4-2014

A PHYLOGENETIC REEVALUATION OF THE GENUS GAVIA (AVES: GAVIIFORMES) USING NEXT-GENERATION SEQUENCING

Quentin D. Sprengelmeyer
qsprenge@nmu.edu

A PHYLOGENETIC REEVALUATION OF THE GENUS *GAVIA* (AVES: GAVIIFORMES) USING NEXT-GENERATION SEQUENCING

By

Quentin D. Sprengelmeyer

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

Office of Graduate Education and Research

2014

SIGNATURE APPROVAL FORM

Title of Thesis: A PHYLOGENETIC REEVALUATION OF THE GENUS *GAVIA* USING NEXT-GENERATION SEQUENCING

This thesis by Quentin Sprengelmeyer is recommended for approval by the student's Thesis Committee and Department Head in the Department of Biology and by the Assistant Provost of Graduate Education and Research.

_____

Committee Chair: Dr. Alec Lindsay Date

_____

First Reader: Dr. Kurt Galbreath Date

_____

Second Reader (if required): Dr. Neil Cumberlidge Date

_____

Department Head: Dr. John Rebers Date

_____

Dr. Brian D. Cherry Date
Assistant Provost of Graduate Education and Research

A PHYLOGENETIC REEVALUATION OF THE GENUS *GAVIA* USING NEXT-GENERATION SEQUENCING

ABSTRACT

Avian phylogenetic analysis based on DNA sequences, rather than morphological characters, has been used in recent decades to resolve systematic relationships. Advancements in molecular techniques have improved avian phylogenetics and have led to new insights on the relationships between and within taxa. Loons (Aves: Gaviiformes) are one of the oldest living lineages of birds, and the order includes five extant species. The morphological cladogram of *Gavia* placed *G. arctica* as a sister species to *G. pacifica*. However, a more recent study based on mtDNA resulted in a discordant tree splitting the *G. arctica*/*G. pacifica* clade, and placed *G. pacifica* as sister to the (*G. immer*, *G. adamsii*) clade. These hypotheses were tested using next-generation sequencing (NGS) data in the form of a RAD-tag dataset comprising 232,094 bps from 2502 variable loci. Bayesian inference, Maximum Likelihood, and Maximum Parsimony phylogenetic analyses of a concatenated dataset strongly supported the traditional phylogeny (*G. stellata*, ((*G. arctica*, *G. pacifica*), (*G. adamsii*, *G. immer*))), and differed from the largely mitochondrially-based hypothesis that placed *G. pacifica* sister to the (*G. immer*, *G. adamsii*) clade. Both internally- and externally-calibrated molecular clock based estimates of divergence dates placed the most recent common ancestor of modern loons in the early Miocene, which is earlier than previously thought, ~21.4 mya (20-22.8 mya) provides a more parsimonious explanation for body size evolution in loons.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Alec Lindsay, and committee members Drs. Kurt Galbreath and Neil Cumberlidge for their guidance. I would also like to thank all members of the Lindsay lab, past and present, for their indispensable assistance. Special thanks to my wife, Emily Sprengelmeyer, for her constant encouragement, enthusiasm, and support.

The Swedish Museum of Natural History Department of Vertebrate Zoology, Fish and Wildlife Service (Alaska Region), Common Coast Research and Conservation, and Biodiversity Research Institute all have my eternal gratitude for providing tissue and blood samples. Without their generosity this project would never have happened. A big thanks to Dr. Michael Sorenson and Jeff DaCosta, for without their help this project would have taken a considerable longer time to complete.

Thanks also to Northern Michigan University and the Department of Biology for financial support through an Excellence in Education Grant, Development fund grant, and a Peter White Scholar Award fund to Dr. Lindsay. Many thanks go to the College of Arts and Sciences and Graduate Education and Research for funding travel costs to present this work at the 2013 International Loon and Diver Workshop, in Hanko, Finland.

I would like to thank Dr. John Bruggink and Michigan Waterfowl Association for financial assistance.

This thesis is dedicated to all the Sweat Hogs of the world who refuse to give up on their quest for self-improvement no matter what the odds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Avian Phylogenetics and Next-Generation Sequencing

Phylogenetic trees estimate relationships between species based on inferences about patterns of characteristics inherited from a series of common ancestors (Lemey et al. 2009). Robust phylogenetic trees can provide insight about biodiversity and species responses to habitat loss, new diseases, and management plans (Baker 2002). Understanding the phylogenetic relationships among avian taxa is important for understanding the evolution of behavioral and life-history traits, the timing of diversification (Fain and Houde 2004) and how phylogenetic similarity influences community structure and species coexistence (Lovette and Hochachka 2006).

Avian phylogenetic analyses based on DNA sequences rather than morphological characters have been used in recent decades to inform taxonomic questions (Edwards et al. 2005). The relationships between different species, based on aligned sequences, can be assessed using different methods that examine characters (each position within the sequence) and states (nucleotides or amino acids found at the position) (Salemi et al. 2009). Many avian phylogenies have been constructed from a single gene or concatenated datasets of multiple genes (Jacobsen and Omland 2011, Alstrom et al. 2011, McCormack et al. 2013, Pulgarin et al. 2013). Concatenated data sets have led to the discovery of previously unknown avian relationships (Hackett et al. 2008) and global diversification rates (Jetz et al. 2012). Previous avian phylogenetic studies have relied on either mitochondrial DNA (mtDNA) (Harlid et al. 1997), a small number of nuclear loci, or limited taxon sampling (Prychitko and Moore 1997, Groth and Barrowclough1999, Lindsay 2002). Both mtDNA and the use of a limited number of nuclear loci

have drawbacks that pose limitations for phylogenetic inference. Sequences taken from mtDNA can weaken data analysis because they are only inherited from the mother (Morin et al. 2004) and sampling a small amount of loci may not provide strong support or well-resolved relationships (Ericson at al. 2005).

Although phylogenetic studies based on concatenated data can be informative, there are drawbacks to using sequences from a limited number of loci. A single concatenated dataset from limited number of loci can produce a statistically well supported but incorrect tree (Kubatko and Degnan 2007). Also genealogies can differ from gene to gene and produce completely different trees from one another (Nordberg and Rosenberg 2002). For example, genes from mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) can produce conflicting trees (Ballard and Whitlock 2004). However, increased sequence lengths collected from an increased number of loci generally produce more accurate phylogenies (Maddison and Knowles 2006).

Recent advancements in DNA sequencing technology have helped to refine avian phylogenetic relationships that were previously based on either morphological characters (Fain and Houde 2004), or on a limited number of genes (Livezey and Zusi 2006). Increases in sequence length and number of loci have grown dramatically in a short amount of time. In only a few years genetic datasets went from 5007 bp collected from five genes (Ericson et al. 2006), to 32,000bp from 19 loci (Hackett et al. 2008), to 539,526 bps from 1541 loci (McCormack et al. 2013).

Next-generation sequencing (NGS) technologies make it possible to collect large genetic datasets (i.e. megabases across thousands of loci) in a fast and cost-efficient manner (Ansorge 2009). One method relies on the use of the Illumina Genome Analyzer. In this method bridge PCR amplification takes place on a solid surface called a flow cell where thousands of copies of

DNA fragments form clusters (Glenn 2011). Hundreds of millions of different clusters can be sequences on a single lane with up to eight lanes per flow cell (Shendure 2008). This technique can produce 1.5 Gigabases (Gb) of single-read, or 3 Gb of paired-end data per run with each read 100-bp in length (Ansorge 2009). Utilizing engineered and inserted barcodes for each sample, multiple uniquely-barcoded samples (96+ per lane), can be pooled and sequenced simultaneously, saving cost and time compared to the traditional capillary method (Glenn 2011). The Illumina platform also has one of the lowest error rates, ~0.1% per base, of all NGS instruments. Large genetic datasets collected from NGS machines are now gaining momentum in phylogenetic studies (Morin et al. 2004). McCormack et al. (2013) collected 539,526bp from 1,541 loci in their study on the evolutionary relationships among Neoaves, and produced a robust phylogeny.

A modified RAD-tag (Restriction-site Associated DNA-tag) technique provides a method for capturing a broad sample of genetic data from many loci from across the genome. In this type of protocol (Baird et al. 2008), homologous short DNA fragments are generated by restriction digestion of the genome. Those fragments are then size-selected to optimize an Illumina read, and then appropriately modified and amplified for loading onto an Illumina flow cell. The use of RAD-tags in phylogenetic and population genetic analyses is beneficial because with NGS thousands of loci can still be assayed from across the genome with high depth per locus creating a condensed snapshot of the genome (Baird et al. 2008, Rubin et al. 2012). Hohenlohe et al. (2011) used RAD-tags to gain better understanding of the genetic diversity among populations of threespine stickleback (*Gasterosteus aculeatus*) and concluded that freshwater populations diverged from oceanic populations. RAD-tags provided Wanger et al. (2012) a large enough genetic dataset to provide phylogenetic resolution for species of cichlid fishes.

A Phylogenetic Reevaluation of the Genus Gavia (Aves: Gaviiformes) Using Next-generation Sequencing

Advancements in molecular techniques have improved avian phylogenetics and have led to new insights on several relationships between and within taxa. The traditional *Gavia* cladogram based on morphological characters, and it places *G. arctica* and *G. pacifica* together in a clade sister to a clade containing *G. immer* and *G. adamsii* (Boertmann 1990). However, a recent study based on mtDNA and nuclear DNA characters resulted in a conflicting tree splitting the *G. arctica*/*G. pacifica* clade (Lindsay 2002). The Lindsay (2002) study was based on only two linkage groups – 4500bp of mitochondrial DNA and 500bp of nuclear intron DNA - and the taxonomic sampling included only single individuals of *G. stellata*, *G. immer* and *G. adamsii*. The limitations of that dataset raise the concern of incomplete linage sorting in that study. The primary goal of this study was to use genetic data collected from next-generation sequencing (NGS) to construct a robust phylogenetic tree of the genus *Gavia*. Specifically, this study looked at the genome-wide phylogenetic signal to evaluate the evolutionary relationships between the five species of *Gavia*. Also, the estimated divergence times generated from the genetic data were used to help better understand the evolutionary and ecological history of the genus *Gavia*.

Gavia species

Loons are one of the oldest living lineages of birds; the ancient status of Gaviiformes is based on a tarsometatarsus fossil of *Neogaeornis wetzeli* from the late Cretaceous period (Olson 1992). Since *N. wetzeli* fossils were discovered in Chile and Antarctica it is presumed that loons originated in the southern hemisphere over 70 MYA and then dispersed to the northern hemisphere. The genus *Colymboides* appears in the Eocene and links the earlier fossils of *N. wetzeli* to the extinct and extant members of the genus *Gavia* (Storer 1956). The early species of

Gavia were smaller in size compared to their modern relatives (Olson and Rasmussen 2001). *Gavia egeriana*, from the early Miocene ~24 Mya, is considered to be the earliest species of the extant genus *Gavia* (Olson and Rasmussen 2001, Mayr and Poschmann 2009). There are fossil records of four other species of *Gavia* from the Miocene: *G. schulzi* (~16 Mya), *G. moldavica* (~11 Mya), *G. brodkorbi* (~10 Mya), and *G. paradoxa* (~10 Mya) (Mlikovsky 1994).

There are currently five extant species of loons that are all aquatic birds that breed on freshwater lakes (very rarely on rivers) in the northern hemisphere. Loons are found on lakes in the tundra, the taiga and northern mixed forests during the summer breeding months. During the winter months loons migrate south to coastal regions (Roselaar et al. 2006). Although historically there have been both four and five recognized species of loons, there is modern consensus of five extant lineages of *Gavia* species: *G. stellata*, *G. arctica*, *G. pacifica*, *G. adamsii*, and *G. immer*.

Gavia stellata is the smallest and the most morphologically distinct of the five extant species (Johnsgard 1987). The small size of *G. stellata* is thought to be a derived character, with large size the ancestral state found in the other four extant species (Boertman 1990). However, the dark grey-brown breeding plumage of *G. stellata* is considered to be the ancestral state, while the more-derived white-checkered pattern the other four species. *G. stellata* breeds on the arctic coasts and interior lakes across northern Alaska, Northwest Territories, and northern Eurasia including Iceland, the British Isles, Scandinavia and Russia. *Gavia stellata'*s wintering range in North America stretches from the Aleutian Islands south to northern Baja Californiain Mexico and all along the Atlantic coast; and in Eurasia south to the Mediterranean, Black, and Caspian Seas, western Pacific coast south to China and Taiwan. The extinct species *G. howardae* from the middle Pliocene is similar in size and shape to *G. stellata* and is thought to be a closely

related sister taxon, if not one of its direct antecedents (Brodkorb 1953, Olson and Rasmussen 2001).

Gavia arctica and *G. pacifica* are almost morphologically identical, although *G. arctica* is distinguished in all plumages by having more exposed white on the flanks, a sleeker and more gray head than the puffy white head of *G. pacifica* and a green iridescence rather than purple on the throat (Birch and Lee 1997). *Gavia pacifica* was previously considered a subspecies of *G. arctica* until the American Ornithologists' Union (1985) recognized it as a valid species based on records of assortative breeding in areas of sympatry in Russia (Stepanian 1975 and Kishchinskij 1980 as in Lindsay 2002). *Gavia arctica* breeds in northern Europe, across northern Siberia, and in a small portion of northwestern Alaska and winters in the Baltic, Black and eastern Mediterranean seas (Birch and Lee 1997). *Gavia pacifica* breeds in Alaska, northern parts of Saskatchewan, Manitoba, Ontario and in eastern Siberia. *Gavia pacifica* mainly winters along the Pacific coast of western North America. *Gavia arctica* and *G. pacifica* are thought to share a common ancestor from the Pliocene, *G. concinna* (Olson and Rasmussen 2001).

Gavia immer and *G. adamsii* are the largest species of *Gavia* (Johnsgard 1987), are morphologically similar and share similar behaviors and vocalizations (Sjölander and Agren 1976). *Gavia immer* and *G. adamsii* have a more pronounced checkered breeding plumage than *G. arctica* and *G. pacifica. Gavia immer* breeds across the northern portions of North America, Greenland, Iceland, and Great Britain (Johnsgard 1987), winters on the Pacific and Atlantic Coasts. *Gavia adamsii* breeds in the far northern portions of Alaska, Nunavut, Northwest Territories, and northwestern and northeastern Siberia and winters along the Pacific coast of Alaska, China, Korea, and Japan. *Gavia immer* and *G. adamsii* are believed to share a common ancestor with *G. fortis* from the early Pliocene (Olson and Rasmussen 2001). Johnsgard (1987)

postulated that *G. adamsii* represents a population that became restricted to the region around the Bering Sea during the Pleistocene and became adapted to breeding in the arctic, leading to its genetic isolation from *G. immer*.

All five species of loons can overlap in range with one another and there have been reports of hybridization between the different species of loons, but only one (between *G. immer* and *G. adamsii*) based on the heterogeneous set of characters has been confirmed (Roselaar et al. 2006).

Gavia Phylogeny

The traditional phylogenetic tree for *Gavia* is based on 21 morphological character states (Boertmann 1990), but recent work based on genetic data has led to a revision of this phylogenetic tree. Traditionally, the genus *Gavia* has been composed of five species placed into three clades: (*G. stellata*, ((*G. arctica*, *G. pacifica*), (*G. adamsii and G. immer*))). The new proposed tree splits *G. arctica* and *G. pacifica* as a monophyletic clade, and places *G. pacifica* sister to the *G. immer* and *G. adamsii* clade (Figure 1: Lindsay 2002).

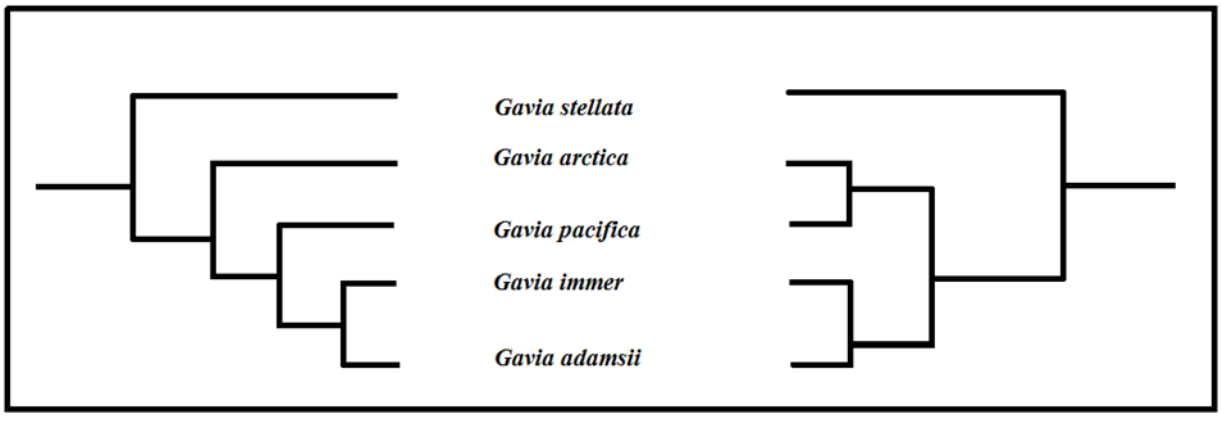

Figure 1. Phylogenetic tree of the genus *Gavia*. Left side of tree represents Lindsay's (2002) hypothesis based on genetic data from ~5000bp of mitochondrial and nuclear intron DNA. Right side of tree represents traditional hypothesis based on 21 morphological character states (Boertmann 1990).

Although support for the tree constructed by Lindsay (2002) was robust, the data used to generate the tree had two main drawbacks. First, the sequence data come from only two linkage groups, ~4500 base-pairs of mitochondrial DNA and ~500 base-pairs of a single nuclear intron. All the nucleotides from the mtDNA data represent just one linkage group and share the same evolutionary history. Zink and Barrowclough (2008) point out that mtDNA-based phylogenies are particularly prone to misleading inferences about evolutionary history, and they do not necessarily always represent the true history of the group under examination (Ballard and Whitlock 2004). With the use of only two loci in the Lindsay (2002) phylogeny raises the possibility that by chance the genes sampled have different relationships than the overall genetic pattern. The second shortcoming of the Lindsay (2002) dataset was possible taxon-sampling problem since only single individuals of *G. stellata, G. immer* and *G. adamsii* were used in the analysis. Kubatko and Degnan (2007) found that sampling only one individual can lead to an incorrect phylogeny even if the analysis is based on concatenated genetic data from multiple loci.

Based on these two drawbacks, the phylogeny that split *G. pacifica* and *G. arctica* could be a result of incomplete linage sorting (Maddison and Knowles 2006). If the Lindsay (2002) phylogeny was a result of incomplete linage sorting and does not represent the true evolutionary history of this group of birds, an analysis of genetic data from multiple loci, sequences from several exemplar individuals of each species may better resolve this tree. The primary goal of this study was to construct a robust phylogenetic tree of the genus *Gavia* allowing a better understanding of the historic events in this lineage. The use of NGS to produce large amount of RAD-tags from multiple individuals of each species (>5) should provide the data necessary to resolve the Gaviiformes phylogenetic tree. The new phylogenetic tree will be compared to the Lindsay (2002) tree.

METHODS

Taxon Sampling

Gavia arctica tissue specimens from seven individuals from Russia (*N*=1) and Sweden (*N*=6) were obtained from the Swedish Museum of Natural History Department of Vertebrate Zoology (Table 1). The US Fish and Wildlife Service (Alaska Region) provided blood samples from 23 individuals: *G. adamsii* (*N*= 9), *G. stellata* (*N*=8), and *G. pacifica* (*N*= 6) all from Alaskan populations. *Gavia immer* samples (*N*=7) were obtained from Biodiversity Research Institute (ME, USA) with individuals from: New York (*N*=1), New Hampshire (*N*=1), Massachusetts (*N*=1), Washington (*N*=1), Maine (*N*=1) and Alaska (*N*=2).

Table 1. Samples used, location of collection, specimen number, and lending institution SMNH: Swedish Museum of Natural History (Stockholm, Sweden); BRI: Biodiversity Research Institute (Maine, United States); and USFW: U.S. Fish and Wildlife Service (Alaska Region, United States) and the date of sample collection.

Species	Sample Number	Sample Location	Specimen Number	Lending Institution	Type	Collection Year
Gavia adamsii	1564	Inigok, AK	YBLO PTT 32936	USFW	Blood	2002
Gavia adamsii	1565	Inigok, AK	YBLO PTT 36401	USFW	Blood	2002
Gavia adamsii	1566	Inigok, AK	YBLO PTT 32934	USFW	Blood	2002
Gavia adamsii	1567	Chippawea River,AK	YBLO C03-36402	USFW	Blood	2003
Gavia adamsii	1568	Chippawea River,AK	YBLO C03-36400	USFW	Blood	2003
Gavia adamsii	1569	Chippawea River,AK	YBLO C03-32951	USFW	Blood	2003
Gavia adamsii	1586	Colville River Delta,AK	YBLO CPD1020 B	USFW	Blood	2000
Gavia stellata	1570	Point Lay,AK	RTLO 1517-78809	USFW	Blood	2009
Gavia stellata	1571	Point Lay,AK	RTLO 1517-78848	USFW	Blood	2009
Gavia stellata	1572	Point Lay,AK	RTLO 1997-10116	USFW	Blood	2009
Gavia stellata	1578	Point Lay,AK	RTLO PCF009	USFW	Blood	2002
Gavia stellata	1579	Point Lay,AK	RTLO ERB034	USFW	Blood	2002
Gavia pacifica	1573	Point Lay,AK	PALO NS10-34	USFW	Blood	2010
Gavia pacifica	1574	Point Lay,AK	PALO 1517-78735	USFW	Blood	2009
Gavia pacifica	1575	Point Lay,AK	PALO 1517-78842	USFW	Blood	2009
Gavia pacifica	1576	Point Lay,AK	PALO 1517-79040	USFW	Blood	2009
Gavia pacifica	1580	Point Lay,AK	PALO CPD1019 B	USFW	Blood	2000
Gavia pacifica	1581	Point Lay,AK	PALO YKD05 B	USFW	Blood	2000
Gavia arctica	1456	Chukotka, Russia	NRM 946654	SMNH	Tissue	1994
Gavia arctica	1457	Uppland,Sweden	NRM 976202	SMNH	Tissue	1997
Gavia arctica	1458	Angermanland, Sweden	NRM 986671	SMNH	Tissue	1998
Gavia arctica	1459	Lake Malaren, Sweden	NRM 20006595	SMNH	Tissue	1987
Gavia arctica	1460	Dalarna, Sweden	NRM 20026057	SMNH	Tissue	2001
Gavia arctica	1461	Skane, Sweden	NRM 20086653	SMNH	Tissue	2008
Gavia arctica	1462	Oland, Sweden	NRM 20116325	SMNH	Tissue	2007
Gavia immer	967	Alaska	93866701	BRI	DNA	2009
Gavia immer	968	Alaska	93866702	BRI	DNA	2009
Gavia immer	984	Massachusetts	93815266	BRI	DNA	2011
Gavia immer	986	Maine	93844808	BRI	DNA	2011
Gavia immer	1031	New Hampshire	93815288	BRI	DNA	2006
Gavia immer	1046	New York	93878821	BRI	DNA	2011
Gavia immer	1589	Washington	938-44840	BRI	DNA	2009

DNA Extraction for NGS

Genomic DNA (gDNA) was extracted from the blood and tissue samples using a silica-based filter purification DNA extraction kit (DNeasy kit; Qiagen, Valenica, CA, USA) in accordance with the manufacturer's protocol. The extracted DNA was quantified with a NanoDrop, and samples were concentrated in a vacuum centrifuge until all were of a concentration of at least 30 ng/µl.

RAD-tag Library Construction

The RAD-tag library was prepared following a double-digested protocol (DaCosta and Sorenson in review). A positive control ("LCAT") was used for the digestion, ligation and amplification steps that were composed of an amplified region of the redhead duck (*Aythya americana*) mitochondrial genome which contained cut sites for both restriction enzymes. Primers designed to target the LCAT fragment are given in Appendix A. Appendices C-G show typical gel images for each steps described below. All thermocycling and thermal incubations were performed in an Eppendorf Mastercycler Gradient thermocycler.

The gDNA was first double-digested by both *Sbf*I-HF and *EcoR*I-HF restriction enzymes (New England BioLabs Inc.) according to the manufacturer's protocols. Each reaction contained 1µg of DNA sample, 1µL of *Sbf*I-HF and *EcoR*I-HF, 5 µL of 10x NEBuffer 4, and brought to 50 µL with dH_2O as needed. The gDNA samples were digested at 37ºC for 30 minutes before enzymes were deactivated by incubation at 65ºC for 20 minutes. Each digested gDNA sample was ligated with a P1 adapter that had a unique six base pair barcode and a "divergent Y" P2 pair–end compatible adapter (Table 1A). The functional nature of the P1 and P2 adapter sequences are described elsewhere (Sorenson and DaCosta, *in review*). Each 70 µL reaction

contained 50 µL digested gDNA, 2 µL 10x NEBuffer 2, 0.6 µL rATP, 4 µL P1 adapter, 12 µL P2 adapter, 0.4 µL water and 1 µL T4 ligase. The samples were ligated at 22ºC for 30 minutes before enzymes were deactivated by incubation at 65ºC for 20 minutes. Custom internal size standards (2µL) of 300 and 450 bp were spiked to each ligation product. Ligated samples were separated on 1.0% low-melt agarose lithium borate (LB) gels (100V for ~150 min), stained with ethidium bromide (EtBr), and visualized under ultraviolet light (Appendix F). Ligated fragments were size selected by performing a wedge cut between the 300 and 450bp fragments, such that only half as much gel was taken from the 300bp end of the cut as compared to the 450bp end of the cut. This helped reduce any bias that may favor the amplification of smaller fragments. The DNA was cleaned using a Qiagen QIAQuick® Gel Extraction Kit following the manufacturer's protocol. The only deviation from the protocol was that 20 µL of the Buffer EB was used instead of 10 µL. The size selected DNA was amplified through a round of PCR in a 60 µL reactions that contained 30 µL of Phusion Mix (New England BioLabs Inc.), 8 µL of water, 3 µL of RAD.F primer, 3 µL of RAD.R primer (Appendix A), and 16 µL of the purified, size-selected DNA template. The DNA fragments were amplified with the following profile: 30 seconds at 98ºC, 26 cycles of the following: 10 seconds at 98ºC, 30 seconds at 60ºC, 40 seconds at 72ºC, after the cycles were completed 5 minutes at 72ºC. The PCR products were purified with magnetic solid-phase reversible immobilization beads. The final product was quantified via real time PCR using a KAPA Biosystems PCR quantification kit. The final library was pooled in equimolar amounts and sent (Tufts University, USA) for sequencing on an Illumina HiSeq 2000 Genome Analyzer.

Bioinformatics Analyses

All the raw data that passed the Illumina filters were processed at Boston University on a custom-designed parallel-processing computer cluster housed in Dr. Michael Sorenson's lab (Dacosta and Sorenson in review) using a Python-scripted pipeline along with several freely available analysis packages. First, the reads were assigned to individual samples based on corresponding barcodes and low quality reads were filtered out (average base Phred score <20) All remaining identical reads from each sample were condensed and counted, and the number of identical reads were recorded for those sequence reads. Condensed sequences from each sample were sorted into "clusters" based on similarity using the UCLUST method in USEACH v5 (Edgar 2010) with an identity threshold of 85%. The sequence with the highest quality read from each cluster was then mapped to the *Columba livia* reference genome using BLASTN v2.2.25 (Altschul et al. 1990), and then combining similar clusters. The sequences in each cluster were then aligned using MUSCLE v3.8.31(Edgar 2004). Genotypes for each individual were generated with the customized python script RADGenotypes.py (https://github.com/BU-RAD-seq/ddRAD-seq-Pipeline).

The final dataset used in these analyses contained no missing data for any individual, although up to 5 low depth or flagged genotypes were allowed for each sample. To generate a concatenated sequence for each sample, when a sample having good heterozygous allele calls, one allele was drawn at random and included in the concatenated dataset. If the sample had a low depth or flagged allele as a part of its genotype, then the allele with the most reads was used. Each consensus locus was mapped to the *Gallus gallus* reference genome via BLASTN v2.2.25 (Altschul et al. 1990).

Phylogenetic Analyses Using Concatenated Data

The concatenated dataset was analyzed using maximum likelihood (ML) in RAXML (Stamatakis 2006) via raxmlGUI (Silvestro and Michalak 2012), maximum parsimony (MP) performed in PAUP* v6.1.7 (Swofford 2002) and Bayesian inference (BI) performed in Mr. Bayes (Ronquist and Huelsenbeck, 2003). *Gavia stellata* is the earliest diverging species of modern *Gavia* (Lindsay 2002, Boertman 1990) and was therefore used as the outgroup to root the trees in all three phylogenetic analyses. Based on MODELTEST results (Posada and Crandall 1998) the BI an ML analyses used a GTR+i+Γ substitution model with parameters estimated from the dataset (Appendix B).

The BI analysis had two independent runs performed simultaneously with each chain length set to 2 million generations that had three heated chains with a chain temperature of 0.2 and a cold chain that is sampled. The subsample frequency was set at every 1000 generations creating 1000 sampled trees (burn-in length of 50%).

In RAxML (Stamatakis 2006) a rapid bootstrap (Stamatakis et al. 2008) and maximum likelihood (ML) analyses were performed simultaneously to search for the best-scoring ML tree. The number of bootstrap replications was set to 1000.

The MP analysis was performed in PAUP* (Swofford 1998). Heuristic searches were conducted with 1000 replicates with 100 bootstrap replicates with random stepwise addition of taxa.

Genetic Distance Calculations

The genetic distance (p-distances) were estimated in PAUP* (Swofford 1998). The uncorrected p-distances were used to estimate divergence dates based on assumed nuclear intron evolution rates of 0.12% (Lerner et al. 2011) and 0.36% (Axlesson et al. 2004) . Divergence

dates were also estimated from a molecular clock calibrated with: 1) *G. concinna* as the most common recent ancestor of *G. pacifica* and *G. arctica* (~ 4.8 Mya), 2) oldest known fossil of *G. pacifica* (~2 Mya), 3) *G. fortis* as the most common recent ancestor of *G. immer* and *G. adamsii* (~4 Mya), and 4) the oldest known fossil of *G. immer* (~2 Mya).

RESULTS

From the Illumina run we retrieved ~70.6 million reads that had high quality scores across all bases of each fragment (Figure 2). On average, each sample contained ~1.5 million reads (Figure 3), and ranged from 690,524- 2,050,224 reads per sample. The average depth of read was ~317 reads per locus per individual. In total there were 3521 putative loci with 318 that were invariant, although not all loci were found in all individuals. The length of each locus was ~ 91 bps.

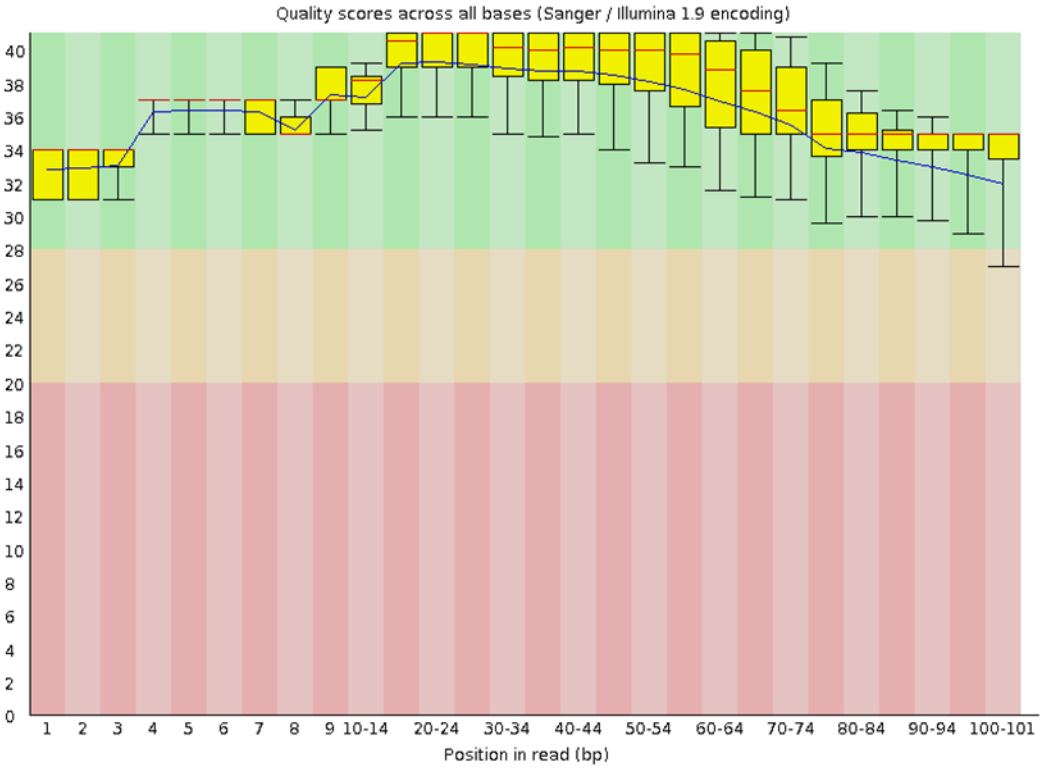

Figure 2: Quality scores from the initial Illumina run for the 70.6 Million reads. All reads, each read is up to 100bp in length, passed with the highest quality score.

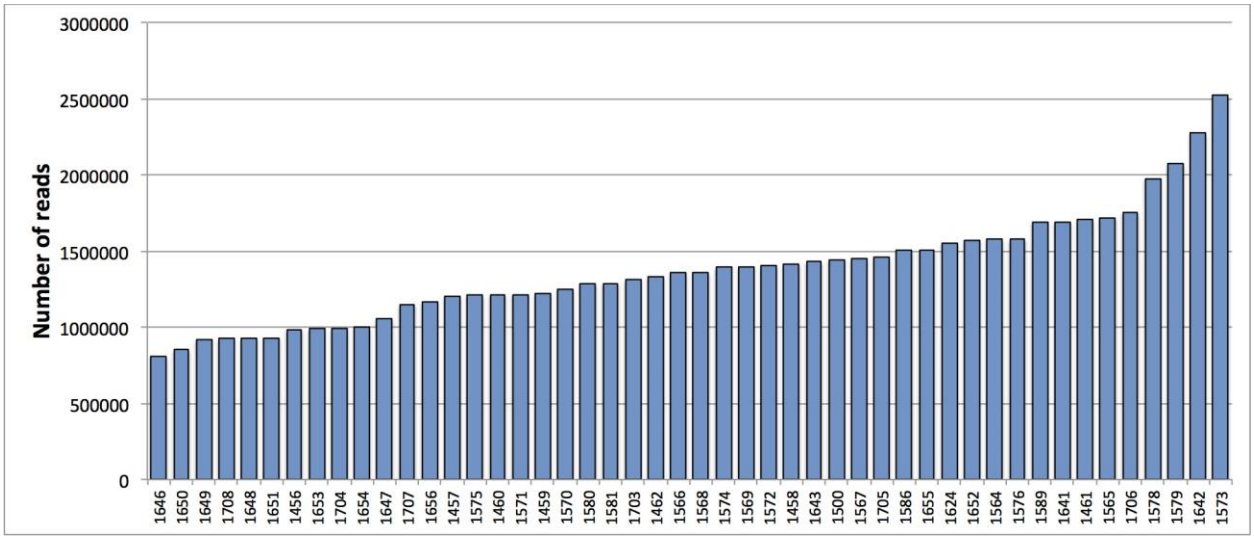

Figure 3: The number of reads, listed on the y-axis, for each sample that is listed on the x-axis.

A final dataset for phylogenetic analyses was constructed from the 3203 variable loci. This dataset comprised genotypes for all 35 samples built from 2502 variable loci with no missing data, but included loci with up to five low-depth or flagged genotypes for each sample. If the sample had unflagged and high-depth alleles at a locus, one allele or another was randomly chosen from the genotype. Conversely, if a sample had a low-depth or flagged allele, that allele was dropped and the unflagged high-depth allele was kept in the dataset. This is a haploid dataset with mostly random selection of alleles at each locus kept for each individual. The BLAST results of the consensus sequences yielded 1087 hits spread out among 28 different chromosomes mapped to the *Gallus gallus* reference genome (Figure 4).

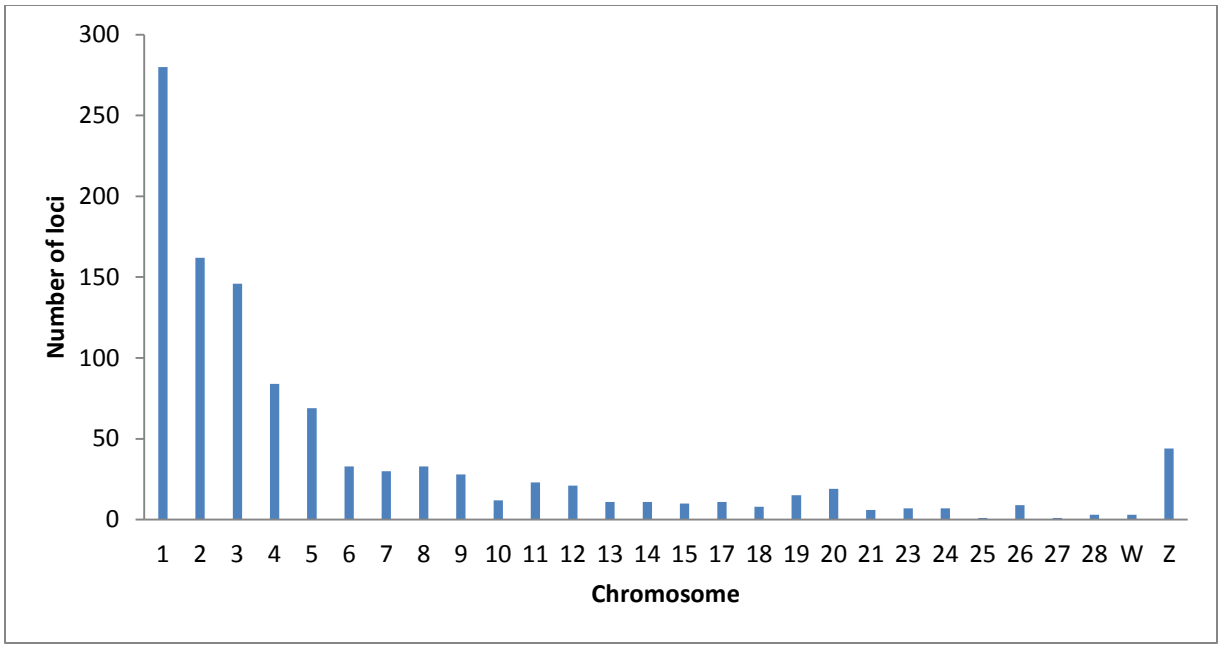

Figure 4. The BLAST results of the consensus sequences of each locus from the dataset. The sequences were mapped to the *Gallus gallus* reference genome.

The BI, ML, and MP phylogenetic analyses of the concatenated dataset yielded the same topology (Figure 5). All three analyses had strong support for a (*G. stellata*, ((*G. arctica*, *G. pacifica)*, (*G. adamsii* and *G. immer)))* phylogeny: with a ML bootstrap of 100% (Figure 6), and a MP bootstrap of 100% (Figure 7), and a Bayesian posterior probabilities of 100% at each of the four nodes (Figure 8).

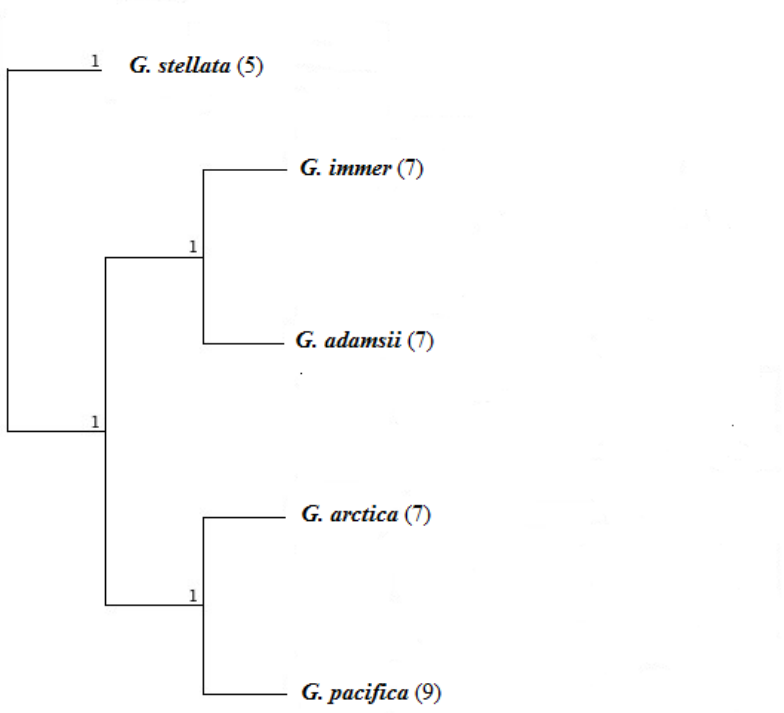

Figure 5. Consensus tree for the five species of *Gavia* from MP, ML, and BI phylogenetic analyses. The trees were constructed with the LF dataset. The numbers of individuals of each species are shown in parentheses. All three techniques had the high support at each node (1); Bayesian Inference a posterior probability of 100%, a Maximum likelihood bootstrap value of 100% and a Maximum Parsimony bootstrap value of 100%.

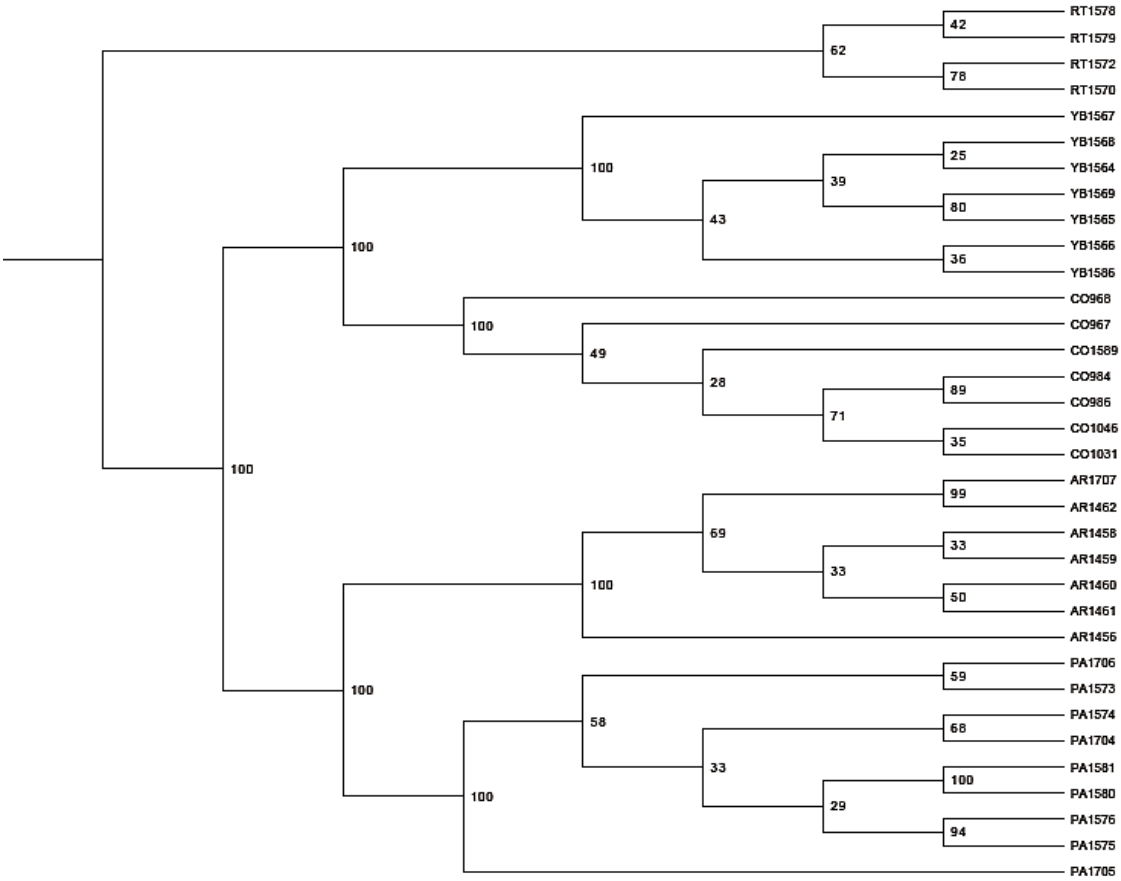

Figure 6. Phylogeny for the five species of *Gavia*. The tree was constructed with LF data set in RAxML using a rapid bootstrap and maximum likelihood method. Maximum likelihood bootstrap values are listed at each node. Samples are labeled at the tip of the branches; RT: *G. stellata*, AR: *G. arctica*, PA: *G. pacifica*, CO: *G. immer*, and YB: *G. adamsii*.

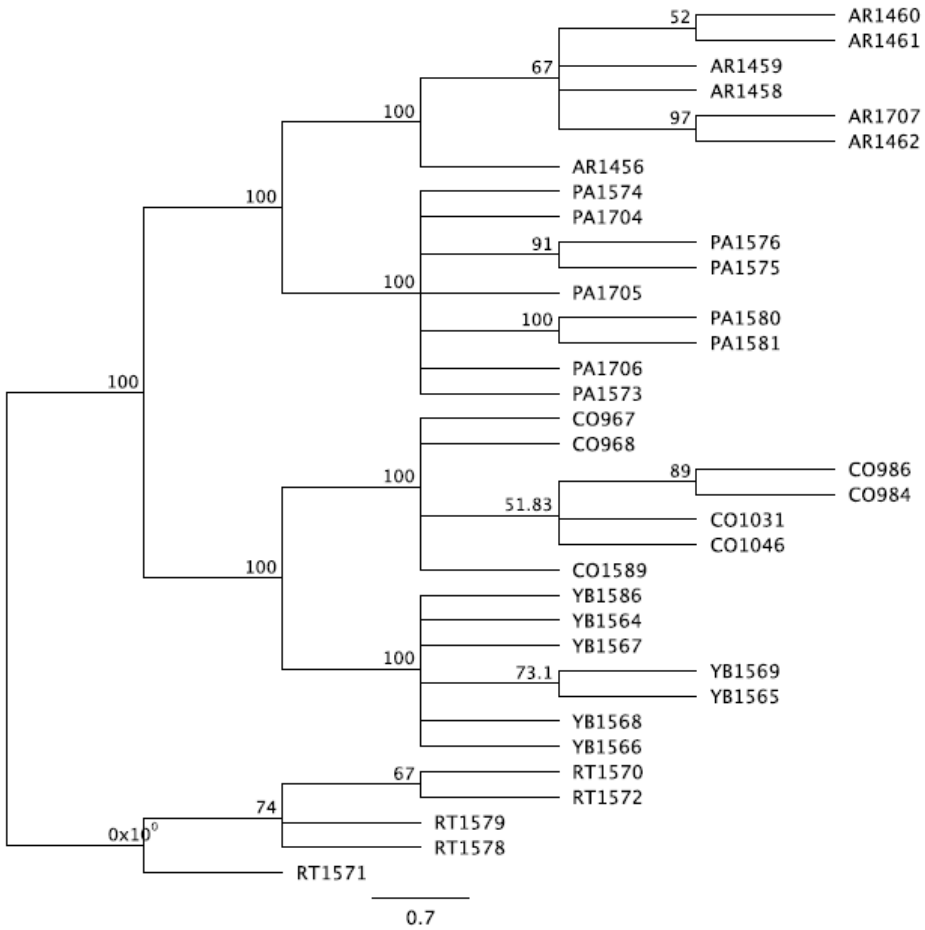

Figure 7. Phylogeny for the five species of *Gavia* from the MP analysis constructed with LF data in PAUP*. MP bootstrap values are listed at each node. Samples are labeled at the tip of the branches; RT: *G. stellata*, AR: *G. arctica*, PA: *G. pacifica*, CO: *G. immer*, and YB: *G. adamsii*.

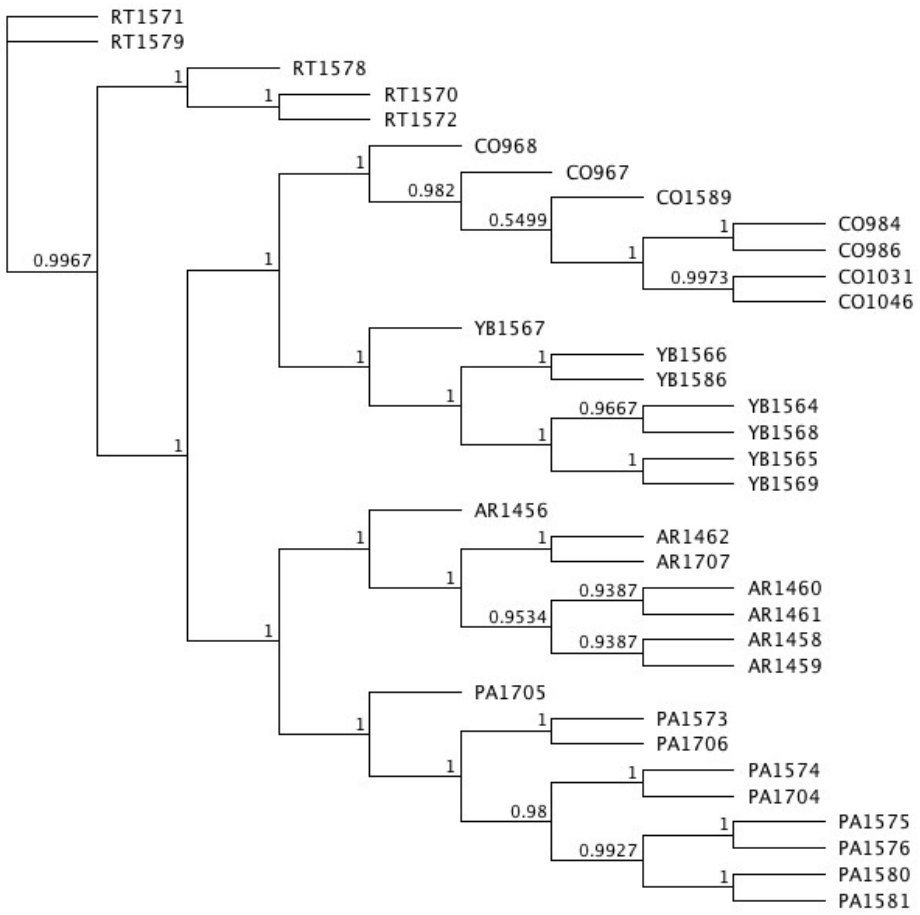

Figure 8. Phylogeny for the five species of *Gavia* from the BI analysis constructed in Mr. Bayes. The BI posterior probabilities are labeled at each node. Samples are labeled at the tip of the branches; RT: *G. stellata*, AR: *G. arctica*, PA: *G. pacifica*, CO: *G. immer*, and YB: *G. adamsii*.

Table 2 shows the average uncorrected patristic distances. All the intraspecific distances had relatively small ranges providing confidence that the samples were correctly identified and appropriate to use in the analyses (Table 3). All intraspecific distances were less than interspecific distances; *G. adamsii* had the smallest intraspecific distance most likely due to the fact that all samples came from a fairly small geographic region (Table 1). *Gavia stellata* is the most diverged of the five species, ~ 3.33% differences with the other four species of (Table 2).

There is ~0.85% divergence between *G. arctica* and *G. pacifica*, while *G. immer* and *G. adamsii* share the smallest distance between any of the species with an average divergence of 0.45%.

Table 2. The average uncorrected patristic distance between each species. Distances were calculated in PAUP*.

	G. stellata	*G. arctica*	*G. pacifica*	*G. immer*	*G. adamsii*
G. stellata	0.0034				
G. arctica	0.033	0.0063			
G. pacifica	0.033	0.0085	0.0057		
G. immer	0.033	0.0109	0.0115	0.003	
G. adamsii	0.033	0.0098	0.0104	0.0045	0.0008

Table 3. Ranges of uncorrected patristic distance between each species. Distances were calculated in PAUP*.

	G. stellata	*G. arctica*	*G. pacifica*	*G. immer*	*G. adamsii*
G. stellata	0.003-0.004				
G. arctica	0.033-0.035	0.004-0.006			
G. pacifica	0.033-0.035	0.007-0.009	0.002-0.006		
G. immer	0.033-0.035	0.009-0.012	0.009-0.012	0.002-0.004	
G. adamsii	0.032-0.034	0.009-0.01	0.009-0.011	0.003-0.005	0-0.001

Calibrating the molecular clock with *G. concinna* (MRCA of *G. pacifica* and *G. arctica* ~ 4.8 mya) and the earliest known fossil of *G. pacifica* (~2mya) gives a range of molecular evolution of 0.18-0.42% per million years. If the molecular clock is calibrated with *G. fortis* (MRCA of *G. immer* and *G. adamsii* ~4.8 mya) along with the earliest known fossil of *G.immer* (~2mya) a range of 0.11%-0.23% is generated. Averaging of the two faster and slower rates gives a range of 0.15-0.33% per million years. The slower rate of 0.15% is comparable to the average mutational rate found in Hawaiian honeycreeper nuclear introns (Fringillidae: Drepanidinae; 0.12 %) (Lerner et al. 2011) and the faster rate of 0.33% is close to the 0.36% per million years Axelsson et al. (2004) found in chickens and turkeys. The faster molecular evolution rates of 0.33-0.36% per million year puts the divergence of *G. stellata* from the other four species of *Gavia* at middle to late Miocene ~ 8.4mya (Table 4 and Table 6), whereas the

slower rates place this event in the early Miocene ~21.4 mya (Table 4 and Table 5). The faster rates place the divergences of *G. arctica* and *G. pacifica* from one another and from the *G.immer* /*G. adamsii* clade in the late Pliocene ~2-3mya (Table 4 and Table 6), which differs from the slower rates that place these speciation events in the late Miocene ~6-9mya (Table 4 and Table 5).

Table 4. The estimated minimum and maximum time of divergence in millions of years between each species of *Gavia.* The divergence rate was based on molecular mutation rate range of 0.15-0.33% per million year. The molecular clock was calibrated from *G. concinna* (~4.8mya), *G. pacifica* (~2mya), *G. fortis* (~4mya), and *G. immer* (~4mya).

	G. stellata	*G. arctica*	*G. pacifica*	*G. immer*
G. stellata				
G. arctica	10-22.8			
G. pacifica	10-22.8	2.6-5.9		
G. immer	10-22.8	3.3-7.5	3.5-7.9	
G. adamsii	10-22.8	2.9-7.8	3.2-7.1	1.4-3.1

Table 5. The estimated time of divergence in millions of years between each species of *Gavia.* The divergence rate was based on a nuclear intron evolution rate of 0.12% (Lerner et al. 2011).

	G. stellata	*G. arctica*	*G. pacifica*	*G. immer*
G. stellata				
G. arctica	20			
G. pacifica	20	7.1		
G. immer	20	9.2	9.2	
G. adamsii	20	8.2	8.3	3.8

Table 6. The estimated time of divergence in millions of years between each species of *Gavia.* The divergence rate was based on a nuclear intron evolution rate of 0.36% (Axelsson et al. 2004).

	G. stellata	*G. arctica*	*G. pacifica*	*G. immer*
G. stellata				
G. arctica	6.7			
G. pacifica	6.7	2.4		
G. immer	6.7	3.1	3.1	
G. adamsii	6.7	2.7	2.8	1.3

DISCUSSION

Phylogenetic Analysis

The results of all three phylogenetic analyses (BI, MP, and ML) (Figures 6-8) are in agreement with the traditional *Gavia* phylogeny (*G. stellata*, ((*G. arctica*, *G. pacifica)*, (*G. adamsii* and *G. immer)))* (Boertmann 1990) and conflict with the DNA-based phylogeny that splits the *G. arctica* and *G. pacifica* clade (Lindsay 2002). The BI yielded better resolution and stronger support for intraspecific relationships (Figure 8) compared to the ML and MP phylogenies (Figures 6 and 7). Suzuki et al. (2002) demonstrated that the posterior probabilities in Bayesian analysis can be more generous while bootstrap probabilities can be stricter. The low support and lack of resolution of intraspecific relationships are most likely due to the samples coming from the same geographical region and not having enough genetic variation to provide a strong signal (Table 1). For example, *G. adamsii* populations found in Alaska had the lowest intraspecific p-distance. On the other hand, *G. arctica* samples from Sweden and Russian populations had the highest intraspecific p-distances, and had more resolved intraspecific relationship. Given that the BI, ML, and MP trees all had strong support (100%) it is likely that the interspecific relationships presented in the trees in this study are robust. Although the mtDNA phylogeny also had strong support (Lindsay 2002), phylogenies constructed from multiple loci are considered more robust when compared to a conflicting tree based on mtDNA (Zink and Barrowclough 2008).

The large number of loci collected in the present study, more than three orders of magnitude greater than the Lindsay (2002) dataset, aided in capturing more genetic variation and provided a more robust *Gavia* phylogenetic tree (Figures 6-8). The trees from the present study were constructed from genetic data from 2502 loci that contained 232,094 bps, whereas the

Lindsay (2002) dataset had ~5,000 bps from two linkage groups. The use of NGS to collect such a massive amount of data helps to provide better coverage of the genome (Figure 4) and to get a more accurate representation of the genetic variation among species.

Divergence Date Estimation

The estimated divergence dates of the five extant species of *Gavia* differed considerably from one another depending on how the molecular clock was calibrated (Table 4-6). The faster rates of 0.33-0.36% do not seem to be appropriate for *Gavia*, because they would put the estimated divergence dates earlier than previously thought, and would not be consistent with the fossil record (Olson and Rasmussen 2001, Emslie 1998, Brodkorb 1953). The faster evolving rates would also put the splitting of *G. stellata* from the four other species of *Gavia* in the late Miocene, 7-10mya, and the diversification of the other four species in the late Pliocene- early Pleistocene. In particular the estimated divergence of *G. immer* and *G. adamsii* would be ~1.3mya. The fossil records of *G. immer* from ~2mya (Emslie 1998) would make it impossible for *G. immer* and *G. adamsii* to have diverged ~1.3 mya (Table 6). Finally, the faster nuclear intron evolution rate of 0.36% (Axelsson et al. 2004) was based on an assumed divergence time estimated from a mtDNA–based molecular clock obtained from limited fossil data (Dimcheff et al. 2002), make it less reliable than the 0.12% (Lerner et al. 2011). The latter authors used the known ages of the different Hawaiian Islands, and ages of the fossils found on the islands to calculate the substitution rates in the Hawaiian honeycreeper.

Given that it is known that different genes can evolve at different rates, basing divergence times solely on a constant substitution rate can result in inaccurate dates and may lead to an overestimation of divergence dates (Burbrink and Pyron 2011, Pulquerio and Nichols 2006,

Wertheim and Sanderson 2010, Ayala 1999). However, increasing the sequence length and calibrating with a fossil from a known date can provide more precision (Wertheim and Sanderson 2010). Calibrating the molecular clock with *G. concinna* and *G. fortis* provides a molecular evolution rate of 0.15%, which aligns with the 0.12% nuclear intron rate (Lerner et al. 2011). Averaging the estimated dates from the two slower rates (0.15% and 12%) would then place the majority of the speciation in the Miocene (Tables 4 and 5). The early Miocene was a time of great diversification of Northern Hemisphere avifauna with new ecological niches being created as temperatures cooled, mountain ranges formed and Eurasia and North America became more separated (Blondel and Mourer-Chavlre 1998). At that time species of *Gavia* increased in body size and became more morphologically specialized as foot-propelled underwater divers (Olson and Rassmussen 2001, Boertman 1990). The *Gavia* lineage also started to increase in body size during the Miocene. However, *G. stellata* is smaller than the other four extant species of *Gavia* and *G. howardae* from the late Pliocene (~2mya) (a presumed ancestor of *G. stellata*) is also smaller than *G. concinna* (4.8 mya) and *G.fortis* (4 mya) and is not as robust as the three other extinct *Gavia* species from the Miocene (~10.5) (*G. moldavica*, *G. brodkorbi*, and *G. paradoxa*) (Olson and Rassmussen 2001, Boertman 1990, Brodkorb 1953). This would mean that if the *G.stellata/ G. howardae* lineage diverged from the other bigger- bodied species of *Gavia* ~7mya (Olson and Rassmusen 2001, Boertman 1990) the lineage became smaller after the split. The new estimated date of ~21.4 mya (Figure 9) provides a parsimonious explanation that the ancestral small body size of the *G. stellata*/*G. howardea* lineage is ancestral and that the larger body sized of the other four modern species of *Gavia* is derived.

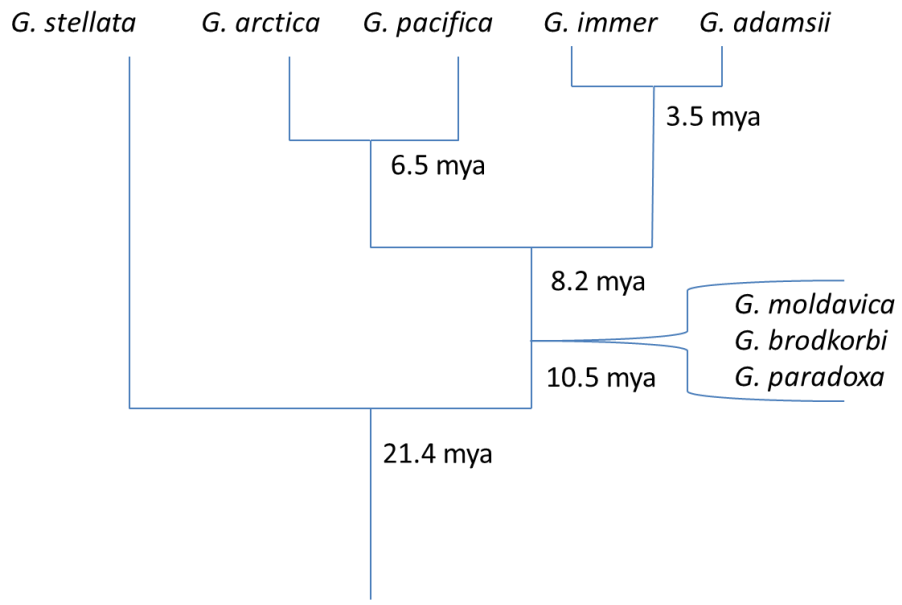

Figure 9. Estimated times of divergence of the five extant species of *Gavia*: *G. stellata, G. arctica, G. pacifica, G. immer,* and *G. adamsii.* The estimated dates were based on calibrating the molecular clock with *G. concinna* (4.8 mya) and *G. fortis* (4 mya) and an assumed 0.12% nuclear intron rate (Lerner et al. 2011). The estimated time of 10.5 mya for *G. moldavica*, *G. brodkorbi*, and *G. paradoxa* are based on the fossil record (Olson and Rassmussen 2001, Mlikosky 1994).

The small body size and low wing-loading (Boertman 1990) allow *G. stellata* to take advantage of fast thawing small lakes in the arctic as breeding habitat (Olson and Rassmussen 2001).The ability to utilize these lakes provides an earlier start to the breeding season and would exert a strong selection pressure to maintain a smaller size. However, *G. stellata'*s use of small shallow lakes with limited prey availability forces them to make daily foraging flights to nearby coastal regions or larger lakes (Bergman and Derksen 1977). On the other hand the bigger-bodied species of *Gavia* have to wait longer for the bigger lakes to thaw, but they can forage locally on their larger nesting lakes (Johnsgard 1987). Schreer et al. (2001) found that large body size allows divers to reach greater depths and to forage for longer. The ability to dive deep would create new foraging niches and would provide a selection pressure for bigger body size. The early species of *Gavia* that were getting bigger in body size could exploit this new feeding niche.

The smaller *G. stellata* mainly forges on pelagic and semi-pelagic fish, whereas the bigger bodied species (*G. pacifica*, *G. arctica*, *G. immer*, and *G. adamsii*) feed more on the benthic species of fish (Johnsgard 1987 and Boertman 1990). *Gavia immer* spends 82% of its foraging time at ~3.7-7.3m, compared to *G. stellata* (54% of time forgaing at ~1.8-3.7m) (Johnsgard 1987). The bigger species of *Gavia* can also stay underwater for longer than *G. stellata*. *Gavia immer* have been recorded foraging at up to ~11m for ~68 seconds, whereas *G. stellata* forages for ~48 seconds down to ~9m (Johnsgard 1987). The maximum dive time for *G. pacifica* is ~5min., for *G. immer*~3min., and for *G. stellata* ~1.5 min (Johnsgard 1987).

As the body size of *Gavia* increased, their requirements in lake size would also have increased. The bigger bodied loons have a high wing load (Boertman 1990) and require large lakes for flight takeoff. For example, *G. pacifica* needs a running start of 120-200 m to achieve takeoff, while *G. stellata* only needs a running start of 15-40 m to reach takeoff speed (Johnsgard 1987). The difference in lake requirements would have led to increasing isolation between the smaller and bigger bodied populations contributing to the species diverging.

As a lineage of birds, loons have survived though a mass extinction event, separating of continents, ice ages, and inter-glacier periods. All these forces have played a part in the evolutionary history of *Gavia*. The data from this study have helped to better understand the relationships between the extant species of *Gavia* and the ecological factors that might have played a role in speciation events.

CONCLUSION

This study is one of the first to use next-generation sequencing to create thousands of RAD-tags for phylogenetic analyses. Based on these results, future avian phylogenetic studies should feel confident in the use of both RAD-tags and a double-digest protocol. The data collected from this study provide strong support for the morphological tree that has five species placed into three clades: (*G. stellata*, ((*G. arctica*, *G. pacifica*), (*G. adamsii and G. immer*))). Other important findings indicated modern loons share a common ancestor from the early Miocene, and the *G. stellata*/ *G. howardae* lineage retains the ancestral state of small body size. The estimated dates were based on a molecular evolution rate of 0.13%. The estimated dates of divergence were based on calibrating the molecular clock with fossils of *G. concinna* (4.8 mya), G. fortis (4 mya*), G.pacifica* (2mya), and *G. immer* (2mya). Future discovery of additional *Gavia* fossils will further increase the precision of molecular clock calibration. Further investigations of the rate at which each lineage or species are evolving will provide more insight into the evolutionary history of *Gavia*.

LITERATURE CITED

Alström, P., Höhna, S., Gelang, M., Ericson, P. G., & Olsson, U. (2011). Non-monophyly and intricate morphological evolution within the avian family Cettiidae revealed by multilocus analysis of a taxonomically densely sampled dataset.BMC evolutionary biology, 11(1), 352.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Andersen, B. G., & Borns, H. W. (1994). The Ice Age world: an introduction to Quaternary history and research with emphasis on North America and northern Europe during the last 2.5 million years. A Scandinavian University Press Publication.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. New Biotechnology, 25:195-203.

Axelsson, E., Smith, N. G., Sundström, H., Berlin, S., & Ellegren, H. (2004). Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Molecular Biology and Evolution, 21(8), 1538-1547.

Ayala, F.J. 1999. Molecular clock mirages. BioEssays 21: 71–75.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., and Selker, E. U. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3(10), e3376.

Baker, G. M. (2002). Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. Biological Journal of the Linnean Society, 76:165-194.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. Molecular ecology, 13(4), 729-744.

Bergman, R. D., & Derksen, D. V. (1977). Observations on Arctic and red-throated loons at Storkersen Point, Alaska. Arctic, 30(1), 41-51.

Birch, A., and Lee, C. T. (1997). Arctic and Pacific loons: Field identification. Birding, 29, 106-115.

Boertmann, D. (1990). Phylogeny of the diver, family Gaviidae (Aves). Steenstrupia 16:21-36.

Brito, P. H., & Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica, 135(3), 439-455.

Brodkorb, P. (1953). A review of the Pliocene loons. The Condor, 55(4), 211-214.

Burbrink, F. T., & Pyron, R. A. (2011). The impact of gene-tree/species-tree discordance on diversification-rate estimation. Evolution, 65(7), 1851-1861.

DeGiorgio, M., & Degnan, J. H. (2014). Robustness to divergence time underestimation when inferring species trees from estimated gene trees. Systematic biology, 63(1), 66-82.

Dimcheff, D. E., Drovetski, S. V., & Mindell, D. P. (2002). Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. Molecular phylogenetics and evolution, 24(2), 203-215.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460-2461.

Edwards, S. V., Bryan Jennings, W., and Shedlock, A. M. (2005). Phylogenetics of modern birds in the era of genomics. Proceedings of the Royal Society B: Biological Sciences, 272(1567), 979-992.

Ericson, P. G., Jansén, A. L., Johansson, U. S., & Ekman, J. (2005). Inter-generic relationships of the crows, jays, magpies and allied groups (Aves: Corvidae) based on nucleotide sequence data. Journal of Avian Biology, 36(3), 222-234.

Emslie, S. D. (1998). Avian community, climate, and sea-level changes in the Plio-Pleistocene of the Florida Peninsula. Ornithological Monographs, 1-113.

Fain, M. G., & Houde, P. (2004). Parallel radiations in the primary clades of birds. Evolution, 58(11), 2558-2573.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology, 59(3), 307-321.

Hackett, S. J., et al. (2008). A phylogenomic study of birds reveals their evolutionary history. science, 320(5884), 1763-1768.

Härlid, A., Janke, A., & Arnason, U. (1997). The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Molecular Biology and Evolution, 14(7), 754-761.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PloS Genetics, 6(2), e100862.

Hohenlohe, P. A., Amish, S. J., Catchen, J.M., Allendorf, F.W., and Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources, 11:117-122.

Jacobsen, F., & Omland, K. E. (2011). Species tree inference in a recent radiation of orioles (Genus I*cterus*): Multiple markers and methods reveal cytonuclear discordance in the northern oriole group. Molecular phylogenetics and evolution, 61(2), 460-469.

Johnsgard, P. A. (1987). Diving Birds of North America: Species Accounts--Loons (Gaviidae). Diving Birds of North America, by Paul Johnsgard, 9.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology, 56(1), 17-24.

Kubatko, L. S., Gibbs, H. L., & Bloomquist, E. W. (2011). Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in Sistrurus rattlesnakes. Systematic Biology, 60(4), 393-409.

Lerner, H. R., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology, 21(21), 1838-1844.

Lindsay, A. R. (2002). Molecular and Vocal Evolution in Loons (Aves:Gaviiformes) Doctoral dissertation, University of Michigan.

Liu, L., & Yu, L. (2010). Phybase: an R package for species tree analysis. Bioinformatics, 26(7), 962-963.

Livezey, B. C. (1988). Morphometrics of flightlessness in the Alcidae. The Auk, 681-698.

Livezey, B. C., and Zusi, R. L. (2007). Higher-order phylogeny of mondern birds (Theropoda, Aves:Neornithes) based on comparative anatomy. II. Analysis and discussion. Zoological Journal of the Linnean Society, 149:1-95.

Lovette, I. J., and Hochachka, W. M. (2006). Simultaneous effects of phylogenetic niche conservatism and completion on avian community structure. Ecology, 87:S14-S28.

Maddison, W. P., & Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. Systematic biology, 55(1), 21-30.

Mayr, G. (2004). A partial skeleton of a new fossil loon (Aves, Gaviiformes) from the early Oligocene of Germany with preserved stomach content. Journal of Ornithology, 145: 281-286.

Mayr, G., and Poschmann, M. (2009). A loon leg (Aves, Gaviidae) with crocodilian tooth from the late Oligocene of Germany. Waterbirds, 32(3), 468-471.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. (2011). Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution:1-13.

McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One, 8(1), e54848.

Mlikovsky, J. (1998). A new loon (Aves: Gaviidae) from the middle Miocene of Austria. Ann Natur Mus Wien, 99, 331-339.

Monroe Jr, B. L., Banks, R. C., Fitzpatrick, J. W., Howell, T. R., Johnson, N. K., Ouellet, H., Remsen, J.V., and Storer, R. W. (1985). Thirty-fifth supplement to the American Ornithologists' Union Check-list of North American birds. The Auk, 680-686.

Morin, P. A., Luikart, G., and Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution, 19: 208-216.

Olson, S. L. (1992). Neogaeornis wetzeli Lambrecht, a cretaceous loon from Chile (Aves: Gaviidae). Journal of Vertebrate Paleontology, 12(1), 122-124.

Olson, S. L., and Rasmussen, P. C. (2001). Miocene and Pliocene Birds from the Lee Creek Mine; North Carolina. Smithsonian Contributions to Paleobiology, (90).

Prychitko, T. M., & Moore, W. S. (1997). The utility of DNA sequences of an intron from the β-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Molecular phylogenetics and evolution, 8(2), 193-204.

Pulgarín-R, P. C., Smith, B. T., Bryson Jr, R. W., Spellman, G. M., & Klicka, J. (2013). Multilocus phylogeny and biogeography of the New World *Pheucticus* grosbeaks (Aves: Cardinalidae). Molecular phylogenetics and evolution, 69(3), 1222-1227.

Posada, D., & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. Bioinformatics, 14(9), 817-818.

Pulquerio, M. J., & Nichols, R. A. (2007). Dates from the molecular clock: how wrong can we be?. Trends in Ecology & Evolution, 22(4), 180-184.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12), 1572-1574.

Roselaar, C., Prins, T. G., Aliabadian, M., and Nijman, V. (2006). Hybrids in divers (Gaviiformes). Journal of Ornithology, 147:24-30.

Rosenberg, N. A., & Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nature Reviews Genetics, 3(5), 380-390.

Groth, J. G., & Barrowclough, G. F. (1999). Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Molecular phylogenetics and evolution, 12(2), 115-123.

Rubin, B. E., Ree, R. H., and Moreau, C. S. (2012). Inferring Phlyogenies from RAD Sequence Data. PLoS ONE, 7(4).

Russell, R. W., & Lehman, P. E. (1994). Spring migration of Pacific Loons through the Southern California Bight: nearshore flights, seasonal timing and distribution at sea. Condor, 96(2), 300-315.

Salemi, M., Vandamme, A. M., & Lemey, P. (Eds.). (2009). The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press.

Sjölander, S., & Ågren, G. (1976). Reproductive behavior of the Yellow-billed Loon, Gavia adamsii. The Condor, 78(4), 454-463.

Shapiro, M. D., Kronenberg, Z., Li, C., Domyan, E. T., Pan, H., Campbell, M., ... & Wang, J. (2013). Genomic diversity and evolution of the head crest in the rock pigeon. Science, 339(6123), 1063-1067.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26:1135-1145.

Silvestro, D., & Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. Organisms Diversity & Evolution, 12(4), 335-337.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21), 2688-2690.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. Systematic biology, 57(5), 758-771.

Suzuki, Y., Glazko, G. V., & Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proceedings of the National Academy of Sciences, 99(25), 16138-16143.

Swofford, D. L. (2002). PAUP* version 4.0. Phylogenetic analysis using parsimony (and other methods).

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular ecology, 22(3), 787-798.

Wertheim, J. O., & Sanderson, M. J. (2011). Estimating diversification rates: how useful are divergence times?. Evolution, 65(2), 309-320.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology & Evolution, 11(9), 367-372.

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology, 17(9), 2107-2121.

APPENDIX A

Table A1. Sequences of adapters and PCR primers used in RAD-tag library preparation

Name	Sequence
LCAT.2F	5’-GTGGTGAACTGGATGTGCTACCG-3’
LCAT.5R	5’-GCACCCAGNGAGATGAAGCC-3’
P1 Adapter Top	5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxTGCA-3’
P1 Adapter Bottom	3’-TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAxxxxxx-Phos-5’
P2 Adapter Top	5’-Phos AATTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCAGAACAA-3’
P2 Adapter Bottom	3’-TCTAGCCTTCTCGCCAAGTCGTCCTTACGGCTCTGGCTAGAGCATACGGCAGAAGACGAAC-5’
RH300.F	5’-TGAGTAACTTGGGGCCACATC-3’
RH300.R	5’-TGATTGCGCTACCTTTGCAC-3’
RH450.F	5’-CACAAGATGCACCTAAACACACC-3’
RH450.R	5’-CTGCTAAATCCGCCTTCCAG-3’
RAD1.F*	5’-AATGATACGGCGACCACCGAG-3’

*On each adapter the xxxxxx = barcode

APPENDIX B

Table 1B. Mean estimates of parameters for the concatenated data set. Substitution rate parameters are calculated with GTR+i+Γ substitution model. For the ML analysis the rate for G <–> T set to 1.

Program	Analaysis	Mean lnL	piA	piC	piG	piT	rA-C	rA-G	rA-T	rC-G	rC-T	rG-T	pinVar	gamma
Mr. Bayes	BI 1	-436525.435	26.7	23.3	23.6	26.4	0.069	0.374	0.045	0.081	0.363	0.067	0.924	0.669
Mr. Bayes	BI 2	-43625.756	26.7	23.3	23.6	26.4	0.069	0.374	0.045	0.081	0.363	0.067	0.924	0.666
RAxML	ML	-436612.295	26.7	23.3	23.6	26.4	1.025	5.506	0.658	1.198	5.34	1	0.923	0.658

APPENDIX C

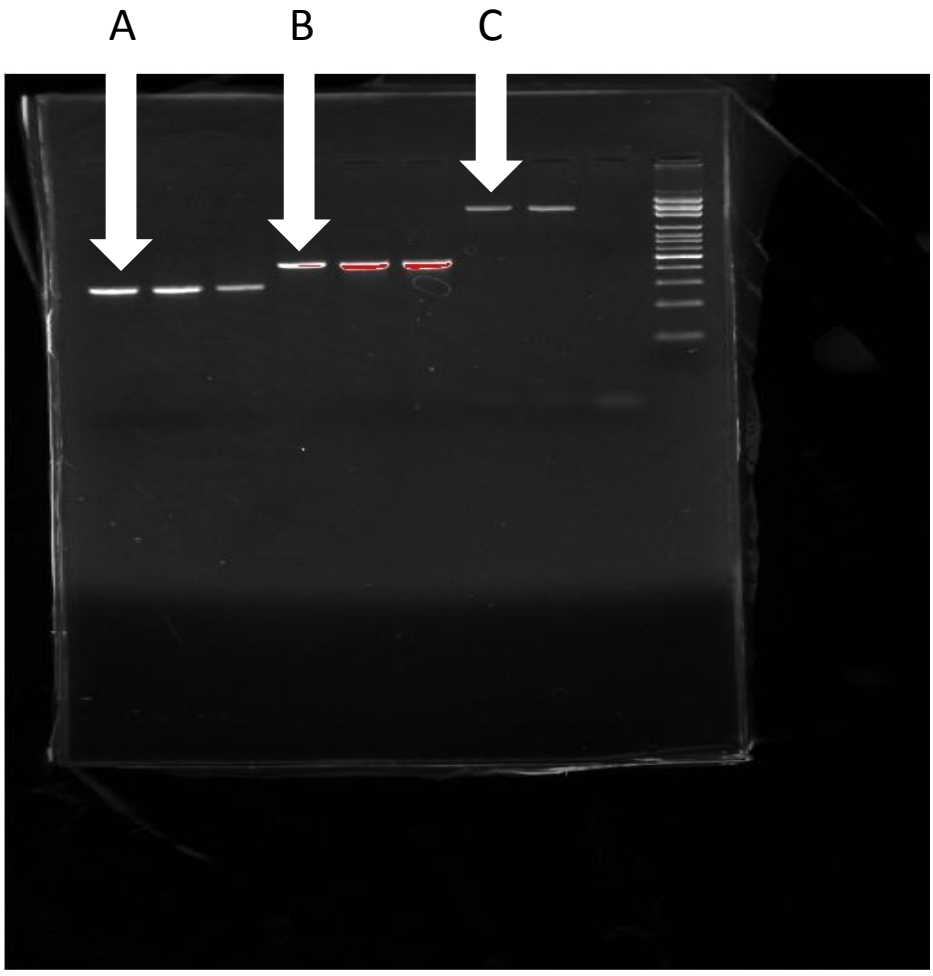

Figure C1. LCAT Gel image. “A” points to the 300bps size standard.
“B” points to the 450 bps size standards. “C” points to the LCAT fragment.

APPENDIX D

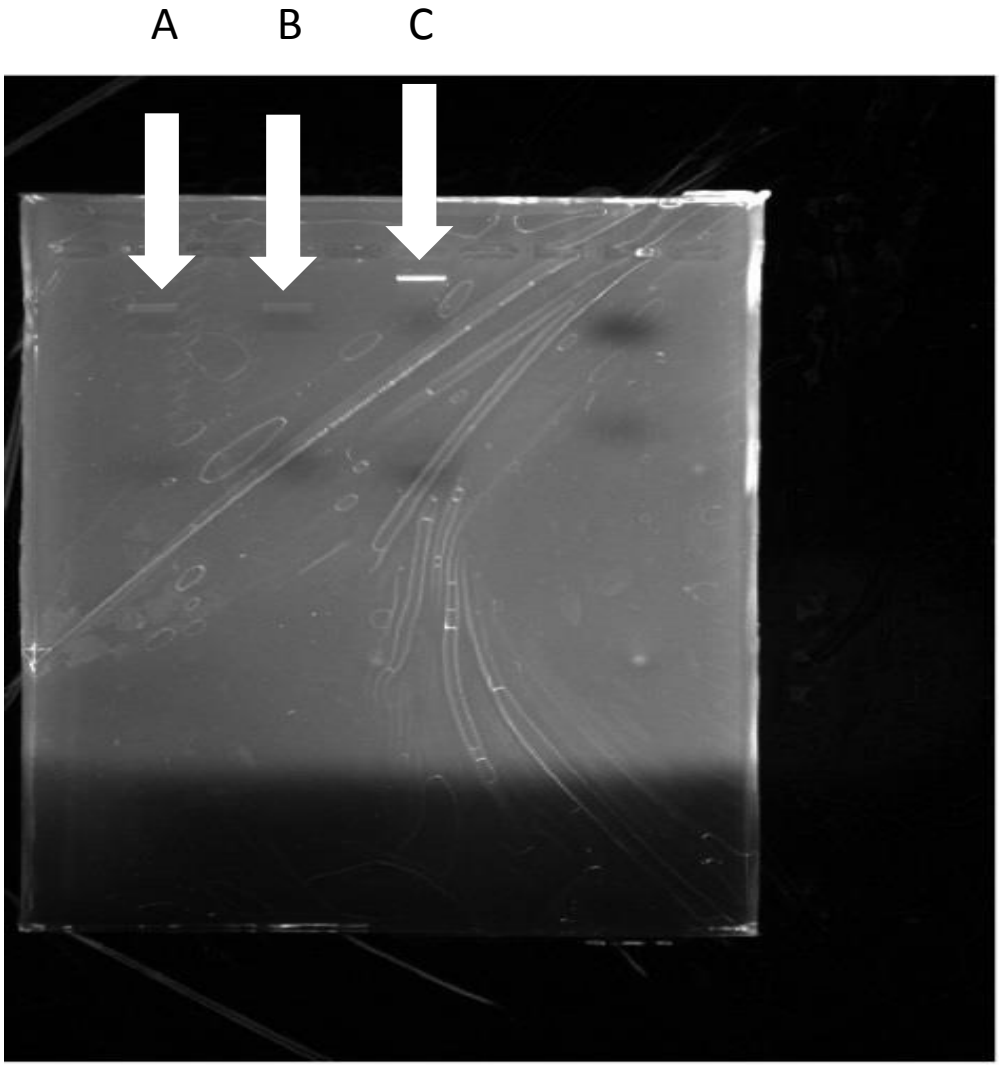

Figure 1D.Check gel image from double-digest LCAT. “A” and “B” both point to the bands of digested LCAT positive control. “C” points to negative control non-digested LCAT.

APPENDIX E

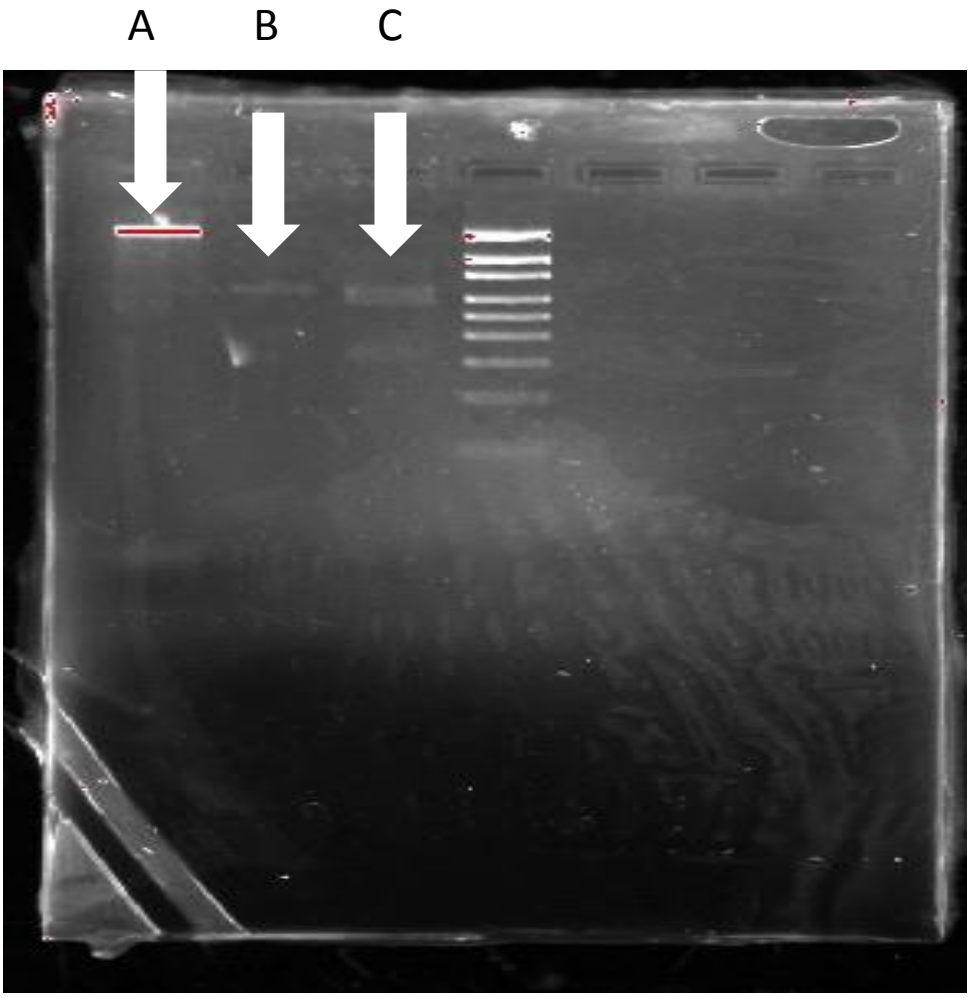

Figure 1E.Check gel image of adapter-ligated LCAT. “A” points to negative LCAT that was not digested or adaptor-ligated. “B” points to digested but non-ligated LCAT. “C” points to digested and adapter-ligated LCAT control.

APPENDIX F

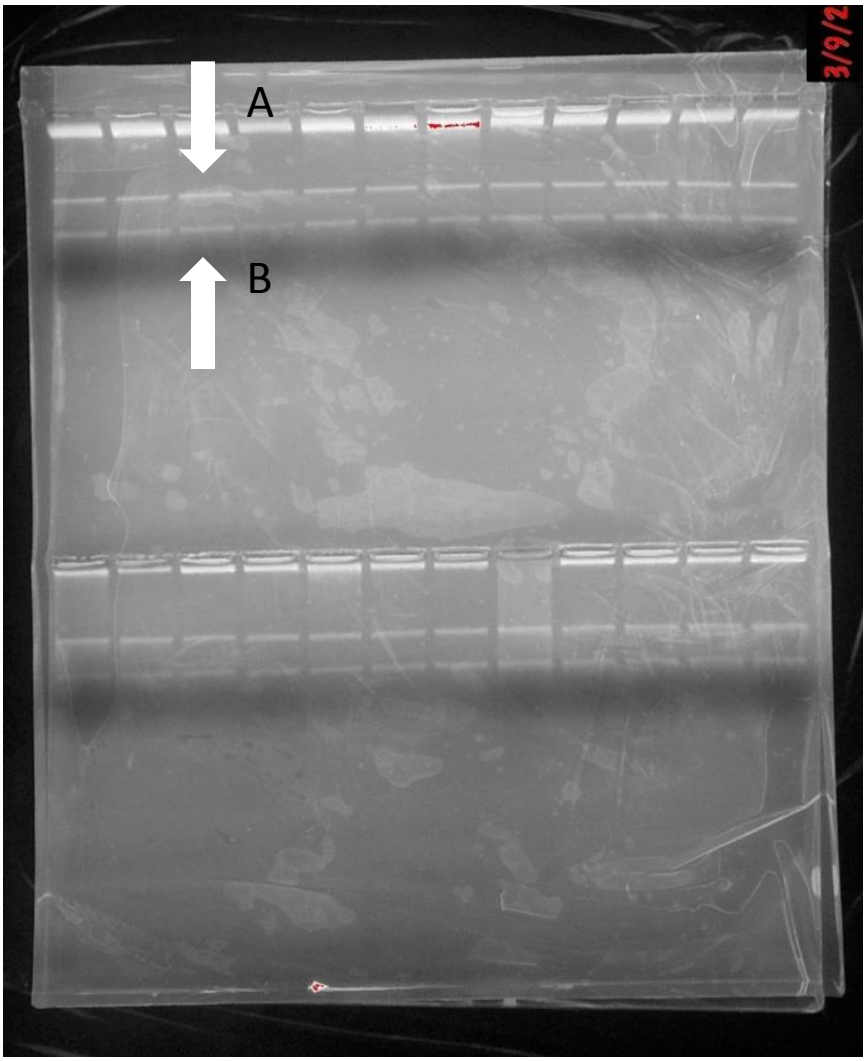

Figure 1F. Example gel image of the digested and adapter-ligated samples. “A” points to the band of the 450 bps internal size standard. “B” points to the band of the 300 bps internal size standard. A wedge cut was performed by cutting on the internal edge of each size standard, i.e. just below 450 bps and just above 300 bps. The entire width of the 450 bps was cut but only half of the 300 bps was cut.

APPENDIX G

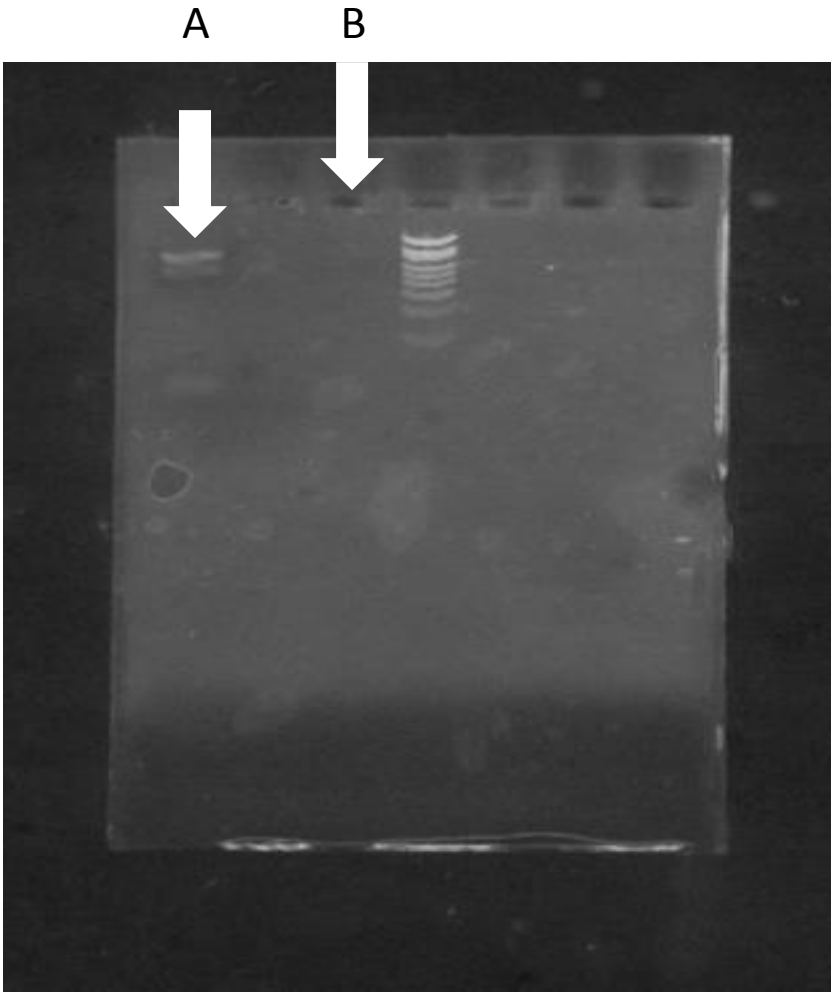

Figure 1G. Check gel image of amplified RAD-Tag LCAT. “A” points to the band of the amplified LCAT. “B” points to negative control (water).