DO CANDIDATE GENES DISTINGUISH MIGRATORY AND NONMIGRATORY AVIAN SPECIES?

An Undergraduate Research Scholars Thesis

by

OLIVIA LENNON

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Kira Delmore

May 2022

Major: Biology

Copyright © 2022. Olivia Lennon.

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Olivia Lennon, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Faculty Advisor prior to the collection of any data used in this final thesis submission.

This project did not require approval from the Texas A&M University Research Compliance & Biosafety office.

TABLE OF CONTENTS

ABSTRACT

Do Candidate Genes Distinguish Migratory and Nonmigratory Avian Species?

Olivia Lennon
Department of Biology
Texas A&M University

Research Faculty Advisor: Dr. Kira Delmore
Department of Biology
Texas A&M University

Migration is a behavioral syndrome that integrates morphological, behavioral and physiological traits. This syndrome exhibits considerable variation within and between species. This variation in migratory behavior has been demonstrated to have a heritable component, and several candidate genes have been identified in previous studies that are linked to migration. The candidate gene approach seeks to uncover genes responsible for variation in migratory behavior by using prior knowledge of characteristics of genes and characteristics of migratory traits. Here, I reviewed the recent results of candidate gene studies related to migration using the Web of Science database and generated an updated list of likely candidates for this study and future studies. The literature search generated 458 candidate genes, which were then used to test for evidence of positive selection (using a dN/dS analysis) among a dataset of migratory and nonmigratory lineages. The dataset used included 13 species of the genus *Catharus*, including both migratory and nonmigratory species. Eleven candidate genes exhibited evidence of positive selection in at least one species, however, there is not significant evidence of a relationship between the significant candidate genes and migratory behavior.

ACKNOWLEDGEMENTS

Contributors

I would like to thank my faculty advisor, Dr. Kira Delmore, for her guidance and support throughout the course of this research. I would also like to thank Dr. Matthew McKim Louder for his patience and assistance throughout the process.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

The data used for "Do Candidate Genes Distinguish Migratory and Nonmigratory Avian Species?" were provided by the Delmore Lab in the Biology Department at Texas A&M University.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

This undergraduate research project required no funding.

1. INTRODUCTION

Migration is a syndrome that integrates behavioral, morphological and physiological traits. These traits are adaptations that allow individuals to capitalize on seasonal fluctuation of resources (Åkesson & Hedenström, 2007; Dingle & Drake, 2007; Liedvogel, Akesson, & Bensch, 2011). For instance, adaptive traits in the context of migration include endogenous initiating and inhibiting of locomotion, shifts in hormones and metabolism, and changes in social behavior (Barbara, Theunis, & Henk, 2006; Dingle, 2006; Ramenofsky & Wingfield, 2007). Although these adaptations are well studied, the connection between genotype and migratory trait is less investigated. Quantitative genetic work, crossbreeding, and selection experiments have shown that variation in migratory behavior and associated morphological, physiological, and behavioral traits is due to genetic differences, especially in smaller songbirds that migrate at night and thus cannot follow experienced individuals on migration (Berthold, 1996; Berthold & Helbig, 1992; Chernetsov, Berthold, & Querner, 2004; Pulido, 2007). Despite knowledge of the existence of a genetic basis to migration, there remains the question of the actual identity of the genes that are involved in the observed differences.

Among approaches to increase knowledge on the identity of specific genes involved in expression of migratory traits is the use of candidate gene analyses in which researchers (1) identify genes with functions that could be related to migration to and (2) survey variation at these genes (differences in expression or sequence) in migratory species, as demonstrated by previous studies (Christensen et al., 2018; Delmore, Toews, Germain, Owens, & Irwin, 2016; Lundberg et al., 2017; Zhan, Merlin, Boore, & Reppert, 2011). These candidate genes are chosen based upon their function, often in other model organisms.

Among the most widely studied candidate genes examined in bird migration are *Clock* and *Adcyap1* for circadian rhythm, and fatty acid binding proteins (*FABP*s) for fat metabolism. Migratory timing and associated variations in genes are linked to the circadian clock. The *Clock* gene (*Circadian Locomotor Output Cycles Kaput*) codes for the CLOCK protein which forms a heterodimer with BMAL1 to produce a transcription factor which combined drive core circadian oscillators (Panda, Hogenesch, & Kay, 2002). Despite high conservation of *Clock*, the poly-glutamine (poly-Q) polymorphism in the promoter region of this gene accounts for variation in circadian rhythms via the transcription-activating activity of the CLOCK/BMAL1 heterodimer complex (Peterson et al., 2013; Saino et al., 2015). Another circadian rhythm gene, *Adcyap1* encodes pituitary adenylate cyclase-activating polypeptide which stimulates the release of melatonin, directly activating the *Clock* gene. A microsatellite polymorphism in *Adcyap1* predicted variation in *Zugunruhe* (migratory restlessness) in blackcap populations (Mueller, Pulido, & Kempenaers, 2011); longer alleles were positively correlated with higher migratory activity. Similarly, a strong positive association was found between *Adcyap1* allele size and breeding latitude of male Wilson's warblers (G. Bazzi, Galimberti, et al., 2016). In the context of rapid fat metabolism, an adaptive trait apparent among most migratory species, *FABP*s represent a likely candidate gene. *FABP*s are responsible for the transport and metabolism of fatty acids, an important role in migratory birds that are physiologically specialized to accumulate fat stores prior to migration and utilize fatty acids to sustain prolonged flights (Guglielmo, 2018). The apparent involvement of candidate genes such as *Clock*, *Adcyap1*, and *FABPs* in adaptation of migratory traits make these genes likely candidates for a role in shaping the avian migratory phenotype.

A major limitation of the candidate gene approach is that researchers must have a preexisting understanding of the mechanisms underlying the phenotype in question. When little knowledge is available, a potential candidate cannot be identified, and this approach is unsuited. Furthermore, since many candidates are initially discovered in nonmigratory model organisms, many candidates are likely missing. Next-generation sequencing (NGS) represents an alternative and unbiased approach to studying the genetics of migration. NGS platforms were recently developed and reduce the time and cost associated with sequencing allowing their use with non-model organisms like migratory birds. With the many NGS platforms, hundreds of thousands or millions of reads can be produced in a single run. Several studies have combined NGS technology with reduced-representation techniques such as the creation of reduced-representation libraries (RRLs), which sequences a set of genome-wide regions in order to obtain a large number of polymorphisms (Bers et al., 2010). Similarly, another reduced-representation technique is amplified fragment length polymorphism (AFLP), a PCR-based technique that amplifies specific digested DNA fragments in order to compare genomes of interest (Paun & Schönswetter, 2012). A final reduced-representation technique is RAD-tags (Restriction-site Associated DNA tags), an SNP typing method in which fragments are cut using a restriction enzyme and used to identify and genotype SNPs. Utilization of reduced representation of the genome increases cost efficiency and may be more inclusive than the original candidate gene approach. However, reduced-representation techniques still only represent a portion of the entire genome and therefore candidate genes may still be missed in the analysis. Whole-genome sequencing (WGS) is an alternative approach which involves scanning entire genomes in search of genetic markers, regardless of the presence of restriction sites. One study combined whole-genome sequencing and migration tracking technology to identify one gene linked to migration

in two species of warblers, demonstrating the potential for WGS in future studies on migration genetics (Toews, Taylor, Streby, Kramer, & Lovette, 2019). Future development of NGS tools can provide a more comprehensive and unbiased approach in the continued study of the genetic basis underlying migratory orientation.

In the present study, I used a literature review to generate a list of all the genes that have been linked to migration in the literature. I used this list to update our knowledge of migration's genetic basis and the techniques being used to study this topic. In order to determine if any biological functions were over-represented in this list of genes linked to migration, a GO enrichment analysis was performed. An enrichment analysis will identify which, if any, GO terms are either over-represented or under-represented in this particular set of genes. Migration is expected to have a common genetic basis in songbirds. Accordingly, I also tested for positive selection at these genes (using dN/dS analysis to compare synonymous and nonsynonymous substitution rates) in a genus of songbirds that differs in migratory behavior – *Catharus* thrushes. These small, omnivorous songbirds are distributed across the Americas and demonstrate a variety of migratory and non-migratory behavior. Species from this genus have been used to explore multiple aspects of avian evolution including seasonal migration and speciation (Everson et al., 2019). I test the hypothesis that variation in sequences at one or multiple common candidate genes for migration have been selected for in an adaptive response to the requirements of migratory behavior in several avian lineages. If genetic variation in candidate genes included from the literature search account for differences in migratory phenotype, I predict that positive selection should be demonstrated in migratory species lineages. The full dataset included 13 avian species total: 6 migratory *Catharus* genus species (*C. bicknelli, C. guttatus [east], C. minimus, C. ustulatus, C. fuscescens*, and *C. guttatus [west*]), 6 *Catharus* genus non-migratory

species (*C. frantzii*, *C. occidentalis*, *C. mexicanus*, *C. gracilirostris*, *C. aurantiirostris*, and *C. fuscater*), and the migratory wood thrush *Hylocichla mustelina* as an outgroup (Figure 1.1).

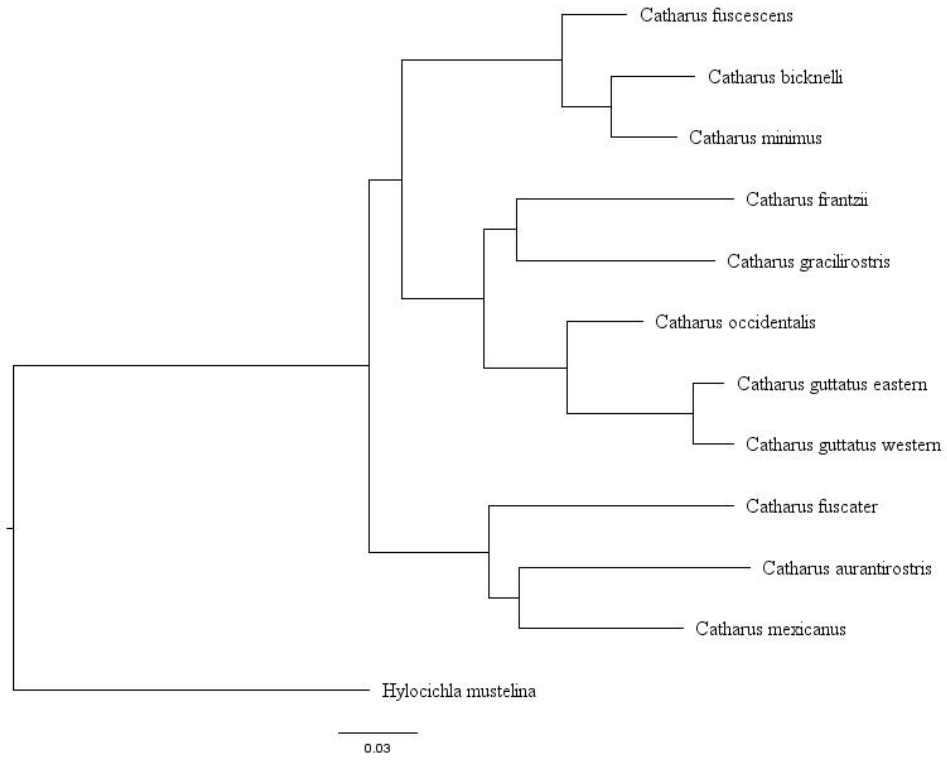

Figure 1.1: Species topology of Catharus thrushes. Hylocichla mustelina was used to root the phylogeny.

2. METHODS

2.1 Literature Review

In this review, I utilized the Clarivate Web of Science database to search for articles related to the genetics of avian migration using studies published in the years 2015-2021 through an all fields keyword search: "((bird OR avian) AND (migration OR migratory) AND gene)." Studies were only used from 2015 onwards because a study by Ruegg, Anderson, Boone, Pouls, and Smith (2014) performed a similar literature search prior to this year. The search returned 923 articles which I individually evaluated for relevance to our focus. Studies were evaluated and picked based on the following criteria: (i) the study focused on molecular differences between migratory and nonmigratory species (e.g. gene polymorphisms and differences in gene expression); (ii) the subject of the study was avian species.

2.2 GO Analysis

In order to find which GO terms are over-represented in the dataset, a GO enrichment analysis was performed using the PANTHER Classification System on the gene list. The GO aspect selected was biological process using *Gallus gallus* as the reference species. *Gallus gallus* was used as the closest relation to bird species. Significant GO terms were assessed using $p \leq 0.05$.

2.3 Alignment of WGS Data and Coding Sequence Extraction

Whole genome sequence libraries were constructed for each individual and sequenced to 20X coverage using Illumina's NovaSeq Sequencing System. The WGS reads of each species were aligned to the Swainson's thrush reference genome using the Burrows-Wheeler Alignment tool (BWA) (Li & Durbin, 2010). The identity of each DNA base within the alignment was

determined and filtered for quality and depth with the program ANGSD (Korneliussen, Albrechtsen, & Nielsen, 2014). DNA sequences for each gene were then extracted for each species with the program GFFRead (Pertea & Pertea, 2020). Using custom scripts in Bash, the sequences of candidate genes for each species were then combined into a single file (using fasta format) for further analysis. Genes for which DNA sequences could not be extracted for *Catharus* species were omitted from the list.

2.4 dN/dS Analysis

Sequences for each candidate gene were analyzed for missing portions and stop codons removed using the software MEGA (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). To test if positive selection has occurred on lineages of migratory species within the *Catharus* genus, I performed an analysis of dN/dS on each branch in the phylogeny on the adaptive Branch-Site Random Effects Likelihood program (aBSREL) of the Datamonkey application (Smith et al., 2015; Weaver et al., 2018). The dataset for each candidate gene included migratory and resident species of the *Catharus* genus. After a priori selection of the migratory branches to test for positive selection on the migratory phenotype, the Likelihood Ration Test was performed at each migratory branch. P-values at each branch were corrected for multiple testing using the Holm-Bonferroni correction and branches under positive selection identified with $p \leq 0.05$.

3. RESULTS

3.1 Literature Review and GO Analysis

My literature review identified 26 papers and 455 genes linked to migratory behavior (see Table A.1). Genes in this list cover a wide array of gene functions including energy metabolism, calcium ion transport, glucose uptake, and circadian oscillators. Core clock components that were mentioned more than once included *CLOCK*, *BMAL1*, *PER2*, *CRY1*, and *ADCYAP1*. Genes involved in metabolism that were repeated included *PDK1*, *FABP3*, *CD36*, *DIO2*, *DIO3*, *FASN*, and *GAD1*.

Among the papers found through the literature search, candidate gene approaches appeared in several studies, most notably through the use of PCR and qPCR techniques to amplify target candidate genes. Table 3.1 provides an overview of papers identified in the literature search along with the techniques employed. Some researchers also used NGS in their search for genes involved in migration. For example, Singh, Swarup, Le, and Kumar (2018) utilized RNA-seq to find differentially expressed genes associated with specific biological processes in the liver between migratory and nonmigratory states in black headed buntings. Similarly, Sharma et al. (2020) used RNA-seq in conjunction with qPCR to show differences in gene expression between migratory states in the hypothalamus, Horton et al. (2019) used RNA-seq to uncover changes in gene expression in the heart and liver tissue, and Johnston, Paxton, Moore, Wayne, and Smith (2016) identified transcriptome-wide gene expression changes in the ventral hypothalamus and optic chiasma. Other NGS techniques employed the use of WGS by Toews et al. (2019), Genotyping-by-sequencing (GBS) by Delmore et al. (2016), and 454 Pyrosequencing by Sharma, Singh, Das, and Kumar (2018).

Table 3.1: Overview of papers found in literature search.

Paper	Species	Dataset (e.g., genomic, RNA-seq, qPCR of candidate genes, sequencing of candidate genes)
bazzi_2016	willow warbler	PCR
bazzi_2015	barn swallow	PCR
bazzi_2016_2	23 trans-Saharan migratory bird species	PCR
boss_2016	willow warblers (2 subspecies)	microarray
contina_2016	11 species of songbird (focus on painted buntings)	PCR
delmore_2015	swainsons thrush	WGS
delmore_2016	swainsons thrush	GBS
demoranville_2019	gray catbird	qPCR
frias-soler_2020	northern wheatear	RNA-seq; qPCR
horton_2019	white-throated sparrows	RNA-seq
johnston_2016	swainsons thrush	RNA-seq
mishra_2018	blackheaded buntings	qPCR
ralston_2019	blackpoll warblers	PCR
saino_2015	4 trans-saharan songbirds (Luscinia megarhynchos; Ficedula hypoleuca; Anthus trivialis; Saxicola)	PCR
sharma_2018	blackheaded buntings	QPC
sharma_2018_2	blackheaded buntings	454 pyrosequencing
sur_sharma_2019	blackheaded buntings	qPCR
sharma_2019_2	blackheaded buntings	qPCR
sharma_2020	blackheaded buntings	RNA-seq
singh_2015	blackheaded buntings	qPCR
singh_2017	blackheaded buntings	qPCR
singh_2018	blackheaded buntings	RNA-seq
toews_2019	vermivora warbler	WGS
trivedi_2015	blackheaded buntings	qPCR

Table 3.1: Overview of papers found in literature search continued.

Paper	Species	Dataset (e.g., genomic, RNA-seq, qPCR of candidate genes, sequencing of candidate genes)
trivedi_2019	blackheaded buntings	qPCR
trivedi_2019_2	blackheaded buntings	qPCR

Note: Genomic refers to datasets composed of electronic collections of genomic data derived from biological samples. RNA-seq refers to RNA-sequencing, and qPCR refers to quantitative polymerase chain reaction.

A GO analysis on all 458 genes yielded significant terms. The most frequent biological functions, apart from over-arching biological processes (100.00%) and cellular processes (78.04%), of the dataset were those related to metabolic processes (Table A.2). Circadian rhythm function had a frequency of only 0.04%, however, this is not surprising given the fewer number of genes involved in circadian rhythm among the entire gene list.

3.2 dN/dS Analysis

In order to identify positive selective forces in migratory species lineages, I completed a gene-wide dN/dS analysis (Table A.1). The analysis for significant nonsynonymous to synonymous mutation rates across the migratory *Catharus* species indicates that *CDH11*, *MYO10*, *NPNT*, *CD36*, *ANKRD54*, *FMN2*, *PDGFRA*, *PEPT1*, *CTRC*, *MADPRT*, and *MAP2* are under positive selection in at least one of the species lineages (Table 3.2). In five of the eleven significant genes, the outgroup *Hylocichla mustelina* was the only significant species detected. Also, *NPNT*, *FMN2*, and *PEPT1* uniquely exhibited positive selection in both migratory and nonmigratory species.

Table 3.2: Gene-wide dN/dS analysis of significant migratory candidate genes.

Gene name	Phenotype	LRT	p-value	Species
CDH11	Migratory	**8.2757**	**0.0391****	*ustulatus*
	Nonmigratory	0.0000	1.0000	*n/a*
MYO10	Migratory	**12.7233**	**0.0041****	*bicknelli*
	Nonmigratory	0.0000	1.0000	*n/a*
NPNT	Migratory	**8.5755**	**0.0048****	*guttatus_west*
	Nonmigratory	**7.3224**	**0.0544**	*fuscater*
CD36	Migratory	**9.0148**	**0.0269****	*ustulatus*
	Nonmigratory	0.0000	1.0000	*n/a*
ANKRD54	Migratory	**8.1595**	**0.0415****	*minimus*
	Nonmigratory	0.0000	1.0000	*n/a*
FMN2	Migratory	**11.9246**	**0.0062****	*hylocichla*
	Nonmigratory	**7.3177**	**0.0546**	*frantzii*
PDGFRA	Migratory	**8.9756**	**0.0314****	*guttatus_east*
	Nonmigratory	0.0000	1.0000	*n/a*
PEPT1	Migratory	**10.4005**	**0.0133****	*hylocichla*
	Nonmigratory	**7.3963**	**0.0524**	*occidentalis*
CTRC	Migratory	**9.8449**	**0.0177****	*hylocichla*
	Nonmigratory	0.0000	1.0000	*n/a*
MADPRT	Migratory	**8.6369**	**0.0326****	*hylocichla*
	Nonmigratory	0.0000	1.0000	*n/a*
MAP2	Migratory	**9.5013**	**0.0210****	*hylocichla*
	Nonmigratory	0.0000	1.0000	*n/a*

Note: Significance of positive selection was assessed with $p<0.1$ indicated with bold font and $p<0.05$ indicated with **. For species labelled “n/a”, none of the species in the given phenotype displayed significance.

4. CONCLUSION

The literature search demonstrated that candidate gene analysis is a continued method of studying the genes underlying migratory behavior. Among the most prevalent recurring genes were those related to the function of internal timing mechanisms (Gaia Bazzi et al., 2015; G. Bazzi, Cecere, et al., 2016; G. Bazzi, Galimberti, et al., 2016; Delmore et al., 2016; Johnston et al., 2016; Ralston et al., 2019; Saino et al., 2015). Genes involved in metabolism were also recurrent throughout several studies (DeMoranville et al., 2019; Horton et al., 2019; Sharma et al., 2018; A. Sharma & Kumar, 2019; Sur, Sharma, Trivedi, Bhardwaj, & Kumar, 2019). In particular, repeated genes are involved in metabolic mechanisms such as fatty acid transport (*PDK1*, *FABP3* and *CD36*), thyroid-hormone response (*DIO2* and *DIO3*), and fatty acid biosynthesis (*FASN* and *GAD1*). Most papers published from 2015-2021 on migration in birds used a candidate gene approach. Furthermore, there is continued support for the role of genes involved in circadian rhythms and metabolism on migratory behavior. Yet it should be noted that recent studies have begun to utilize NGS techniques. Particularly, RNA-seq has demonstrated promise in recent years, often in conjunction with other techniques such as qPCR, in uncovering evidence for candidate genes related to migration using the transcriptome. The advent of NGS will benefit the continued research on the genetic basis of migration as interest in genomics and transcriptomics persists. It is also important to note that the migratory phenotypic variation may likely have complex influences such as epigenetics or may not share a common genetic basis among some species.

Patterns of phenological candidate gene-phenotype associations may be complex as shown by variations in results between sexes and geographically distinct populations (G. Bazzi,

Galimberti, et al., 2016). Furthermore, studies of candidate genes for migration exhibit different results on the significance of focal candidate genes. Female blackcaps with rounder wings and longer *Adcyap1* 3'UTR allele length arrived later in the spring than their counterparts (Mettler, Segelbacher, & Schaefer, 2015). Similarly, Ralstone et al. (2019) found that longer *Adcyap1* alleles in Blackpoll Warblers was correlated with earlier departure of wintering grounds in the spring and later arrival in the fall (Ralston et al., 2019). However, a recent study on migratory variation in bluebirds provided no evidence that *Adcyap1* was associated with a propensity to migrate (Sauve, Dale, Tigano, Ratcliffe, & Friesen, 2021). Evidently, the genetic component of migration is more complex than a single gene or group of genes can explain.

In this study, I tested the hypothesis that candidate genes could distinguish migratory and resident phenotypes within the *Catharus* genus. The dN/dS analysis demonstrated that candidate genes do not determine migratory phenotype in this group. Although positive selection appears to be occurring in particular genes, selection is limited to a single species in each case rather than in all migratory species, indicating that a migratory phenotypic correlation is unsupported. Furthermore, the species within the migratory lineages analysis in five of the eleven significant genes is *Hylocichla mustelina*, an outgroup if the dataset which would therefore be expected to have greater dissimilarity and evolve independently from the *Catharus* species. As such, significant evidence of selection on this species does not further indicate positive selection occurring in the migratory lineages. Finally, *NPNT*, *FMN2*, and *PEPT1* each demonstrated significant positive selection occurring on nonmigratory species as well as migratory. Thus, despite selection on some species within select genes, none of the genes undergoing positive selection exhibited significance in more than one species in either of the independent datasets of separately analyzed migratory and nonmigratory species.

One alternative the hypothesis that candidate genes distinguish migratory and resident phenotypes within the *Catharus* genus is that there is a different genetic basis for migration among bird species within the genus. In other words, a number of genes evolved independently in separate species to influence the migratory phenotype. However, this occurrence is improbable among species so closely related. Alternatively, other genes may not show evidence of positive selection because selection is occurring outside of the region being analyzed. Instead, the genes could be evolving in regulatory regions which are not picked up by this study. This represents a shortcoming of candidate gene analysis in which portions of the genome are selectively analyzed, leading to the omission of possible candidates. Finally, migration may not have a common genetic basis in thrushes. This study was performed under the assumption that all bird species share genetic underpinnings controlling migratory traits. However, in varying species of birds, the genes involved could have evolved separately and therefore have no commonality.

A more thorough understanding of the genetic basis of migration will benefit from further studies encompassing genomics, transcriptomics, and epigenetics in order to take into account the likely complexity of migratory behavior. Continued use of candidate gene analysis and modern NGS techniques including WGS in application to migration remain useful methods of increasing the knowledge base on migration in avian species.

REFERENCES

Åkesson, S., & Hedenström, A. (2007). How Migrants Get There: Migratory Performance and Orientation. *BioScience, 57*(2), 123-133. doi:10.1641/b570207

Barbara, H., Theunis, P., & Henk. (2006). Sociable schedules: interplay between avian seasonal and social behaviour. *Animal Behaviour, 72*(2), 245-262. doi:https://doi.org/10.1016/j.anbehav.2005.12.007

Bazzi, G., Ambrosini, R., Caprioli, M., Costanzo, A., Liechti, F., Gatti, E., . . . Rubolini, D. (2015). Clock gene polymorphism and scheduling of migration: a geolocator study of the barn swallow Hirundo rustica. *Scientific Reports, 5*(1), 12443. doi:10.1038/srep12443

Bazzi, G., Cecere, J. G., Caprioli, M., Gatti, E., Gianfranceschi, L., Podofillini, S., . . . Rubolini, D. (2016). Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. *Molecular Ecology, 25*(24), 6077-6091. doi:10.1111/mec.13913

Bazzi, G., Galimberti, A., Hays, Q. R., Bruni, I., Cecere, J. G., Gianfranceschi, L., . . . Rubolini, D. (2016). Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (Cardellina pusilla). *Ecol Evol, 6*(10), 3226-3239. doi:10.1002/ece3.2053

Bers, N. E. M. V., Oers, K. V., Kerstens, H. H. D., Dibbits, B. W., Crooijmans, R. P. M. A., Visser, M. E., & Groenen, M. A. M. (2010). Genome-wide SNP detection in the great tit Parus major using high throughput sequencing. *Molecular Ecology, 19*(s1), 89-99. doi:https://doi.org/10.1111/j.1365-294X.2009.04486.x

Berthold, P. (1996). *Control of bird migration*.

Berthold, P., & Helbig, A. J. (1992). The genetics of bird migration: stimulus, timing, and direction. *Ibis, 134*(s1), 35-40. doi:https://doi.org/10.1111/j.1474-919X.1992.tb04731.x

Chernetsov, N., Berthold, P., & Querner, U. (2004). Migratory orientation of first-year white storks (Ciconia ciconia): inherited information and social interactions. *Journal of Experimental Biology, 207*(6), 937-943. doi:10.1242/jeb.00853

Christensen, K. A., Leong, J. S., Sakhrani, D., Biagi, C. A., Minkley, D. R., Withler, R. E., . . . Devlin, R. H. (2018). Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. *PloS one, 13*(4), e0195461-e0195461. doi:10.1371/journal.pone.0195461

Delmore, Kira E., Toews, David P. L., Germain, Ryan R., Owens, Gregory L., & Irwin, Darren E. (2016). The Genetics of Seasonal Migration and Plumage Color. *Current Biology, 26*(16), 2167-2173. doi:https://doi.org/10.1016/j.cub.2016.06.015

DeMoranville, K. J., Corder, K. R., Hamilton, A., Russell, D. E., Huss, J. M., & Schaeffer, P. J. (2019). PPAR expression, muscle size and metabolic rates across the gray catbird's annual cycle are greatest in preparation for fall migration. *J Exp Biol, 222*(Pt 14). doi:10.1242/jeb.198028

Dingle, H. (2006). Animal migration: is there a common migratory syndrome? *Journal of Ornithology, 147*(2), 212-220. doi:10.1007/s10336-005-0052-2

Dingle, H., & Drake, V. A. (2007). What Is Migration? *BioScience, 57*(2), 113-121. doi:10.1641/b570206

Everson, K. M., McLaughlin, J. F., Cato, I. A., Evans, M. M., Gastaldi, A. R., Mills, K. K., . . . Winker, K. (2019). Speciation, gene flow, and seasonal migration in Catharus thrushes (Aves:Turdidae). *Molecular Phylogenetics and Evolution, 139*, 106564. doi:https://doi.org/10.1016/j.ympev.2019.106564

Guglielmo, C. G. (2018). Obese super athletes: fat-fueled migration in birds and bats. *J Exp Biol, 221*(Pt Suppl 1). doi:10.1242/jeb.165753

Horton, W., Jensen, M., Sebastian, A., Praul, C., Albert, I., & Bartell, P. (2019). Transcriptome Analyses of Heart and Liver Reveal Novel Pathways for Regulating Songbird Migration. *Scientific Reports, 9*. doi:10.1038/s41598-019-41252-8

Johnston, R. A., Paxton, K. L., Moore, F. R., Wayne, R. K., & Smith, T. B. (2016). Seasonal gene expression in a migratory songbird. *Molecular Ecology, 25*(22), 5680-5691. doi:https://doi.org/10.1111/mec.13879

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics, 15*(1), 356. doi:10.1186/s12859-014-0356-4

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics, 26*(5), 589-595. doi:10.1093/bioinformatics/btp698

Liedvogel, M., Akesson, S., & Bensch, S. (2011). The genetics of migration on the move. *Trends in Ecology & Evolution, 26*(11), 561-569. doi:10.1016/j.tree.2011.07.009

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., . . . Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters, 1*(3), 155-168. doi:https://doi.org/10.1002/evl3.15

Mettler, R., Segelbacher, G., & Schaefer, H. M. (2015). Interactions between a Candidate Gene for Migration (ADCYAP1), Morphology and Sex Predict Spring Arrival in Blackcap Populations. *PloS one, 10*(12), e0144587. doi:10.1371/journal.pone.0144587

Mueller, J. C., Pulido, F., & Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B-Biological Sciences, 278*(1719), 2848-2856. doi:10.1098/rspb.2010.2567

Panda, S., Hogenesch, J. B., & Kay, S. A. (2002). Circadian rhythms from flies to human. *Nature, 417*(6886), 329-335. doi:10.1038/417329a

Paun, O., & Schönswetter, P. (2012). Amplified fragment length polymorphism: an invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. *Methods in molecular biology (Clifton, N.J.), 862*, 75-87. doi:10.1007/978-1-61779-609-8_7

Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research, 9*, 304. doi:10.12688/f1000research.23297.1

Peterson, M. P., Abolins-Abols, M., Atwell, J. W., Rice, R. J., Milá, B., & Ketterson, E. D. (2013). Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research, 2*, 115-115. doi:10.12688/f1000research.2-115.v1

Pulido, F. (2007). The Genetics and Evolution of Avian Migration. *BioScience, 57*(2), 165-174. doi:10.1641/b570211

Ralston, J., Lorenc, L., Montes, M., DeLuca, W. V., Kirchman, J. J., Woodworth, B. K., . . . Norris, D. R. (2019). Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (Setophaga striata). *Ecology and Evolution, 9*(15), 8840-8855. doi:https://doi.org/10.1002/ece3.5436

Ramenofsky, M., & Wingfield, J. C. (2007). Regulation of Migration. *BioScience, 57*(2), 135-143. doi:10.1641/b570208

Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology, 23*(19), 4757-4769. doi:https://doi.org/10.1111/mec.12842

Saino, N., Bazzi, G., Gatti, E., Caprioli, M., Cecere, J. G., Possenti, C. D., . . . Spina, F. (2015). Polymorphism at the Clock gene predicts phenology of long-distance migration in birds. *Molecular Ecology, 24*(8), 1758-1773. doi:10.1111/mec.13159

Sauve, D., Dale, C. A., Tigano, A., Ratcliffe, L. M., & Friesen, V. L. (2021). Do candidate genes for migration and behavior explain migratory variation in bluebirds (Sialia spp.)? *The Wilson Journal of Ornithology, 132*, 820 - 829.

Sharma, Das, S., Komal, R., Malik, S., Rani, S., & Kumar, V. (2020). Seasonal reproductive state determines gene expression in the hypothalamus of a latitudinal migratory songbird during the spring and autumn migration. *Mol Cell Endocrinol, 508*, 110794. doi:10.1016/j.mce.2020.110794

Sharma, Singh, D., Das, S., & Kumar, V. (2018). Hypothalamic and liver transcriptome from two crucial life-history stages in a migratory songbird. *Experimental Physiology, 103*. doi:10.1113/EP086831

Sharma, A., & Kumar, V. (2019). Metabolic plasticity mediates differential responses to spring and autumn migrations: Evidence from gene expression patterns in migratory buntings. *Experimental Physiology, 104*(12), 1841-1857. doi:https://doi.org/10.1113/EP087974

Singh, D., Swarup, V., Le, H., & Kumar, V. (2018). Transcriptional Signatures in Liver Reveal Metabolic Adaptations to Seasons in Migratory Blackheaded Buntings. *Front Physiol, 9*, 1568. doi:10.3389/fphys.2018.01568

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution, 32*(5), 1342-1353. doi:10.1093/molbev/msv022

Sur, S., Sharma, A., Trivedi, A. K., Bhardwaj, S. K., & Kumar, V. (2019). Temperature affects liver and muscle metabolism in photostimulated migratory redheaded buntings (Emberiza bruniceps). *J Comp Physiol B, 189*(5), 623-635. doi:10.1007/s00360-019-01229-5

Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R., & Lovette, I. J. (2019). Selection on VPS13A linked to migration in a songbird. *Proceedings of the National Academy of Sciences of the United States of America, 116*(37), 18272-18274. doi:10.1073/pnas.1909186116

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution, 35*(3), 773-777. doi:10.1093/molbev/msx335

Zhan, S., Merlin, C., Boore, Jeffrey L., & Reppert, Steven M. (2011). The Monarch Butterfly Genome Yields Insights into Long-Distance Migration. *Cell, 147*(5), 1171-1185. doi:https://doi.org/10.1016/j.cell.2011.09.052

APPENDIX

Table A.1: Candidate genes for migration found in literature search.

Gene name	Transcript ID	Significant using dN/dS
AADAT	XM_033061337.1	0
AAMP	XM_033065424.1	0
AANAT	XM_033076667.1	0
ABCD4	XM_033062974.1	0
ACOT9	XM_033053022.1	0
ACSL3	XM_033068595.1	0
ACTG1	XM_033076707.1	0
ADAM10	XM_033070819.1	0
ADARB2	XM_033063178.1	0
ADH5	XM_033061088.1	0
AHR	XM_033064745.1	0
ALAS1	XM_033071947.1	0
ALB	XM_033060719.1	0
ALDOA	XM_033084557.1	0
ANK3	XM_033065796.1	0
ANKRD54	**XM_033057847.1**	**1**
ARHGAP21	XM_033054211.1	0
ARHGEF12	XM_033081683.1	0
ARPP21	XM_033083398.1	0
ATF4	XM_033058732.1	0
ATP1A1	XM_033051992.1	0
ATP5F1E	XM_033075002.1	0
ATP6V0D1	XM_033069641.1	0
ATXN7	XM_033071302.1	0
BDH2	XM_033059756.1	0
BMPR1B	XM_033061222.1	0
BRINP2	XM_033068081.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
BRPF3	XM_033080142.1	0
BUD13	XM_033081566.1	0
CFOS	XM_033063761.1	0
CALM2	XM_033053689.1	0
CAMK2A	XM_033073058.1	0
CAMKV	XM_033071663.1	0
CAV1	XM_033059118.1	0
CBR4	XM_033061183.1	0
CCDC174	XM_033071953.1	0
CCNG2	XM_033061170.1	0
CD36	**XM_033059138.1**	**1**
CD81	XM_033062870.1	0
CDC42SE1	XM_033082933.1	0
CDH11	**XM_033069519.1**	**1**
CDKN1B	XM_033059037.1	0
CDKN2AIP	XM_033061168.1	0
CELA1	XM_033083254.1	0
CEP44	XM_033061382.1	0
CISD2	XM_033059757.1	0
CISH	XM_033071443.1	0
CLCN3	XM_033059481.1	0
CLEC16A	XM_033073890.1	0
CLIC4	XM_033080833.1	0
CLOCK	XM_033060004.1	0
CNP	XM_033078758.1	0
COL1A2	XM_033071651.1	0
COL8A2	XM_033080595.1	0
COQ9	XM_033069507.1	0
COX19	XM_033074187.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
CPE	XM_033060853.1	0
CPM	XM_033057107.1	0
CRABP1	XM_033070970.1	0
CREB1	XM_033064265.1	0
CREBRF	XM_033073123.1	0
CRY1	XM_033058093.1	0
CRY2	XM_033061903.1	0
CTSD	XM_033062045.1	0
CUL5	XM_033052857.1	0
DAPP1	XM_033061342.1	0
DCP1A	XM_033071896.1	0
DCUN1D5	XM_033051762.1	0
DGAT2	XM_033052846.1	0
DICER	XM_033061831.1	0
DIO2	XM_033063528.1	0
DIO3	XM_033063731.1	0
DNAJB14	XM_033061345.1	0
DNALI1	XM_033080831.1	0
DNMT3A	XM_033055172.1	0
DNMT3B	XM_033074829.1	0
DOLPP1	XM_033077360.1	0
DPP6	XM_033053988.1	0
DPYSL2	XM_033081042.1	0
EIF4A2	XM_033068445.1	0
ELOVL1	XM_033067083.1	0
ELOVL6	XM_033059950.1	0
EMCN	XM_033061335.1	0
ENO2	XM_033053208.1	0
EPAS1	XM_033054893.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
EPS15	XM_033067853.1	0
ETS1	XM_033081854.1	0
EYA3	XM_033080631.1	0
FABP3	XM_033080836.1	0
FAHD1	XM_033074153.1	0
FAM107B	XM_033058798.1	0
FAM184B	XM_033060016.1	0
FAM20A	XM_033076811.1	0
FASN	XM_033076668.1	0
FBXL4	XM_033055880.1	0
FGF13	XM_033072300.1	0
FMN2	**XM_033056683.1**	**1**
FNIP2	XM_033060081.1	0
FOXO1	XM_033052704.1	0
FSCN1	XM_033074184.1	0
GABRA5	XM_033053090.1	0
GAD1	XM_033064078.1	0
GAD65	XM_033066287.1	0
GALK2	XM_033070979.1	0
GALNT7	XM_033061072.1	0
GGT7	XM_033074848.1	0
GLI3	XM_033071455.1	0
GLS	XM_033064799.1	0
GLUT5	XM_033079485.1	0
GNAO1	XM_033070160.1	0
GNIH	XM_033069325.1	0
GOT2	XM_033069637.1	0
GRIN1	XM_033077647.1	0
GRINA	XM_033056568.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
GRM5	XM_033085942.1	0
GSK3B	XM_033053150.1	0
GTF2IRD1	XM_033078001.1	0
HAND2	XM_033060901.1	0
HDAC3	XM_033073425.1	0
HIVEP2	XM_033054497.1	0
HMG20A	XM_033070297.1	0
HMGB2	XM_033061255.1	0
HNRNPU	XM_033053788.1	0
HPD	XM_033075877.1	0
HSBP1	XM_033070152.1	0
HSPG2	XM_033079615.1	0
ICA1	XM_033060096.1	0
IL4R	XM_033074278.1	0
IMMT	XM_033059893.1	0
INA	XM_033066121.1	0
INPP4A	XM_033052977.1	0
INSR	XM_033082078.1	0
IRF8	XM_033070176.1	0
ITPK1	XM_033063888.1	0
JCHAIN	XM_033060698.1	0
KAT2A	XM_033079121.1	0
KAT6A	XM_033081140.1	0
KAT6B	XM_033066199.1	0
KCNB1	XM_033075138.1	0
KCNQ5	XM_033055747.1	0
KDM1B	XM_033060515.1	0
KDM3A	XM_033060213.1	0
KDM5A	XM_033057009.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
KDR	XM_033061202.1	0
KIRREL3	XM_033081525.1	0
KIT	XM_033060819.1	0
KLHL2	XM_033061099.1	0
KMT2A	XM_033081521.1	0
KPNA3	XM_033052006.1	0
KRAS	XM_033058183.1	0
LAMA4	XM_033055343.1	0
LAMTOR3	XM_033061344.1	0
LATS2	XM_033052471.1	0
LDB2	XM_033060840.1	0
LMO3	XM_033057918.1	0
LPIN1	XM_033054401.1	0
LSM3	XM_033071680.1	0
MACF1	XM_033080489.1	0
MAN1A2	XM_033051198.1	0
MANBA	XM_033059403.1	0
MAP3K7	XM_033054623.1	0
MAPKBP1	XM_033062140.1	0
MDM2	XM_033057105.1	0
Mel1A	XM_033060322.1	0
Mel1B	XM_033053253.1	0
Mel1C	XM_033072732.1	0
MFAP3L	XM_033061338.1	0
MIB2	XM_033079602.1	0
MID2	XM_033072323.1	0
MMP2	XM_033070153.1	0
MRAS	XM_033064575.1	0
MRC2	XM_033078828.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
MREG	XM_033065320.1	0
MSH4	XM_033065753.1	0
MSMO1	XM_033061102.1	0
MSN	XM_033072493.1	0
MTFR1	XM_033079171.1	0
MTTP	XM_033060617.1	0
MYO10	**XM_033066839.1**	**1**
MYO1C	**XM_033078285.1**	**1**
NAF1	XM_033061033.1	0
NBR1	XM_033078708.1	0
NCAM1	XM_033081618.1	0
NCOA2	XM_033053133.1	0
NDRG4	XM_033069390.1	0
NDUFA6	XM_033057239.1	0
NDUFB10	XM_033074156.1	0
NDUFB9	XM_033081079.1	0
NEK1	XM_033059474.1	0
NFKB1	XM_033059400.1	0
NHLRC1	XM_033056121.1	0
NLRC3	XM_033074517.1	0
NONO	XM_033072548.1	0
NPAS1	XM_033083888.1	0
NPNT	**XM_033060820.1**	**1**
NPY	XM_033081506.1	0
NR3C2	XM_033060511.1	0
NRCAM	XM_033057169.1	0
NRSN1	XM_033070544.1	0
NRXN1	XM_033055067.1	0
OGDH	XM_033081119.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
OSTN	XM_033068853.1	0
PAFAH1B2	XM_033081580.1	0
PAK1	XM_033053077.1	0
PALLD	XM_033060250.1	0
PANK3	XM_033072869.1	0
PAX6	XM_033063103.1	0
PBX1	XM_033067448.1	0
PCBP2	XM_033083338.1	0
PCK1	XM_033074951.1	0
PCM1	XM_033059865.1	0
PDC	XM_033067639.1	0
PDCL	XM_033077853.1	0
PDE12	XM_033071971.1	0
PDGFRA	**XM_033059289.1**	**1**
PDGFRB	XM_033073261.1	0
PDLIM5	XM_033060450.1	0
PDPK1	XM_033074470.1	0
PDYN	XM_033075228.1	0
PEAK1	XM_033070391.1	0
PEPT1	**XM_033051222.1**	**1**
per2	XM_033068458.1	0
PGM2	XM_033061173.1	0
PIK3C2A	XM_033062480.1	0
PIK3CA	XM_033068191.1	0
PITPNC1	XM_033076639.1	0
PLCB4	XM_033054226.1	0
PLCD3	XM_033078782.1	0
PLP1	XM_033072309.1	0
POMC	XM_033056804.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
PPARA	XM_033057130.1	0
PPARG	XM_033072010.1	0
PPP3CA	XM_033059998.1	0
PRKCE	XM_033055005.1	0
PSD	XM_033066508.1	0
PSMA7	XM_033074728.1	0
PSMD1	XM_033068856.1	0
PSMD2	XM_033069164.1	0
PSPH	XM_033078537.1	0
PTGES	XM_033077881.1	0
RAC3	XM_033076501.1	0
RALGAPA2	XM_033053656.1	0
RARS	XM_033073215.1	0
RASGRP1	XM_033062453.1	0
RASL10B	XM_033078327.1	0
RBL1	XM_033074815.1	0
RBM39	XM_033074685.1	0
RBP4	XM_033066716.1	0
RCAN2	XM_033054657.1	0
RGS4	XM_033067311.1	0
RHOBTB1	XM_033066660.1	0
RHO	XM_033071598.1	0
RHOJ	XM_033063030.1	0
RNPEPL1	XM_033068908.1	0
RNPS1	XM_033074247.1	0
RORA	XM_033070470.1	0
RPL22L1	XM_033068751.1	0
RPL30	XM_033059333.1	0
RPL7	XM_033055058.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
RPL9	XM_033060238.1	0
RPS11	XM_033051991.1	0
RPS12	XM_033054552.1	0
RPS15A	XM_033074233.1	0
RPS21	XM_033074964.1	0
RSU1	XM_033072511.1	0
RUNDC1	XM_033078719.1	0
RYR2	XM_033055730.1	0
S100B	XM_033065035.1	0
SAP30	XM_033060052.1	0
SBK1	XM_033073994.1	0
SCD	XM_033065552.1	0
SCRG1	XM_033060053.1	0
SDHD	XM_033081617.1	0
SEMA3G	XM_033071905.1	0
SERPINH1	XM_033052148.1	0
SETX	XM_033077452.1	0
SH3RF1	XM_033060684.1	0
SHISA9	XM_033074415.1	0
SIRT1	XM_033065621.1	0
SLC27A4	XM_033077134.1	0
SLC29A1	XM_033054040.1	0
SLC39A8	XM_033061141.1	0
SLC4A11	XM_033060532.1	0
SLC7A5	XM_033069478.1	0
SMAD1	XM_033061136.1	0
SMARCC2	XM_033083105.1	0
SMARCE1	XM_033079053.1	0
SOSTDC1	XM_033065515.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
SPECC1	XM_033078576.1	0
SRD5A3	XM_033060132.1	0
SRPK2	XM_033057639.1	0
SRSF1	XM_033078363.1	0
SST	XM_033068709.1	0
ST3GAL4	XM_033081532.1	0
STAM2	XM_033065209.1	0
STK32C	XM_033066194.1	0
STK4	XM_033075004.1	0
STMN4	XM_033055168.1	0
SUV39H2	XM_033058317.1	0
SZT2	XM_033067632.1	0
TAC1	XM_033073867.1	0
TACR3	XM_033060989.1	0
TALDO1	XM_033062919.1	0
TAOK1	XM_033078396.1	0
TAT	XM_033070031.1	0
TBCID5	XM_033051678.1	0
TET2	XM_033059835.1	0
TGFBR2	XM_033076276.1	0
TH	XM_033062424.1	0
THRA	XM_033078893.1	0
THRB	XM_033052577.1	0
TIGAR	XM_033058623.1	0
TLN2	XM_033070221.1	0
TMA16	XM_033061059.1	0
TMCO3	XM_033052079.1	0
TMEM132B	XM_033075772.1	0
TMEM165	XM_033060859.1	0

Table A.1: Candidate genes for migration found in literature search continued.

Gene name	Transcript ID	Significant using dN/dS
TMEM192	XM_033060206.1	0
TMPRSS6	XM_033058444.1	0
TRAF6	XM_033061757.1	0
TRIOBP	XM_033058952.1	0
TRPM8	XM_033065381.1	0
TSHB	XM_033080456.1	0
TSTD3	XM_033054676.1	0
TUBB1	XM_033075000.1	0
TXNDC12	XM_033067203.1	0
TXNDC16	XM_033061917.1	0
UBAP2L	XM_033082705.1	0
UBE2M	XM_033084067.1	0
UBE3A	XM_033085174.1	0
UBE3C	XM_033059430.1	0
UBR1	XM_033062014.1	0
UNC5C	XM_033060401.1	0
UQCRQ	XM_033073623.1	0
USP46	XM_033060806.1	0
VIP	XM_033056097.1	0
VOPP1	XM_033074895.1	0
WIPF1	XM_033064845.1	0
YWHAG	XM_033078092.1	0
ZBTB38	XM_033068869.1	0
ZC3H12A	XM_033080978.1	0
ZFAND4	XM_033066338.1	0
ZYX	XM_033053548.1	0

Note: Significant genes are denoted by 1 and in bold. Nonsignificant genes are denoted by 0.

Table A.2: Table of significant GO terms and biological function.

Term ID	Name	Frequency
GO:0008150	biological process	100.00%
GO:0009987	cellular process	78.04%
GO:0008152	metabolic process	62.69%
GO:0071704	organic substance metabolic process	55.50%
GO:0044237	cellular metabolic process	51.52%
GO:0044238	primary metabolic process	49.93%
GO:0006807	nitrogen compound metabolic process	45.36%
GO:0043170	macromolecule metabolic process	34.44%
GO:1901564	organonitrogen compound metabolic process	32.20%
GO:0044260	cellular macromolecule metabolic process	25.54%
GO:0009058	biosynthetic process	24.43%
GO:0065007	biological regulation	23.39%
GO:1901576	organic substance biosynthetic process	23.31%
GO:0050789	regulation of biological process	21.77%
GO:0019538	protein metabolic process	19.82%
GO:0044281	small molecule metabolic process	15.77%
GO:0044267	cellular protein metabolic process	15.25%
GO:0006793	phosphorus metabolic process	14.19%
GO:0050896	response to stimulus	13.72%
GO:0019222	regulation of metabolic process	13.47%
GO:0043412	macromolecule modification	11.24%
GO:0051716	cellular response to stimulus	11.09%
GO:0006355	regulation of transcription, DNA-templated	9.89%
GO:0006082	organic acid metabolic process	9.21%
GO:0071840	cellular component organization or biogenesis	9.05%
GO:0007154	cell communication	7.88%
GO:0016310	phosphorylation	7.70%
GO:0023052	signaling	7.36%
GO:0016043	cellular component organization	7.33%
GO:0044283	small molecule biosynthetic process	6.41%
GO:0009056	catabolic process	6.09%
GO:0006950	response to stress	4.77%
GO:0006629	lipid metabolic process	4.39%
GO:0044248	cellular catabolic process	4.29%
GO:0006468	protein phosphorylation	4.19%
GO:0035556	intracellular signal transduction	3.85%
GO:0042221	response to chemical	3.54%

Table A.2: Table of significant GO terms and biological function continued.

Term ID	Name	Frequency
GO:0044255	cellular lipid metabolic process	3.36%
GO:0033554	cellular response to stress	3.07%
GO:0065008	regulation of biological quality	2.87%
GO:0032787	monocarboxylic acid metabolic process	2.74%
GO:0018193	peptidyl-amino acid modification	2.71%
GO:0022607	cellular component assembly	2.62%
GO:0032501	multicellular organismal process	2.58%
GO:0048519	negative regulation of biological process	2.55%
GO:0032502	developmental process	2.45%
GO:0051246	regulation of protein metabolic process	2.38%
GO:0048518	positive regulation of biological process	2.11%
GO:0065009	regulation of molecular function	2.06%
GO:0009605	response to external stimulus	1.85%
GO:0006357	regulation of transcription by RNA polymerase II	1.76%
GO:0048731	system development	1.41%
GO:0031325	positive regulation of cellular metabolic process	1.31%
GO:0050793	regulation of developmental process	1.30%
GO:0048583	regulation of response to stimulus	1.29%
GO:1901615	organic hydroxy compound metabolic process	1.22%
GO:0010033	response to organic substance	1.10%
GO:0023051	regulation of signaling	1.03%
GO:0010646	regulation of cell communication	1.02%
GO:0042592	homeostatic process	0.94%
GO:0071310	cellular response to organic substance	0.74%
GO:0032879	regulation of localization	0.71%
GO:0051253	negative regulation of RNA metabolic process	0.70%
GO:0009628	response to abiotic stimulus	0.60%
GO:0006325	chromatin organization	0.60%
GO:0002376	immune system process	0.59%
GO:0009719	response to endogenous stimulus	0.58%
GO:1901700	response to oxygen-containing compound	0.56%
GO:0051239	regulation of multicellular organismal process	0.55%
GO:0016042	lipid catabolic process	0.47%
GO:0009991	response to extracellular stimulus	0.47%
GO:0031399	regulation of protein modification process	0.46%

Table A.2: Table of significant GO terms and biological function continued.

Term ID	Name	Frequency
GO:0019220	regulation of phosphate metabolic process	0.45%
GO:0051174	regulation of phosphorus metabolic process	0.45%
GO:0010876	lipid localization	0.43%
GO:0016570	histone modification	0.43%
GO:0010941	regulation of cell death	0.36%
GO:0006304	DNA modification	0.33%
GO:0072329	monocarboxylic acid catabolic process	0.32%
GO:0045595	regulation of cell differentiation	0.29%
GO:0051338	regulation of transferase activity	0.28%
GO:1901698	response to nitrogen compound	0.28%
GO:0042773	ATP synthesis coupled electron transport	0.27%
GO:0019953	sexual reproduction	0.25%
GO:0044242	cellular lipid catabolic process	0.23%
GO:0032101	regulation of response to external stimulus	0.23%
GO:0043543	protein acylation	0.21%
GO:0006306	DNA methylation	0.21%
GO:0010817	regulation of hormone levels	0.20%
GO:0040012	regulation of locomotion	0.20%
GO:0006473	protein acetylation	0.18%
GO:0043069	negative regulation of programmed cell death	0.16%
GO:0051347	positive regulation of transferase activity	0.15%
GO:0060322	head development	0.14%
GO:0018298	protein-chromophore linkage	0.14%
GO:0043408	regulation of MAPK cascade	0.13%
GO:0030155	regulation of cell adhesion	0.13%
GO:0007610	behavior	0.12%
GO:0042445	hormone metabolic process	0.12%
GO:0031667	response to nutrient levels	0.12%
GO:0062012	regulation of small molecule metabolic process	0.11%
GO:0104004	cellular response to environmental stimulus	0.11%
GO:0048511	rhythmic process	0.10%
GO:0051235	maintenance of location	0.09%
GO:0030335	positive regulation of cell migration	0.09%
GO:0019932	second-messenger-mediated signaling	0.08%
GO:0006109	regulation of carbohydrate metabolic process	0.08%

Table A.2: Table of significant GO terms and biological function continued.

Term ID	Name	Frequency
GO:0019216	regulation of lipid metabolic process	0.08%
GO:0050804	modulation of chemical synaptic transmission	0.08%
GO:0099177	regulation of trans-synaptic signaling	0.08%
GO:0045785	positive regulation of cell adhesion	0.07%
GO:0030522	intracellular receptor signaling pathway	0.07%
GO:0018209	peptidyl-serine modification	0.07%
GO:0048017	inositol lipid-mediated signaling	0.06%
GO:0018105	peptidyl-serine phosphorylation	0.06%
GO:0006638	neutral lipid metabolic process	0.06%
GO:0019722	calcium-mediated signaling	0.06%
GO:0050863	regulation of T cell activation	0.06%
GO:0006641	triglyceride metabolic process	0.05%
GO:0015908	fatty acid transport	0.05%
GO:0048015	phosphatidylinositol-mediated signaling	0.05%
GO:0051222	positive regulation of protein transport	0.05%
GO:0046890	regulation of lipid biosynthetic process	0.05%
GO:0001101	response to acid chemical	0.05%
GO:0007623	circadian rhythm	0.04%
GO:0070482	response to oxygen levels	0.04%
GO:0010675	regulation of cellular carbohydrate metabolic process	0.04%
GO:1901568	fatty acid derivative metabolic process	0.04%
GO:0001763	morphogenesis of a branching structure	0.04%
GO:0042752	regulation of circadian rhythm	0.03%
GO:0019915	lipid storage	0.03%
GO:0006636	unsaturated fatty acid biosynthetic process	0.03%
GO:1905952	regulation of lipid localization	0.03%
GO:0036092	phosphatidylinositol-3-phosphate biosynthetic process	0.02%
GO:0002274	myeloid leukocyte activation	0.02%
GO:1901983	regulation of protein acetylation	0.02%
GO:0071453	cellular response to oxygen levels	0.02%
GO:0062013	positive regulation of small molecule metabolic process	0.02%
GO:1901654	response to ketone	0.02%
GO:0043200	response to amino acid	0.02%
GO:0120161	regulation of cold-induced thermogenesis	0.02%
GO:0000038	very long-chain fatty acid metabolic process	0.02%

Table A.2: Table of significant GO terms and biological function continued.

Term ID	Name	Frequency
GO:0042632	cholesterol homeostasis	0.02%
GO:0051775	response to redox state	0.02%
GO:0061035	regulation of cartilage development	0.01%
GO:0035162	embryonic hemopoiesis	0.01%
GO:0120162	positive regulation of cold-induced thermogenesis	0.01%
GO:0009299	mRNA transcription	0.01%
GO:0031057	negative regulation of histone modification	0.01%
GO:0035336	long-chain fatty-acyl-CoA metabolic process	0.01%
GO:0010883	regulation of lipid storage	0.01%
GO:0045471	response to ethanol	0.01%
GO:0097006	regulation of plasma lipoprotein particle levels	0.01%
GO:0042789	mRNA transcription by RNA polymerase II	0.01%
GO:0150104	transport across blood-brain barrier	0.00%
GO:0010885	regulation of cholesterol storage	0.00%
GO:0003413	chondrocyte differentiation involved in endochondral bone morphogenesis	0.00%
GO:0010248	establishment or maintenance of transmembrane electrochemical gradient	0.00%
GO:0044539	long-chain fatty acid import into cell	0.00%
GO:2000322	regulation of glucocorticoid receptor signaling pathway	0.00%
GO:1903010	regulation of bone development	0.00%