PHYLOGENETIC RELATIONSHIPS WITHIN SYLVIOIDEA SUPERFAMILY

Doctor of Philosophy
Thesis

SCIENTIFIC COORDINATORS:
PROF. DR. GOGU GHIORGHIȚĂ
PROF. DR. DUMITRU COJOCARU

PhD STUDENT:
MITICĂ CIORPAC

IAȘI
-2015-
Introduction

The taxonomy, initially, developed the animals classification, based on the similarities and differences in the degree of relationships between them. The large number of species and the high morphological diversity, led sometime to wrong taxonomies and debates between zoologists.

The birds (Class Aves) have approximately 10,000 contemporary species, being considered the class of tetrapods vertebrate, with the greatest diversity. All living species belong to a single subclass Neornithes occupying areas all around the globe, ranging from Antarctica in the South to the Arctic Ocean in the North. Class Aves consists of two subclasses, comprising 29 Orders, a great number of species with a high level of morphological diversity, representing a drag in the taxonomic organization.

The largest order of class Aves, Passeriformes includes about 5400 species (Mayr, 1946), about half of all bird species, making it one of the most diverse terrestrial vertebrate orders. This order is organized into three suborders and about 100 families, where the largest group is the Sylvioidea superfamily.

Superfamily Sylvioidea includes reed warblers and Old World babbler, swallows, larks, etc., counting numbering about 1,300 species. Mainly this The large number of species has determined countless taxonomic uncertainties, thus being a common subject for previous studies. The superfamily Sylvioidea was proposed in the first time by Sibley and Ahlquist (1990), Sibley and Monroe (1990), based on DNA-DNA hybridization studies, which were based on measuring the consonance of the genetic material. Further studies have supported the inclusion of some families in this taxon, but denied the larks membership (Alstrom et al., 2006). Some families from this superfamily have been redefined, in particular Old World reed warblers, family Sylviidae and Timaliidae. Since quite often they have been used as a container taxa, which included a large number of species, a fact that has proved that they were not so closely related. Several new families were created, and some species have been moved from one family to another (Boyd, 2010).

Superfamily Sylvioidea is currently made up of 25 families, divided into 7 groups describing separate clades.
Chapter 1. Introduction to Sylvioidea superfamily

Passeriformes order

It is organized into three suborders: Passeri (Oscines, true songbirds), Tyranni (Suboscines) with Acanthisitti forming a common group (Figure 1). The Acanthisitti suborder, endemic to New Zealand, consists in a single family (Acanthisittidae) with two living species, grouped into two different genera and another three extinct genera (Fregin et al., 2012).

Oscines or Passeri, it’s a unique group, characterized by a pronounced morphological uniformity (Fregin, 2013), which makes it difficult to define subgroups based on phenotypic characters that evolved by convergence (Feduccia 1996). Passerida was divided into three main superfamilies by Sibley and Ahlquist (1990): Muscicapoidea, Sylvioidea and Passeroidea.

Sylvioidea Superfamily

The Sylvioidea superfamily (Sibley și Ahlquist, 1990) includes reed warblers and Old World babbler, swallows, larks, etc., numbering about 1,300 species. This taxa currently contains 25 families clustered in 7 main clades (Figure 2): 1- Stenostiridae, Paridae and Remizidae; 2- Nicatoridae, Panuridae and Alaudidae; 3- Macrosphenidae; 4- Acrocephalus group (6 families); 5- Hirundinidae / Pycnonotidae; 6- Hyliiidae, Aegithalidae, Cettiidae and Phylloscopidae; 7- babblers group, (Boyd, 2013).
The purpose and objectives of the thesis

The purpose of the thesis

Through this study, we aimed to identify the phylogenetic relationships within Sylvioidea superfamily, using genetic markers: mitochondrial DNA (mtDNA) nuclear DNA (ADNn) and microsatellite (SSR). Phylogenetic analyzes had as a main target, the removing of existing taxonomic uncertainties within the superfamily Sylvioidea. Further more, we estimate the divergent time of the major lineages of sylvioids in order to clarify their evolutionary history. Fundamental evolutionary processes were identified and correlated, like gene flow, demographic and spatial expansion, speciation and interspecific relationships, representing decision markers in taxonomic delimitation of the species concept. They were identified and correlated by fundamental evolutionary processes, like gene flow, demographic and spatial expansion, speciation and interspecific relationships, representing decision makers in taxonomic delimitation of the species concept.

Research objectives

- The optimal combination of molecular markers enables the description of a better song to a better time scaled of the song birds phylogeny.
Chapter 2. Material and Methods

2.1 Sampling

The biological material used in this study is represented by a total of 100 individuals within the superfamily Sylvioidea, captured in Larga Jijia area (Iasi county) during ring sessions.

The biological material consists of blood samples from the brachial vein, preserved in Queen’s Lysis Buffer (Seutin et al., 1991) The blood samples were stored at -80°C until DNA extraction.

2.2 DNA Extraction and PCR Amplification

The isolated and purified DNA samples will be used in the polymerase chain reaction (PCR), whereby a sequence of the genetic information to be studied is multiplied exponentially. There were amplified and sequenced DNA strands from two mitochondrial genes (cytochrome b and cytochrome oxidase subunit I), a nuclear gene (CHD Cromo DNA helicase - used to determine the sex) and three nuclear microsatellite for the analyzed species.

2.3 Sequence Analysis

The sequence alignment of each phylogenetic marker was performed using MEGA 6 software (Tamura et al., 2013).

Further, the sequence alignments were used to the population and phylogeographical analysis.

Chapter 3. Markers section for passerines multilocus phylogeny

In this chapter we wanted to test the phylogenetic performance of five nuclear markers (MB, RAG1, GAPDH, ODC1, CHD-Z) and a mitochondrial one (cytb).
3.1 **Mutation mean rate dynamics**

The mean mutation rates and the variation range shift of the nuclear genes, describes this markers like additional phylogenetic tools for birds phylogeny. Therefore, the individual use of nuclear and mitochondrial markers is insufficient to describe a correct and stable topology of the phylogenetic relationships within songbirds. The performance of this genes in the single locus phylogeny is inappropriate and their contribution to the different tree topologies in multi locus phylogeny is unknown. The lack of accuracy of these markers is determined mainly by the short length sequences and by the mutation rates.

3.2 **The accuracy of individual genes in estimating divergence time**

For a better view of the individual genes estimation performance, we checked the correlation between the divergence time estimated from the individual genes and from the concatenated analysis. So, the correlation between the estimated divergence times split the analyzed markers in two groups. The first one consist in genes with a good estimation of recent MRCA and the second one that estimates more precisely the age of older MRCA.

3.3 **Optimal combination of markers to a better fit phylogeny**

In this section, we aim to identify the optimal combination of the minimum number of markers with the best phylogenetic model fit.

Since, *cytb* gene, imposes the best model for topology and for divergence time, we tested the performance of each nuclear gene combined with *cytb* gene. The tree topologies generally remain the same like concatenated tree with some exceptions: *cytb* + MB, *cytb* + GAPDH and *cytb* + ODC1 which places the Petroicidae family in Corvida clade. In addition, *cytb* + ODC1, also places the Sparrow basal in the Eupasseri clade and the Paroidea family in a sister clade of Sylvioidea. The Models comparison of each nuclear gene combined with *cytb* gene show the optimal combination of markers in song birds phylogeny.

Therefore, the optimal combination of markers in song birds phylogeny is given by *cytb* + CHD-Z. In this case, mitochondrial marker, *cytb* gene, offer the time frame and a good node heights and CHD-Z gene acts more like a topology correction factor.
Chapter 4. The phylogenetic reconstruction of the Sylvioidea superfamily evolutionary history

Clarification of the taxonomic uncertainty is the main objective of this study, therefore they could be removed only by the construction of a phylogenetic super-tree (SA) to include all species components (an ideal case) or a number as close as possible.

The data set (sequences) used in the phylogenetic analysis consisted in mitochondrial marker sequence, cytochrome b gene (cytb) from 1918 individuals of 25 families, most sequences being downloaded from the international database GenBank. Taxons selection and data set construction, aimed to include all species available from a genus and at least 5 sequences from each species. In those cases when some species had subspecies, there were included at least three sequences per subspecies.

4.1 Sylvioidea superfamily phylogenetic SuperTree

Therefor according to the main objective, removing existing taxonomic uncertainties within the superfamily Sylvioidea, the obtained phylogenetic super-tree is able to clear the majority of taxonomic uncertainty. As a result, the supertree topology supports numerous previous studies and contradict punctually certain assumptions.

4.1.1 Superfamily Sylviidae: position and components

Superfamily Sylviidae is placed in the suborder Passeri, in the Passerida clade, alongside the Muscicapoidea superfamily, in accordance with previous studies (Ericson et al., 2014; Fregin et al., 2012; Jønsson și Fjeldså, 2006). Regarding the superfamily Sylvioidea composition, 24 families are clustered together in a main clade, (Figure 3), divided into 8 groups which describe separate subclades.
The Stenostiridae, Remizidae and Paridae families inclusion in Sylvioidea superfamily raised numerous debates in the last years about this taxon monophyly. The Fergin et. al. (2012) and Jønsson și Fjeldså (2006) studies reject the monophyly hypothesis describing a phylogeny with the Stenostiridae, Remizidae and Paridae families clustering in a basal clade alongside Reguloidea/Passeroidea clade.

Monophyly hypothesis is supported by numerous studies, and still none of them strongly confirmed their membership, nor reject this hypothesis with solid studies (Alström et al., 2006; Barker et al., 2002; Barker et al., 2004; Beresford et al., 2005; Ericson et al., 2014; Sheldon și Ellegren, 1996).
Phylogenetic reconstruction carried out in this paper confirms the superfamily Sylvioidea monophyly, placing basal the Stenostiridae family, followed by two subclades, one for Paridae and Remizidae families and the other one for Sylvioidea superfamily "sensu stricto".

### 4.2 Sylvioids Major Lineages Diversification

Zoogeographical, palaeontological and biochemical data support the origin hypothesis of passerine birds in the southern hemisphere, also, the molecular data suggest that the majority of the living bird orders were formed in Late to Middle Cretaceous (Ericson et al., 2002). Further we will present both phylogenetic tree construction method and the time scale calibration points, as well as the evolutionary history of the Passeriformes order and Sylvioidea superfamily.

#### 4.2.1 Timescale Calibration Points

The Phylogenetic supertree timescale calibration was made using multiple calibration points. We use four points to ensure the tree time scale: an vicariance event and three fosils calibration (Figure 4).

![Figure 4 - Calibration points for time scale models used in the superfamily Sylvioidea chronogram.](image)
4.2.2 Life history of Sylvioidea superfamily

The cytochrome b gene analysis has shown a monophyletic relationship within the Passeri suborder. The MCC tree has a strong topology and a divergent genealogy.

Oscines spread around the globe from Gondwana to Asia via Australia hypothesis, which is widely accepted in the evolution of this suborder. This hypothesis is justified by limiting the basal oscine lineages (Menuroidea and Meliphagoidea) and the basal members of Corvoidea in the Australo-Papuan region (Ericson et al., 2002) and also by the large number of species confined in Indochina peninsula.

The life history of Sylvioidea superfamily has started most likely in South Asia around 51.5 ma (Figure 5). A rapid diversification is determined by the Eocene Optimum global warming. The families ancestors spread through the Eurasia, Africa, Australia and Pacific Islands.
Figura 5 – Major lineages diversification of superfamily Sylvioidea. The MCC tree genealogy resulted from a Yule Speciation Process simulated in BEAST v.1.8.2 software using four calibration points.
5.1 Phylogenetic relationships within family Acrocephalidae

The BI tree obtained (Figure 6) describes a typology of phylogenetic relationships in accordance with previous studies, being made up of two superclades: one for large body size species - subgenus *Acrocephalus* (A) and a second superclade with *Acrocephalus* small body size species and basal genera like *Iduna*, *Nesillas*, *Hippolais* and *Calamonastides* (B).

![Figure 6 – The MCC tree of family Acrocephalidae, the tree genealogy results from a Yule Speciation Process simulated in BEAST v 1.8.2 software for 304 individuals over three MCMC chains at 50 millions states. *C* – Calamonastides; **I** – Phragmaticola](image)

5.1.1 Charting the colonisation pattern of the family Acrocephalidae across the Old World

The Acrocephalidae family dating time is not entirely estimated and well accepted. Both, previous studies and this paper, show a passerine evolution, triggered in Gondwana supercontinent (Cracraft, 2001). Therefore, we suggest that the Acrocephalidae family ancestor was formed
in Paleogene, around 42 and 34 million years ago, most likely across Africa.

The family diversification and genera formation (Figure 7) occurred after a mass extinction event from Eocene-Oligocene boundary (~ 34 Ma), followed by a slower divergence process, colonization and new environment adaptations.

![Figure 7 - Major lineages diversification of family Acrocephalidae. The MCC tree genealogy resulted from a Yule Speciation Process simulated in BEAST v.1.8.2 software using two calibration points.](image)

**5.2 Modeling the Temporal Evolution A. scirpaceus Species**

**5.2.1 Population Structure and Gene Flow**

*Acrocephalus scirpaceus* complex population structure assessment is one of the prime factors in shaping the temporal evolution, playing a major role in the experimental design of modeling the populations demographic evolution.

The genetic structure presented in Figure 8 shows the belonging probability of an individual to a defined cluster in Structure software. Non-random association of the individuals from a specific subspecies to a cluster denotes a boundary at the genetic level between these subspecies. Although there is an exception for *A. s. avicenniae* and *A. s. baeticatus*, which showed a population structure in mosaic.
5.2.2 Temporal evolution modeling

The populations temporal evolution modeling was performed using 115 cytochrome b gene partial sequences (879bp) from 5 subspecies of *Acrocephalus scirpaceus*. The analysis of mitochondrial gene cytochrome b showed a monophyletic relationship among four subspecies of A. scirpaceus. It also showed a strong topology of the MCC tree and a divergent lineage of individuals of each subspecies. Meanwhile, we see an exception, the hallae farm of subspecies A. s. baeticatus is placed basal in the A. s. baeticatus guiers and A. s. ambiguus clade (Figure 9). The explanation for this unusual arrangement could be the result of a speciation process updating the A. s. baeticatus hallae status to a full subspecies of the complex. The MCC tree also showed that the subspecies
A. s. avicenniae is closer to A. c. fuscus than to the A. s. scirpaceus, as previously thought. While, A. s. fuscus appears to be the earliest subspecies from the A. scirpaceus complex.

The estimated time of divergence for the A. scirpaceus complex showed its evolution in the last 7.7 million years, phylogenetic relationships within the complex being described by the MCC tree.

Data provided by the Bayesian inference carried out in the BEAST program, by simulating two MCMC chains with a length of 10 million iterations, indicating an evolution of this complex into two main phases: complex diversification as an adaptation to the geo-climatic conditions in the new occupied areas and the demographic expansion in the last 1.8 million years (Figure 10).
The *A. scirpaceus* complex life history shows a strong genetic differentiation of the five genetic lineages, originally formed by splitting the ancestral population and occupation of new areas, a spatial expansion without a demographic expansion. Evolution and diversification of the five actual genetic lineages is a consequence of geographic isolation in five refuges during Pleistocene glaciation: in Southwest Asia - *fuscus*; Europe - *scirpaceus*; Iberian Peninsula and North Africa - *ambiguus*; Africa - *baeticatus* and the Arabian Peninsula and Northeastern Africa - *avicenniae*.

5.3 **Molecular confirmation of a Great Reed Warbler X Reed Warbler hybrid (*Acrocephalus arundinaceus* X *A. scirpaceus*) in northeastern Romania**

There has been captured a male individual, in a ring session organized in Larga Jijia, one of the most important areas in Romania for passerine rest during the migration, that presents similarities between *A. scirpaceus* and *A. arundinaceus*. Morphological measurements performed, made us suspect that this individual could be the hybridization result between these two species (Ion et al., 2012).
5.3.1 THE FEMALE GENITOR IDENTIFICATION

In order to identify the female genitor we amplified and sequenced (in PCR and sequencing condition described in Chapter 2) two mitochondrial genes: COX I (600bp) and cytb (1040bp).

Amplicons of the two species and the hybrids were then sequenced, and correctly aligned. Sequences from assumed parental species and from the hybrid were used to infer the similarity and divergence matrix. Thus, they were further used for the congruent phylogenetic tree shown in Figure 11, made by the CustalW method, using Megalign module from the Lasergene suites.

![Figure 11 – Congruente phylogenetic tree based on COX I and cytb gene sequences from parental species and hybrid calculated by CustalW method, using Lasergene suites.](image)

5.3.2 THE MALE GENITOR IDENTIFICATION

Following the microsatellites HrU6, Aar3 and FhU2 amplification (in conditions described in Chapter 2), we found that the parents paternal species is A. scirpaceus.

According to the maternal inheritance of mitochondrial DNA, 100 % identity for 1,063 bp CytB and 93.5 % for 416 bp COI, indicates that the A. arundinaceus is the maternal genitor. The microsatellite analysis confirms the hypothesis of A. arundinaceus X A. scirpaceus hybridization, revealing an intermediate genotype between the two analyzed species (Figure 12), with multiple common loci.

The molecular analysis confirmed that the bird captured in September 2010 in Northeastern Romania, is a Great Reed Warbler X Reed Warbler hybrid. The mother was a Great Reed Warbler and the father was a Reed Warbler.
Conclusions

From the conducted researches on phylogenetic relationships of the superfamily Sylvioidea, were drawn the following conclusions.

The molecular markers performance to illustrate the phylogenetic relationships within passerines concluded as follows:

1. The nuclear genes are generally inconsistent phylogenetic markers usually sensitive to topological constraints. The low mutation rate and its instability on MCMC chain length induces a loss in the phylogenetic tree resolution and topologies with reduced support.

2. The individual use of nuclear and mitochondrial markers is insufficient to describe a correct and stable topology of the phylogenetic relationships within songbirds. The lack of accuracy of these markers is determined mainly by the short length sequences and by the mutation rates.

3. None of the analyzed genes can be considered able to accurately describe both phylogenetic relationships and the divergence time. Commonly, each of these markers, when are used alone, give topologies with reduced support and a wrong estimation of the nodes age.
4. The optimal combination of markers, in song birds phylogeny, is given by cytb + CHD-Z. In this case, mitochondrial marker, cytb gene offers the time frame and good node heights while the CHD-Z gene acts more like a topology correction factor.

The Sylvioida superfamily taxonomic reconstruction has enabled the following conclusions:

5. The Sylvioida phylogenetic super tree presents high-resolution clades, being able to clarify the existing taxonomic uncertainty within superfamily lineages.

6. Sylvioida superfamily is a monofiletic taxa even after “Paroidea” inclusion. In our phylogeny, the Stenostiridae family is a basal member followed by a common node for the Remizidae/Paridae clade and Sylvioida “sensu stricto”.

The evolutionary history, phylogenetic relationships and phylogeographic reconstruction of the family Acrocephalidae have drawn the following conclusions:

7. Phylogenetic relationships within the family Acrocephalidae do not present taxonomic uncertainty regarding genera composition, the current taxonomy being clarified by previous studies.

8. In accordance with previous studies, the genus Acrocephalus is not monophyletic, as species are grouped into three well-defined clades, interspersed with Iduna and Hippolais genera, without a well-defined basal species.

9. The BI tree fossilis calibration has enabled us to estimate the family age (~33,5mya).

10. The geographical origin of the genus Acrocephalus is the African continent, the genera evolution being imposed by the continental drift and climate change, at the same time represents opportunities for spatial expansion and also a condition of evolution, adaptation measurements to new areas.

11. The two main groups of warbler, the large body and the small body size species, had various ways of evolution and
colonization: the large ones originated in Africa and the small ones are rooted in South Asian continent.

12. Diversification of reed-warblers across the Pacific islands followed two phylogeographic pathways: the first one from Indochina Peninsula to Hawaii, through Mariana Islands, and the second one from Indochina Peninsula to the Pitcairn Islands through New Guinea, Kiribati, Solomon Islands and Samoa.

The temporal evolution modeling of the supraspecific complex *Acrocephalus scirpaceus* has enabled the following conclusions:

13. Bayesian inference of the divergence time showed the evolution of *A. scirpaceus* complex in the last 7.7 million years, starting from an ancestral population across central Asia.

14. The evolution of this complex was carried out in two major phases: diversification as an adaptation to the geo-climatic conditions in the new occupied areas and a demographic expansion in the last 2 million years.

15. The low number of descending lineages from the complex appearance until three million years ago is a consequence of the evolution starting from a small population. The demographic expansion is strongly impeded by natural selection (so, by the rate of adaptation to the environment) and the fragmentation of this group in small populations, that have colonize new geographical areas (representing the current subspecies ancestors).

16. The process of gene introgression was present in all subspecies populations, having as a main source, *A. s. scirpaceus* subspecies, as a consequence of the wintering areas overlapping.

17. The initial gene flow and geo-climatic factors pushed to a different evolution for each subspecies.

18. Quantification of intra and interspecific genetic diversity and phylogenetic relationships support the hypothesis that identify current taxonomic status of supraspecific complex, , removing the pre-existing taxonomic uncertainty.
19. Modeling temporal evolution of bird populations can be determined by correlating the divergence time, demographic and spatial expansion and geo-climatic changes.

The molecular identification of the interspecific hybrid *Acrocephalus scirpaceus* X *Acrocephalus arundinaceus* led to the following conclusions:

20. The sex distribution in *Acrocephalus arundinaceus* populations is mostly formed by females, sex ratio 42,85 [(m/f) * 100].

21. For *Acrocephalus scirpaceus* and *A. schoenobaenus* we observed a majority of males in populations, sex ratio value being 311.11 and 166.66, respectively.

22. Due to the deficiency of females, *Acrocephalus scirpaceus* males recourse to reproduce with females of the species *Acrocephalus arundinaceus*, that have a deficit of male individuals, resulting interspecific hybrids with intermediary phenotypes.

**Bibliografie**


Proceedings of the National Academy of Sciences of the United States of America 101, 11040-11045.


