation in birds. *Annual Review of Nutrition* **14**:187–215 DOI 10.1146/annurev.nu.14.070194.001155.

Bairlein F, Norris DR, Nagel R, Bulte M, Voigt CC, Fox JW, Hussell DJT, Schmaljohann H. 2012. Cross-hemisphere migration of a 25 g songbird. *Biology Letters* **8**:505–507 DOI 10.1098/rsbl.2011.1223.

Balakrishnan CN, Chapus C, Brewer MS, Clayton DF. 2013. Brain transcriptome of the violet-eared waxbill Uraeginthus granatina and recent evolution in the songbird genome. *Open Biology* **3**:130063 DOI 10.1098/rsob.130063.

Balakrishnan CN, Mukai M, Gonser RA, Wingfield JC, London SE, Tuttle EM, Clayton DF. 2014. Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. *PeerJ* **2**:e396 DOI 10.7717/peerj.396.

Bastian FB, Chibucos MC, Gaudet P, Giglio M, Holliday GL, Huang H, Lewis SE, Niknejad A, Orchard S, Poux S, Skunca N, Robinson-Rechavi M. 2015. The confidence information ontology: a step towards a standard for asserting confidence in annotations. *Database* **2015**:1–11 DOI 10.1093/database/bav043.

Berthold P. 1996. *Control of bird migration.* London: Chapman and Hall, 355.

Bicudo JEPW, Buttemer WA, Chappell MA, Pearson JT, Bech C. 2010. In: Burggren W, ed. *Ecological and environmental physiology of birds.* 1st ed. New York: Oxford University Press, 336.

Boss J, Liedvogel M, Lundberg M, Olsson P, Reischke N, Naurin S, Åkesson S, Hasselquist D, Wright A, Grahn M, Bensch S. 2016. Gene expression in the brain of a migratory songbird during breeding and migration. *Movement Ecology* **4** DOI 10.1186/s40462-016-0069-6.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C, Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G, Gibbs RA, Gnerre S, Godzaridis É, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S, Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li Y, Li Z, Liu B, Liu Y, Luo R, Maccallum I, Macmanes MD, Maillet N, Melnikov S, Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG, Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J, Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF. 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *GigaScience* **2**:1–31 DOI 10.1186/2047-217X-2-10.

Del Hoyo J, Elliott A, Sargatal J. 2005. *Handbook of the birds of the world. 10: Cuckoo-Shrikes to Thrushes.* Barcelona: Lynx Editions.

Finseth FR, Harrison RG. 2014. A comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq in a non-model bird. *PLOS ONE* **9**:e108550 DOI 10.1371/journal.pone.0108550.

Guglielmo CG, Haunerland NH, Williams TD. 1998. Fatty acid binding protein, a major protein in the flight muscle of migrating Western Sandpipers. *Comparative Biochemistry and Physiology Biochemistry and Molecular Biology Part B* **119**:549–555 DOI 10.1016/S0305-0491(98)00016-9.

Gwinner E. 1996. Circannual clocks in avian reproduction and migration. *Ibis* **138**:47–63 DOI 10.1111/j.1474-919X.1996.tb04312.x.

Honaas LA, Wafula EK, Wickett NJ, Der JP, Zhang Y, Edger PP, Altman NS, Pires JC, Leebens-Mack JH, DePamphilis CW. 2016. Selecting superior *de novo* transcriptome assemblies: lessons learned by leveraging the best plant genome. *PLOS ONE* **11**:1–42 DOI 10.1371/journal.pone.0146062.

Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, Kim H, Gan S, Zhao Y, Li J, Yi K, Feng H, Zhu P, Li B, Liu Q, Fairley S, Magor KE, Du Z, Hu X, Goodman L, Tafer H, Vignal A, Lee T, Kim K, Sheng Z, An Y, Searle S, Herrero J, Groenen MAM, Crooijmans RPMA, Faraut T, Cai Q, Webster RG, Aldridge JR, Warren WC, Bartschat S, Kehr S, Marz M, Stadler PF, Smith J, Kraus RHS, Zhao Y, Ren L, Fei J, Morisson M, Kaiser P, Griffin DK, Rao M, Pitel F. 2013.** The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics* **45**:776–784 DOI 10.1038/ng.2657.

Koglin S, Trense D, Wink M, Sauer-Gürth H, Tietze DT. 2017. Characterization of a *de novo* assembled transcriptome of the common blackbird (Turdus merula). *PeerJ* **5**:e4045 DOI 10.7717/peerj.4045.

Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Pozdnyakov IA, Ioannidis P, Sim FA, Zdobnov M. 2014. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Research* **43**:250–256 DOI 10.1093/nar/gku1220.

Künstner A, Wolf JBW, Backström N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK, Jarvis ED, Warren WC, Ellegren H. 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Molecular Ecology* **19**:266–276 DOI 10.1111/j.1365-294X.2009.04487.x.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv preprint. arXiv:1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079 DOI 10.1093/bioinformatics/btp352.

Lovell PV, Wirthlin M, Wilhelm L, Minx P, Lazar NH, Carbone L, Warren WC, Mello CV. 2014. Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biology* **15**:565 DOI 10.1186/s13059-014-0565-1.

Lugo Ramos JS, Delmore KE, Liedvogel M. 2017. Candidate genes for migration do not distinguish migratory and non-migratory birds. *Journal of Comparative Physiology A* **203**:383–397 DOI 10.1007/s00359-017-1184-6.

Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, Åkesson S, Bensch S. 2017. Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1–3**:155–168.

Maggini I, Bairlein F. 2010. Endogenous rhythms of seasonal migratory body mass changes and nocturnal restlessness in different populations of northern wheatear *Oenanthe oenanthe*. *Journal of Biological Rhythms* **25**:268–276 DOI 10.1177/0748730410373442.

Maggini I, Bulte M, Bairlein F. 2017. Endogenous control of fuelling in a migratory songbird. *Science Nature* **104** DOI 10.1007/s00114-017-1514-0.

**Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC,

He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380 DOI 10.1038/nature03959.

Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* **95**:315–327 DOI 10.1016/j.ygeno.2010.03.001.

O'Leary Nuala A, Wright MW, Brister JR, Ciufo S, Haddad D, Mcveigh R, Rajput B, Robbertse B, Smith-white B, Ako-adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar S, Kodali VK, Li W, Maglott D, Masterson P, Mcgarvey M. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**:733–745 DOI 10.1093/nar/gkv1189.

O'Neil ST, Emrich SJ. 2013. Assessing *De Novo* transcriptome assembly metrics for consistency and utility. *BMC Genomics* **14**:465–465 DOI 10.1186/1471-2164-14-465.

Opitz CA, Kulke M, Leake MC, Neagoe C, Hinssen H, Hajjar RJ, Linke WA. 2003. Damped elastic recoil of the titin spring in myofibrils of human myocardium. *Proceedings of the National Academy of Sciences of the United States of America* **100**:12688–12693 DOI 10.1073/pnas.2133733100.

Peterson MP, Whittaker DJ, Ambreth S, Sureshchandra S, Buechlein A, Podicheti R, Choi J, Lai Z, Mockatis K, Colbourne J, Tang H, Ketterson ED. 2012. *De novo* transcriptome sequencing in a songbird, the dark-eyed junco (*Junco hyemalis*): genomic tools for an ecological model system. *BMC Genomics* **13**:305 DOI 10.1186/1471-2164-13-305.

Piersma T, Pérez-tris J, Mouritsen H, Bauchinger U, Bairlein F. 2005. Is there a "migratory syndrome" common to all migrant birds? *Annals New York Academy of Science* **1046**:282–293 DOI 10.1196/annals.1343.026.

R Core Team. 2017. R: A language and environment for statistical computing. Version 3.2.2. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/* .

Ruegg K, Anderson EC, Boone J, Pouls J. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**:4757–4769 DOI 10.1111/mec.12842.

Sambrook J, Russell DW. 2012. Extraction, purification, and analysis of mRNA from eukaryotic cells. In: *The condensed protocols from molecular cloning: a laboratory manual.* New York: Cold Spring Harbor Laboratory Press, 348–445.

Schliesky S, Gowik U, Weber APM, Brautigam A. 2012. RNA-Seq assembly—are we there yet? *Frontiers in Plant Science* **3**:220–220 DOI 10.3389/fpls.2012.00220.

Shcheglov AS, Zhulidov PA, Bogdanova EA, Shagin DA. 2007. Normalization of cDNA libraries. In: Buzdin A, Lukyanov S, eds. *Nucleic acids hybridization: modern applications.* Dordrecht: Springer.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. Genome analysis BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212 DOI 10.1093/bioinformatics/btv351.

Singhal S. 2013. *De novo* transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Molecular Ecology Resources* **13**:403–416 DOI 10.1111/1755-0998.12077.

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. 2016. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Research* **26**:1134–1144 DOI 10.1101/gr.196469.115.

Theissinger K, Falckenhayn C, Blande D, Toljamo A, Gutekunst J, Makkonen J, Jussila J, Lyko F, Schrimpf A, Schulz R, Kokko H. 2016. *De novo* assembly and annotation of the freshwater crayfish Astacus astacus transcriptome. *Marine Genomics* **28**:7–10 DOI 10.1016/j.margen.2016.02.006.

Totzke U, Hübinger A, Bairlein F. 1998. Glucose utilization rate and pancreatic hormone response to oral glucose loads are influenced by the migratory condition and fasting in the garden warbler (*Sylvia borin*). *Journal of Endocrinology* **158**:191–196 DOI 10.1677/joe.0.1580191.

Wang S, Gribskov M. 2017. Comprehensive evaluation of *de novo* transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics* **33**:327–333 DOI 10.1093/bioinformatics/btw625.

Wink M, Kuhn M, Sauer-Gürth H, Witt HH. 2002. Ein Eistaucher (*Gavia immer*) bei Düren-Fundgeschichte und erste genetische Herkunftsuntersuchungen. *Charadrius* **38**:239–245.

Xiao W, Adhikari S, Xiao W, Adhikari S, Dahal U, Chen Y, Hao Y, Sun B, Sun H. 2016. Nuclear m 6 A reader YTHDC1 regulates mRNA splicing. *Molecular Cell* **61**:507–519 DOI 10.1016/j.molcel.2016.01.012.

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Ioannidis P, Seppey M, Loetscher A, Sim A. 2016. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research* **45**:744–749 DOI 10.1093/nar/gkw1119.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, Mclaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. 2018. Ensembl 2018. *Nucleic Acids Research* **46**:D754–D761 DOI 10.1093/nar/gkx1098.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, Zeng Y, Xiong Z, Liu S, Zhou L, Huang Z, An N, Wang J, Zheng Q, Xiong Y, Wang G, Wang B, Wang J, Fan Y, Da Fonseca RR, Alfaro-Núñez A, Schubert M, Orlando L, Mourier T, Howard JT, Ganapathy G, Pfenning A, Whitney O, Rivas MV, Hara E, Smith J, Farré M, Narayan J, Slavov G, Romanov MN, Borges R, PauloMachado J, Khan I, Springer MS, Gatesy J, Hoffmann FG, Opazo JC, Håstad O, Sawyer RH, Kim H, Kim K-W, Kim HJ, Cho S, Li N, Huang Y, Bruford MW, Zhan X, Dixon A, Bertelsen MF, Derryberry E, Warren W, KWilson R, Li S, Ray DA, Green RE, O'Brien SJ, Griffin D, Johnson WE, Haussler D, Ryder OA, Willerslev E, Graves GR, Alström P, Fjeldså J, Mindell PD, Edwards SV, Braun EL, Rahbek C, Burt D, Houde P, Zhang Y, Yang H, Wang J, Consortium AG, Jarvis ED, Gilbert MTP, Wang J. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**:1311–1320 DOI 10.1126/science.1251385.