“BABEŞ-BOLYAI” UNIVERSITY CLUJ-NAPOCA
BIOLOGY AND GEOLOGY FACULTY
EXPERIMENTAL BIOLOGY DEPARTMENT

Molecular markers for amphibian and reptilian population genetics

Summary of the thesis

Scientific supervisor:
Prof. Dr. Octavian Popescu

PhD student:
Marosi Albert Béla

2011

Table of contents

Acknowledgements 5
Summary 6
Theoretical background 9
1. Use of the MHC genes in population genetics 9
1. 1. Structure and function of the MHC genes 9
1. 2. Comparative genomics of major histocompatibility complexes 12
1. 3. Mechanisms of natural selection that contribute to the maintenance of MHC polymorphism 14
1. 3. 1. The heterozygote advantage hypothesis – overdominant selection 14
1. 3. 2. Frequency dependent selection 15
1. 3. 3. MHC dependent disassortative mating 15
1. 4. Use of the MHC genes in different study areas 16
1. 4. 1. MHC genes in ecological immunology 17
1. 4. 2. MHC genes in sexual selection and mate choice 18
1. 4. 3. MHC genes in parentage 18
1. 4. 4. MHC genes in historical demography of populations 18
1. 4. 5. MHC genes in systematics 18
1. 4. 6. MHC genes in population genetics 19
1. 4. 7. MHC genes and evolution of the immune system 19
1. 4. 8. MHC genes and the evolution of multigene families 20
1. 4. 9. The importance of immune gene variability in conservation biology 20
1. 5. Approaches for isolating MHC genes 22
1. 5. 1. Isolation of proteins with antibodies 22
1. 5. 2. PCR approaches to cloning MHC genes 23
1. 5. 2. 1. Identification of transcribed genes 24
1. 5. 2. 2. Southern blots 24
1. 5. 2. 3. Genomic cloning 24
1. 6. Analysis of MHC polymorphism 25
2. Use of COI genes in phylogeny, phylogeography and population genetics 26
2.1. Approaches for isolating COI genes 29

3. Population genetics 29
4. Typical statistical methods for the MHC genes 35
5. Typical statistical methods for phylogeny and phylogeography, based on mitochondrial markers 36
Research projects and results 40
1. Identification of partial MHC class II B exon 2 sequences in two closely related snake species: *Natrix tessellata* and *Natrix natrix* 40
1. 1. Methods 40
1. 2. Results and discussion 41
2. Identification and characterization of major histocompatibility complex class IIB alleles in three species of European ranid frogs 45
2. 1. Materials and Methods 45
2. 2. Results 48
2. 3. Discussion 52
3. Identification of COI partial sequences in two closely related frog species, *Rana dalmatina* and *Rana temporaria* 55
3. 1. Methods 55
3. 2. Results and discussion 57
4. COI-based phylogeography and intraspecific genetic variation of *Rana dalmatina* populations in the vicinity of the Carpathians 61
4. 1. Materials and methods 62
4. 2. Results 64
4. 3. Discussion 67
5. COI based phylogeny in European ranid frogs: *Rana arvalis*, *Rana temporaria*, *Rana dalmatina*, *Pelophylax kurtmuelleri* and *Pelophylax lessonae* 68
5. 1. Materials and methods 69
5. 2. Results and discussion 70
6. Glacial refugium of *Mesotriton alpestris* in the Apuseni Mountains 74
6. 1. Materials and methods 74
6. 2. Results and discussion 75
7. Phylogeography of the dice snake (*Natrix tessellata*) in South-East Europe 81
7. 1. Materials and methods 82
7. 2. Results and discussion 83
8. Comparison between MHC and COI based phylogeny 88

Conclusions 92
References 94
List of scientific publications 100
Conference participations 100

Acknowledgements

The present PhD study was carried out at the Molecular Biology Centre, Institute for Interdisciplinary Experimental Research, "Babe -Bolyai" University, Cluj-Napoca.

First of all I would like to thank my scientific supervisor *Prof. Dr. Octavian Popescu*, for accepting my PhD application, for providing the opportunity to work in his laboratory, for his understanding and trust during the work, for financial and technical support and valuable guidance.

Secondly this thesis and research project would not have been possible without the help and support of Dr. Ioan V. Ghira. I would like to express how grateful I am for his continuous encouragement, new project ideas and assistance in field work and sample collection.

It is a pleasure to say thank you to my former and present colleagues at the laboratory, who provided me with valuable advice, and taught me new work methods during the laboratory work and data analyses: Dr. Jakab Endre, Dr. Sergiu Chira, Dr. Annette Damert, Dr. Bálint Miklós, Dr. Iulia Lupan, Dr. Beatrice Kelemen, Dr. Beatrix Ferencz, Dr. Mircea Chiriac, Kovács Levente, Bartha László, Sajgó Szilárd, Irina Matetovici and Miruna Ghinia.

Special thanks for Dr. Karen M. Kiemnec-Tyburczy for help with data analysis and results interpretation.

Last but not least I would like to thank for my parents for supporting and encouraging me during my PhD study.

Key words: major histocompatibility complex, antigen binding site, cytochrome oxidase subunit I, phylogeny, phylogeography, population genetics, Ranidae, Natrix, Mesotriton.

Summary

The major histocompatibility complex (MHC) is a relatively large region found in the genomes of all the jawed vertebrates. Most of the genes in this region are necessary for a functioning immune response and are thus vital to the survival of an organism. MHC molecules function in the presentation of peptides derived from intra-and extra-cellular pathogens. MHC class II proteins, in particular, are expressed on the surface of specialized antigen presenting cells, such as macrophages, and their primary function is to present extracellular antigens to helper T cells. The mature MHC class II complex is composed of one alpha and one beta peptide chain that are encoded by the MHC II A and B genes. The antigen binding site of this heteromeric protein complex is encoded by the second exons of the alpha and beta genes. These sites are often extremely polymorphic and show signatures of natural selection. Theories that predict the high variation of these sites include: heterozygote advantage, frequency dependent selection and disassortative mating preferences.

Because these genes and the antigen binding sites specifically are so variable and play such a key role in the vertebrate immune response, they are often used as markers with which to assess the population structure and status. Populations with low MHC variability may be highly threatened by infections and/or by the effects of inbreeding.

In case of amphibians and reptiles, especially frogs and snakes, there are very few studies focusing on MHC gene structure and variability. Many snake and frog species are experiencing population declines due to these emerging infectious diseases. Assessment of the genetic variation at MHC loci in endangered and threatened amphibian and reptile species can be used for developing conservation management strategies that target the maintenance of adaptive variability in wild frog and snake populations.

The COI/COX1 (cytochrome c oxidase subunit 1) is the main subunit of the cytochrome c oxidase protein complex, it is coded in the mitochondria and it has catalytic function. It is the most popular marker for population genetic and phylogeographic studies across the animal kingdom. Its popularity has increased even more since it appears that the M1-M6 partition of the COI gene (The Folmer region) is an efficient indication tool for

Metazoan species. This is the reason why this gene fragment became the subject sequence of DNA barcoding. DNA barcoding can serve as a means of accessing taxonomic information and help in the identification of species as well as highlighting those species for which no data are available. It is not a complete tool for phylogenetic reconstruction, but completing with other nuclear or mitochondrial markers we can obtain detailed and accurate information about species/populations phylogeny, phylogeography and population structure.

The present PhD study focused on the identification, characterization and applicability of two different types of molecular markers (MHC class IIB exon 2 and COI) in amphibian and reptile species. The study species were the following: *Rana dalmatina*, *Rana temporaria*, *Rana arvalis*, *Pelophylax lessonae Pelophylax kurtmuelleri*, and *Mesotriton alpestris* – amphibian species and *Natrix tessellata*, *Natrix natrix* and *Lacerta agilis* – reptile species.

Theoretical background

1. Use of the MHC genes in population genetics

This section is dealing with the molecular structure, classification and function of the MHC genes, general approaches for the isolation of these genes and their use in population genetics. For population genetics studies the most relevant is the exon 2 of the MHC class IIB gene. This exon is coding the antigen binding groove of the MHC class II protein beta chain (Fig. 1.) and it is the most variable site in a vertebrate genome (Penn &Ilmonen 2005).

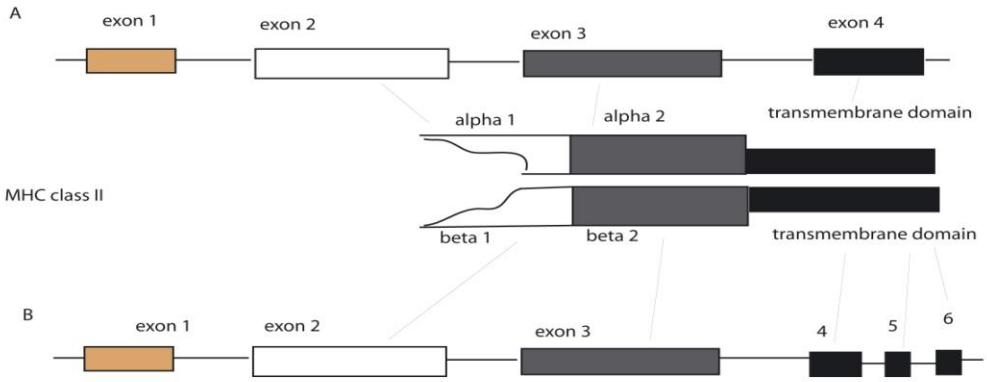

Figure 1: Schematic structure of the MHC class II genes and protein.

After the mammalian model, the MHC is generally divided into regions with similar functions, including class I, class II and class III, extended class I and class II regions (Kelley et al. 2005). The number of genes and the presence of each region vary between species.

Several hypotheses have been proposed to explain how MHC polymorphisms are maintained by natural selection: The heterozygote advantage hypothesis -overdominant selection, frequency dependent selection, MHC dependent disassortative mating, and the inbreeding avoidance hypothesis (Penn 2002).

MHC genes generate a great interest form ecologists, because they are extremely polymorphic and are relevant to many questions of ecology and evolution. Since they are essential in immune recognition of the parasites, they become involved in any area of questions in which parasites play a role. Studies of fitness, expression of secondary sexual characters and mate choice can benefit from the knowledge of MHC variation. MHC genes are potentially becoming important in conservation genetics, and their exceptional diversity can offer novel insights into the population genetic structure and history (Hedrick 2003).

2. Use of COI genes in phylogeny, phylogeography and population genetics

This section is presenting the structure and function of the cytochrome oxidase subunit I genes, approaches to their isolation and their use in animal species barcoding, phylogeny and phylogeography. Studies associated with the Consortium for the Barcode of Life (CBOL) usually sequence a 648 bps long sequence of the COI gene, starting from the 5'end. Intraspecific variation in this sequence is generally 10% less than the interspecific variations and insertions or deletions are rare (Waugh 2007).

3. Population genetics

This section presents and explains general terms, methods and principles applied in population genetics.

4. Typical statistical methods for the MHC genes

The MHC genes are different from other genes regarding the positive selection acting on their antigen binding sites, thus their analysis needs certain specific methods, which are presented in this section.

5. Typical statistical methods for phylogeny and phylogeography, based on mitochondrial markers

This section is dealing with the explanation of general terms of phylogenetics and phylogeography, model choice for sequence analysis, description of models and analytical methods.

Research projects and results

In this section there are presented the research projects separately. For each research the methods results and discussions are separate.

1. Identification of partial MHC class II B exon 2 sequences in two closely related snake species: *Natrix tessellata* and *Natrix natrix*

Two individuals of *Natrix tessellata* (N. tesselata1 and N. tesselata2) and one individual of *Natrix natrix* (N. natrix) were used in our study. To demonstrate that our sequences are indeed MHC sequences and construction of phylogenetic tree, we used homologue MHC sequences from the NCBI. For primer design we used the two *M. corallinus* sequences (FL589895.1 FL590235.1) since alignments with the other reptile sequences yielded few conserved regions. The designed degenerate primers SnakeMHC F: AGC GGG TGC GGT TCC TSS and SnakeMHC R: GTC CRC ATC CGS CTC CCC yielded a 113 bps long product at the annealing temperature of 59ºC.

In case of the 20 *N. tessellata* clones from the 2 individuals we only found 2 alleles (N. tessellata 1 and N. tessellata 2). In case of the single *N. natrix* individual we found 4 alleles (clone1-4). The phylogenetic tree indicates that our MHC sequences cannot be used to distinguish between species. The N. natrix clone1 and clone3 clusters together with the N. tessellata 1, while the N. natrix clone2 and clone4 cluster together with N. tessellata 2, and the

M. corallinus sequences. Our results indicate the presence of trans-species polymorphism, a phenomenon in which the same alleles are retained in different species over evolutionary time (Klein et al. 1998). It is important to note that the snake MHC sequences cluster together with the mammalian MHC sequences instead of the rest of reptile species, which seem to be closer to the bird MHC sequences (Fig. 2.).

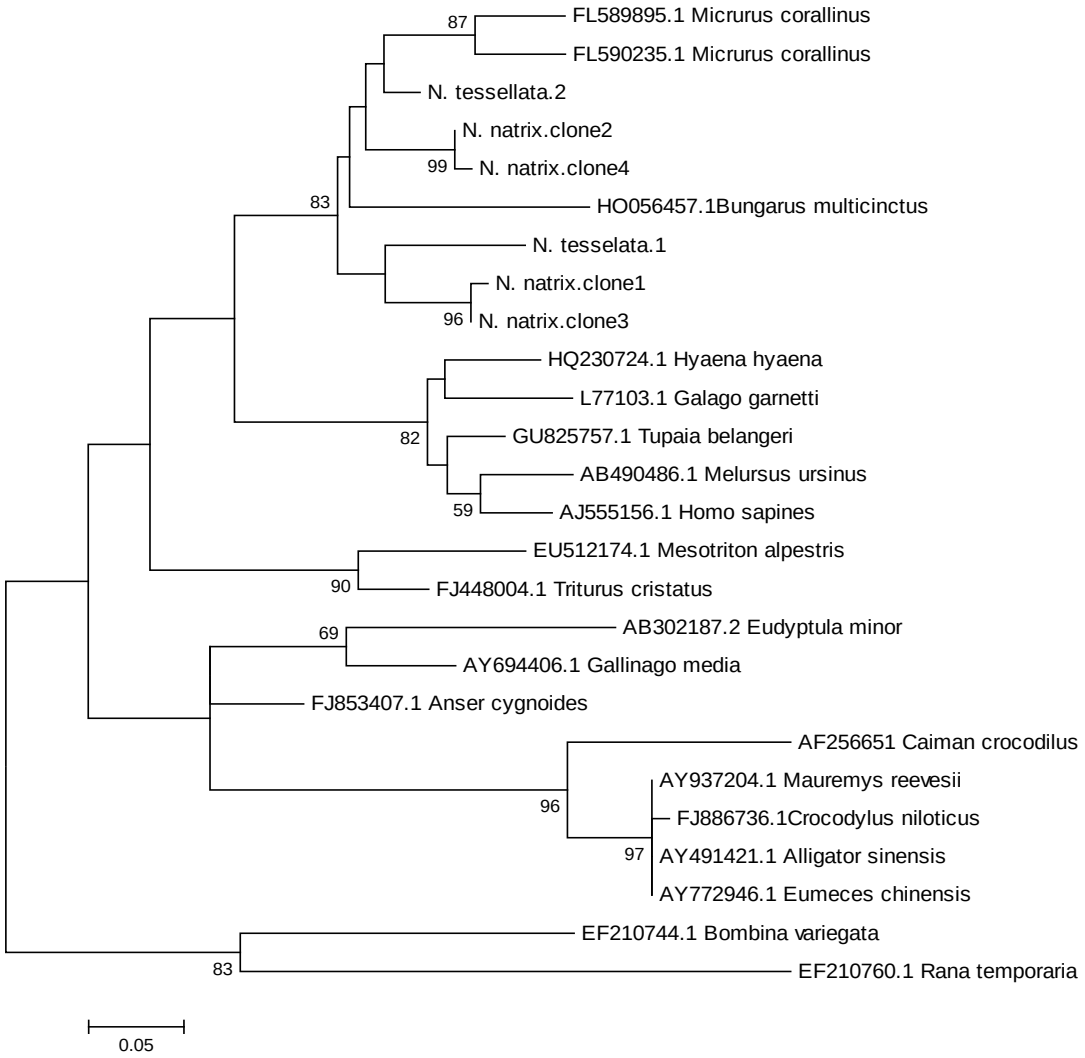

Figure 2: Maximum likelihood phylogenetic tree of the partial MHC class II B exon 2 sequences, generated with 1000 bootstrap replicates.

2. Identification and characterization of major histocompatibility complex class IIB alleles in three species of European ranid frogs

Interest in monitoring amphibian populations, however, is increasing because of the emergence of infectious diseases caused by *Batrachochytrium dendrobatidis* (*Bd*) (Berger et al. 1998) and ranavirus(es) (Granoff 1989). Many European ranid frogs are experiencing population declines due to these emerging infectious diseases (e.g., *Rana klepton esculenta*; Wood et al. 2009 and *Rana temporaria*; Teacher et al. 2010).

The aim of the present study was to identify and characterize the MHC class II beta chain exon 2 in three European ranid species: the Moor frog (*Rana arvalis)*, the Balkan Water Frog (*Pelophylax kurtmuelleri)* and the Pool Frog (*Pelophylax lessonae*). Our goal was to successfully develop and optimize amplification using primers that can be used for population structure analysis and species phylogeography on these and other closely related ranid species.

We used ethanol-preserved tissues from *R. arvalis* (n=3), P. *kurtmuelleri* (n=2), *R. temporaria* (n=2) and *P. lessonae* (n=2). *P. kurtmuelleri* is difficult to distinguish from *R. ridibunda*, we tested our two specimens with the universal primers 16Sar and 16Sbr (Palumbi 1996). We used a degenerate primer pair (MHC-F and MHC-5R) developed by Hauswaldt et al. (2007) to amplify a 235 bps fragment from *R. arvalis* and *R. temporaria* genomic DNA. . Using the sequences of *R. pipiens*, *R. sylvatica*, *R. palustris*, *R. warszewitschii*, *R. yavapaiensis* and *R. catesbeiana* from the NCBI database, we developed a set of degenerate primers to amplify exon2. The sequences of these primers were as follows: RanaF 5′-CAG TGT TAT TAC CGG AAC GGG ACG-3′ and RanaR2: 5′-TTT SMG STC TAT GGC TGY AGG-3′. The expected product length of this primer pair was 243 bps (including primers).

In total were isolated 13 MHC class II exon 2 alleles from the genomic DNA of the three European frog species we sampled. The translated amino acid sequences of all alleles showed significant amino acid similarity to other frog sequences accessioned in GenBank (BLAST e-values less than 2.4e-48). Our 2 *R. temporaria* sequences were only used in the construction of the phylogenetic tree because MHC II B sequences from *R. temporaria* have already been characterized (Zeisset and Beebee 2009) but our identified alleles were longer and not identical to the ones present in the NCBI.

In *R. arvalis*, out of the 61 amino acids in the alignment, there were 12 variable amino acid positions among the eight alleles. Eight of these variable sites were predicted to be in the ABS (antigen binding site) based on the model of Tong et al. (2006).

The rate of nonsynonymous mutations- d_N was significantly higher than the rate of synonymous mutations- d_S, a result which indicates positive selection has acted on these sites (Table 1.). In contrast, the rate of nonsynonymous substitutions was not higher than the rate of synonymous substitutions outside the antigen binding sites.

Table 1: Estimates of nonsynonymous (d_N) and synonymous (d_S) substitution rates of the MHC class II exon 2 sequences from two ranid frog species

Sites	N	d_S	d_N	d_N-d_S	P
R. arvalis ABS	8	0.015 ± 0.014	0.154 ±0.009	0.139 ± 0.043	0.003
R. arvalis Non-ABS	8	0.032 ± 0.016	0.01±0.005	-0.022 ± 0.017	1.0
P. lessonae ABS	4	0.324 ± 0.280	0.431 ±0.214	0.107 ± 0.333	0.375
P. lessonae Non-ABS	4	0.334 ± 0.121	0.199 ± 0.041	-0.135 ± 0.134	1.0

For *P. kurtmuelleri* we found only 1 allele. In the third species, *P. lessonae*, we found 30 variable amino acid sites, with seven of them occurring in the ABS defined by Tong et al. (2006) and 23 in the non–ABS region. We found no significant evidence of positive selection in the ABS and or non-ABS in *P. lessonae* (Table 1.).

The molecular phylogenetic analysis indicated common alleles between species: trans-species polymorphism (Fig. 3.). The *R. arvalis* alleles formed one well-defined clade, but two alleles of *P. lessonae* grouped together with the *R. kurtmuelleri* sequence with very high support. Notably, the two other *P. lessonae* alleles formed a separate clade that was more closely related to the New World ranid species. Surprisingly, the *R. temporaria* alleles we amplified grouped together with MHC sequences from New World *Rana* sequences, although the node separating *R. temporaria* and New World ranids from the *R. arvalis* sequences had low support.

We infer that the *R. arvalis* genome has least two MHC class IIB loci because we recovered four unique alleles from one individual. We also predict that *P. lessonae* has at least two loci, because the amino acid differences between the two groups of alleles are so large that it is likely they represent two loci.

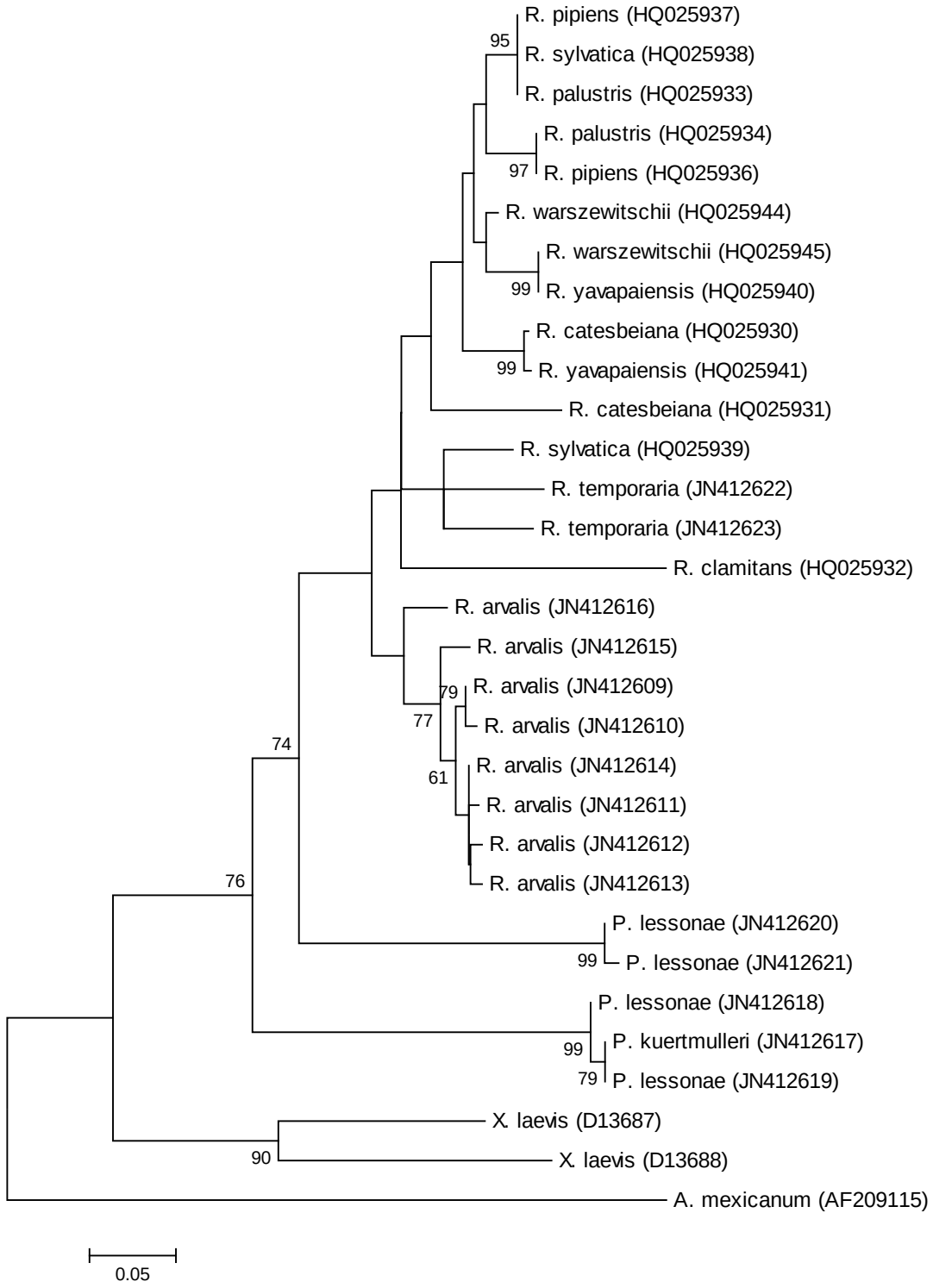

Figure 3: Phylogenetic relationships of the MHC class IIB exon 2 sequences in frogs generated with maximum likelihood. *Ambystoma mexicanum* was used as an outgroup and bootstrap values (1000 replicates) are shown for nodes that received greater than 60% support. GenBank accession numbers are given after the sequence name both for homologue sequences from the NCBI and both for sequences identified by us.

3. Identification of COI partial sequences in two closely related frog species, *Rana dalmatina* and *Rana temporaria*

Here we present COI partial sequences of the two species and analyze the phylogenetic relationships between these sequences and those of other *Rana* species available in the NCBI database. In order to corroborate species assignment based on morphology we also analyzed 16S RNA partial gene sequences of the respective species. Two individuals each of *Rana dalmatina* (dalmatina1 and dalmatina2) and *Rana temporaria* (temporaria1 and temporaria2) were used in our study species.

We used the universal primers 16Sar and 16Sbr (Palumbi 1996) to amplify a sequence of approximately 590 bps of the 16S RNA coding mitochondrial gene.

In order to obtain partial COI gene sequence we used the degenerate primer pair V1F and V1R published by Smith et al (2008). Initially a 710 bps fragment was amplified on dalmatina1 genomic DNA. Based on the sequenced product of this PCR the specific primers were designed: RanaCOIF: 5'TTCTCTACTAACCACAAAGACATTGG 3' and RanaCOIR: 5' TAGACTTCTGGGTGGCCGAAAAATCA 3'. The length of the sequences obtained form the direct sequencing with the RanaCOIF was around 640 bps.

Based on the multiple alignment of *R. temporaria*, *R. dalmatina* and database-derived COI sequences of other *Rana* species a Maximum likelihood phylogenetic tree was constructed (Fig. 4.).

The COI-based tree presents a topology similar to that obtained with 16S RNA sequences as far as *R. dalmatina* and *R. temporaria* are concerned. As in the case of 16S RNA sequences, *R. pyrenaica* is found to be more closely related to *R. temporaria* than to *R. dalmatina*. The clear separation of the European species *Rana temporaria, Rana dalmatina* and *Rana pyrenaica* from the American ones is also in accordance with the data obtained using 16S RNA sequences.

To our knowledge this is the first report of COI sequences for *Rana temporaria* and *Rana dalmatina*. The specific primers developed for direct sequencing should facilitate a fast assessment of population variability in these two species.

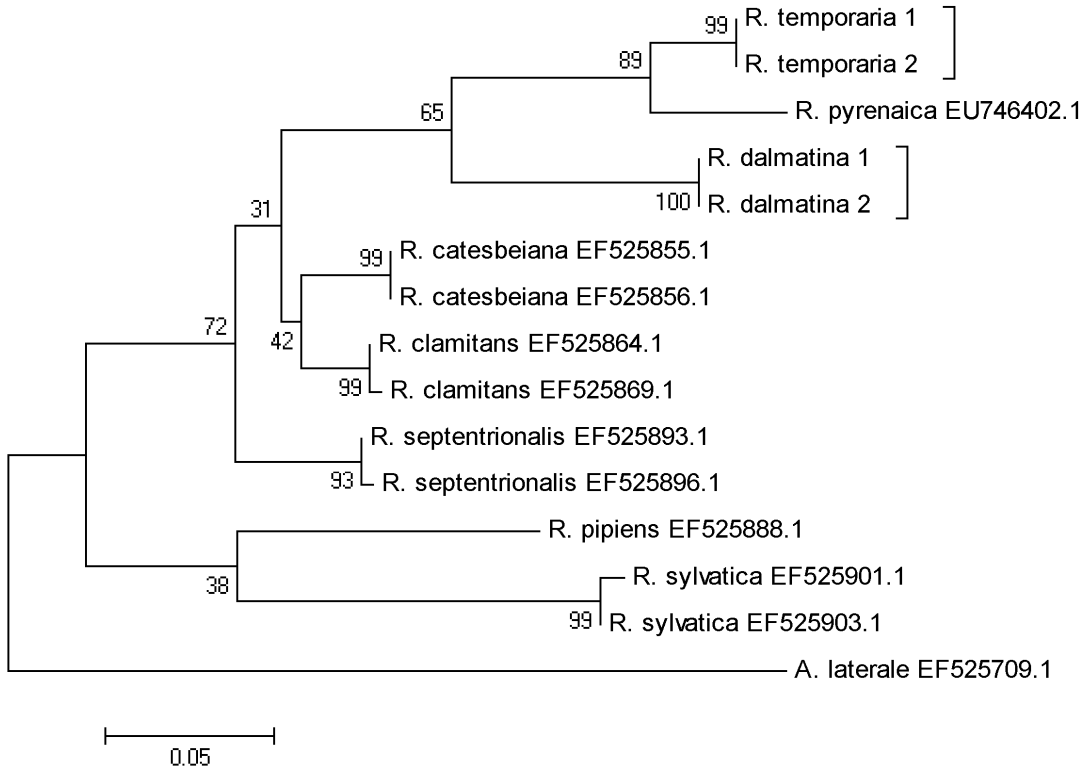

Figure 4: Maximum likelihood phylogenetic tree of COI partial coding sequences of different *Rana* species. Bootstrap values were generated from 1000 replicates. GenBank accession numbers are given near species names.

4. COI based phylogeography and intraspecific genetic variation of *Rana dalmatina* populations in the vicinity of the Carpathians.

The most well known regions that provided refugia for various animal and plat species are the Mediterranean peninsulas and the Caucasus (Taberlet et al. 1998). However it seems that recent glacials could have been survived in more northern cryptic refugia as in case of the cold tolerant *Rana arvalis*, which species shows a high diversity in the Carpathian basin (Babik et al. 2004). Local refugium was identified with genetically distinct haplotypes for the *Rana temporaria* in the south west of Ireland (Teacher et al. 2009). Considering the presence of cryptic refugia for the *Rana temporaria* and *Rana arvalis* outside the Mediterranean peninsulas, we searched for possible indication of cryptic refugia of the *Rana dalmatina* species in the vicinity of the Carpathian Mountains and in the Transylvanian plateau.

Based on the fragment of the cytochrome oxidase subunit I gene COI) that is also used for species barcoding (Smith et al. 2008) we tried to elucidate the genetic variability of the *Rana dalmatina* populations in Romania.

Frog specimens were sampled from 7 geographical regions of Romania (Image 1.). The regions were marked with letters from A – G. Samples belonging to the region A (n = 31) were from the Transylvanian plateau, and they were regarded as one metapopulation since there was no considerable geographical barrier that could split the populations. The region B was west to the Apuseni Mountains and south to the Criș River with sample size of n = 5. The region marked with C was near the southeast part of the Danube River with n = 5 samples. The region D was bordered to the east by the Olt river valley and to the north by the Southern Carpathians with sample size n = 6. The region E included the undulating area in the southeast of the Carpathians with sample size n = 6. From the region marked with F in the vicinity of the Siret River valley we could only collect n = 2 specimens. The region marked with G was the plateau of Dobruja with n = 7 samples. Beside the 62 samples from Romania we analyzed sequences from specimens received from other parts of Europe: n = 1 sample from Spain (North West of the Pyrenees), n = 2 samples from Croatia (South of the Drava River), n = 4 samples from Slovakia, n = 1 sample from Austria and n = 2 samples from Hungary. These sequences were not included in the population structure analyses.

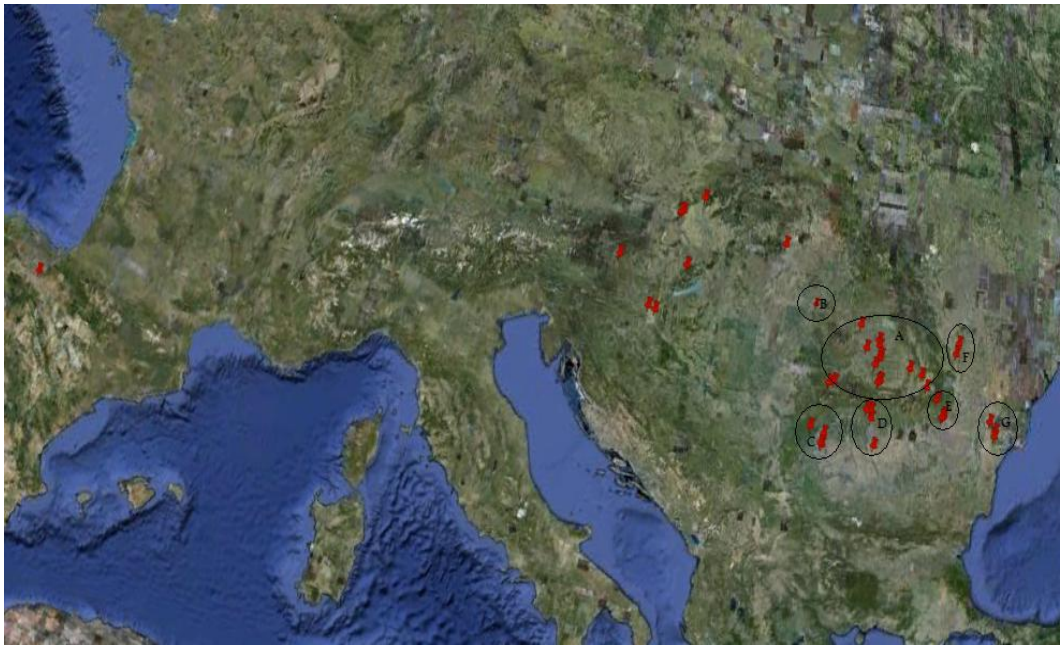

Image 1: Sample collection sites.

For PCR amplifications and for direct sequencing we used the primers RanaCOIF/RanaCOIR from the previous study.

From the analyzed 72 *R. dalmatina* specimens we have identified 13 variable nucleotide positions. All the nucleotide position were synonymous, they did not alter the

amino acid of the translated protein. The sequences differed only in singletons. For the Romanian samples we have identified 9 different alleles.

The haplotype diversity was the highest for samples of the D region, with 5 alleles. In the region A there were 3 alleles and 2 alleles for the B region. In region F we also found 2 alleles from the 2 specimens but due to this low number of the specimens, haplotype diversity must be treated with caution in this case (Table 2.). For the regions C, G, E, we only found the same common allele.

Table 2: Haplotype diversity of the 7 sampled regions; h: haplotype diversity, uh: unbiased haplotype diversity.

Pop	h	uh
C	0.000	0.000
D	0.778	0.933
A	0.231	0.239
B	0.320	0.400
G	0.000	0.000
E	0.000	0.000
F	0.500	1.000

The Pairwise Nei genetic distances were the highest for the samples from the region D (Table 3.). The Mantel test between the genetic distance (GD) and geographic distance (GGD) indicate a weak negative correlation Rxy = -0.108 with a no significant P = 0.07 value.

The AMOVA test did not indicate population differentiation between metapopulation separated by the Carpathian Mountains. However it indicated population differentiations when the 7 regions were analyzed separately (Table 4.), and the highest and significant pairwise PhiPT values were obtained for the D region samples. Our explanation for this high variability experienced in the population D is possible cryptic refugia of the species.

Table 3: Nei genetic distance (GD) and Nei genetic identity (ID) values for the 7 sampled regions

Pop1	Pop2	Nei GD	Nei ID
C	D	0.347	0.707
C	A	0.007	0.993
D	A	0.353	0.702
C	B	0.030	0.970
D	B	0.377	0.686
A	B	0.037	0.964
C	G	0.000	1.000
D	G	0.347	0.707
A	G	0.007	0.993
B	G	0.030	0.970
C	E	0.000	1.000
D	E	0.347	0.707
A	E	0.007	0.993
B	E	0.030	0.970
G	E	0.000	1.000
C	F	0.347	0.707
D	F	0.693	0.500
A	F	0.353	0.702
B	F	0.377	0.686
G	F	0.347	0.707
E	F	0.347	0.707

Table 4: Pairwise population PhiPT and linearized PhiPT values estimations.

Pop1	Pop2	PhiPT	LinPhiPT	P(rand >= data)
C	D	0.261	0.352	0.060
C	A	0.000	0.000	0.220
D	A	0.337	0.508	0.010
C	B	0.000	0.000	0.010
D	B	0.075	0.082	0.160
A	B	0.000	0.000	0.110
C	G	0.000	0.000	1.000
D	G	0.333	0.500	0.010
A	G	0.000	0.000	0.130
B	G	0.073	0.079	0.440
C	E	0.000	0.000	1.000
D	E	0.300	0.429	0.050
A	E	0.000	0.000	0.300
B	E	0.040	0.042	0.560
G	E	0.000	0.000	1.000
C	F	0.474	0.900	0.330
D	F	0.000	0.000	0.520
A	F	0.329	0.490	0.070
B	F	0.015	0.015	0.540
G	F	0.588	1.429	0.260
E	F	0.538	1.167	0.270

5. COI based phylogeny in European ranid frogs: *Rana arvalis, Rana temporaria, Rana dalmatina, Pelophylax kurtmuelleri* and *Pelophylax lessonae*

In the herein study we used the cytochrome oxidase subunit I gene sequence (Folmer region) to determine the phylogenetic relatedness of 5 ranid frog species. Although the phylogeny of ranid and other frog spices was already carried out by the aid of 16S rRNA, 12S rRNA, tRNA-valine, histone H3, rhodopsin (Frost et al. 2006, Veith et al. 2003), the COI gene sequence of ranid spices is still to be determined. The Folmer region of the COI gene is the most wide spread tool for metazoan species identification (Waugh 2007). DNA barcoding can serve as a means of accessing taxonomic information and help in the identification of species as well as highlighting those species that for which no data are available. The phylogenetic tree obtained from COI sequences of the ranid frogs can be compared to the tree constructed on the 16S rRNA sequences.

The specimens were the same as from the previous studies and the RanaCOIF/RanaCOIR primers were used for PCR amplification and direct sequencing.
From the 3 specimens of *R. arvalis* we obtained 3 different COI haplotypes while for the 2 specimens of *P. kurtmuelleri* and *P. lessonae* we obtained only 1 COI haplotype per species. From our previous study we had 1 haplotype of *R. dalmatina* and 1 haplotype of *R. temporaria*.

The estimates of evolutionary divergences between sequence pairs were relatively high, the highest values being between *P. kurtmuelleri* 01 and *R. temporaria* 01 (Table 5.).

Table 5: Estimates of Evolutionary Divergence between the 7 identified haplotypes.

P._kurtmuelleri_01						
P._lessonae_01	0.173					
R._arvalis_01	0.234	0.208				
R._arvalis_02	0.239	0.213	0.006			
R._arvalis_03	0.234	0.208	0.002	0.005		
R._temporaria_01	0.244	0.199	0.102	0.105	0.104	
R._dalmatina_01	0.235	0.225	0.155	0.151	0.153	0.159

Using the available COI sequences from the NCBI and our obtained sequences we constructed a phylogenetic tree of the ranid species. According to the COI based phylogenetic tree the European Rana species form one distinct clade from the New World *Rana* spices, however the *R. catesbeiana* seems to occupy an intermediate position between *Rana* species

of the two continents (Fig. 5.). Within the European Rana species the *R. temporaria* is closest to the *R. pyrenaica* while the *R. dalmatina* represents a basal sequence for the European *Rana* species. The two *Pelophylax* species form a different clade from the *Rana* species, indicating that divergence between the green frogs and brown frogs is older than separation of the European and New World brown frogs.

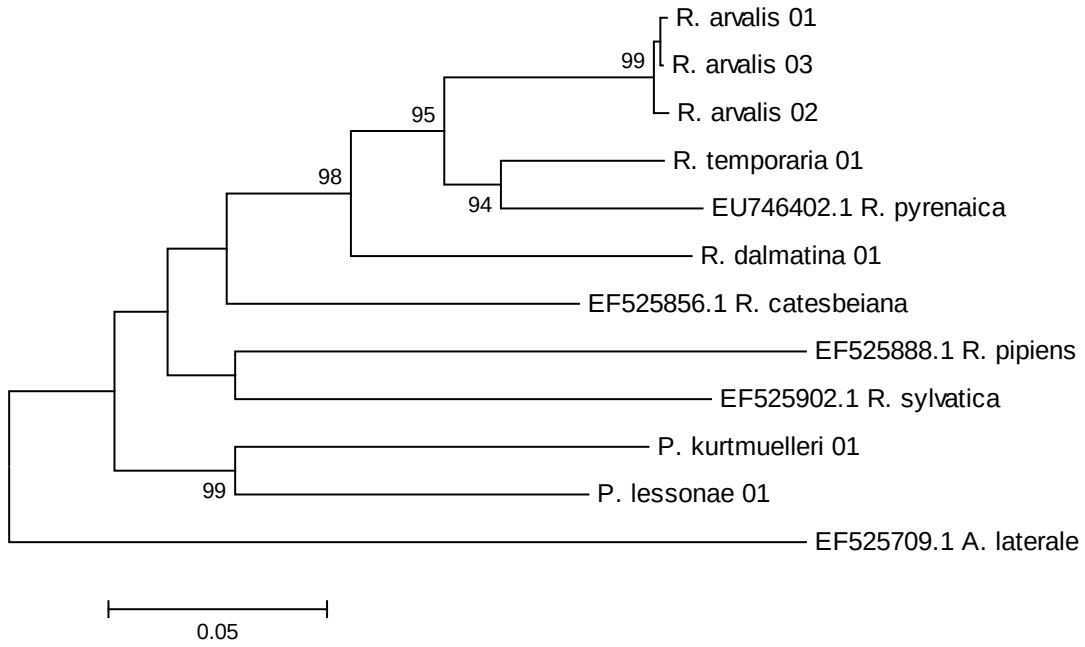

Figure 5: COI based phylogenetic tree of ranid species based on Maximum likelihood method and 1000 bootstrap replicates.

The phylogenetic tree constructed on the basis of 16S rRNA sequences is only slightly different from the COI based phylogenetic tree (Fig. 6.).

Our results indicate that the average evolutionary divergence over all sequences is 0.2074 ± 0.0132 for the COI haplotypes (Kimura 2-parameter model) while for the 16S rRNA sequences it is significantly lower: 0.103 ± 0.01 ($P < 0.01$). We can conclude that the COI sequence having more variability can yield a better resolution and constitutes a better molecular marker for species barcoding and phylogeny.

The alignment of the identified 7 haplotypes indicated a high number of variable sites between the 5 species. The high number of variable sites confirms that the Folmer region of the COI gene is the most convenient tool for species identification and barcoding.

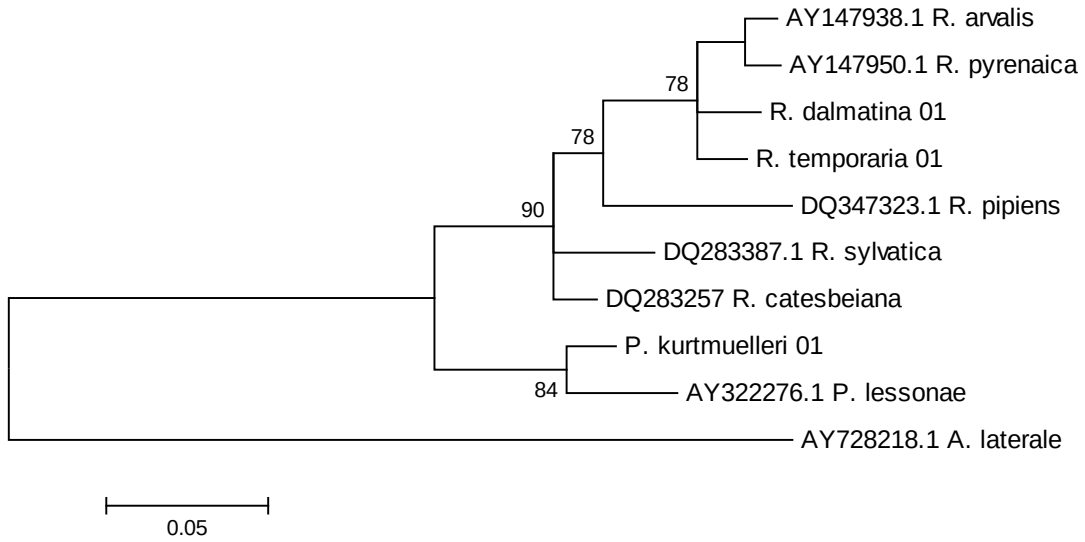

Figure 6: 16S rRNA based phylogenetic tree of ranid species based on Maximum likelihood method and 1000 bootstrap replicates.

6. Glacial refugium of *Mesotriton alpestris* in the Apuseni Mountains

This study aimed the genetic analysis of the alpine newts from Apuseni Mountains and the Southern Carpathians, searching for possible cryptic refugia of the species. Two populations were analyzed, in concordance with the two geographical areas. From the Apuseni Mountains the newts originate from Mădrigeşti locality, from an altitude of 445 m. From the Middle Carpathians, the newts originate from the area of Voineasa locality, from 830 m altitude.

In total 48 mountain newts were sampled, 24 samples from each population (Mădrigeşti: m1-24, Voineasa: v1-24).

After initial amplification with the universal primer set for vertebrates: V1F-V1R (Smith et al. 2008) we designed or specific primers:

MesF: 5’ - TTCTCTACTAACCACAAAGACATTGG - 3’ and

MesR: 5’ - TAGACTTCTGGGTGACCAAAAAACCA - 3’. Using the specific primer set we performed PCR on the 48 samples, and in order to exclude any possible amplification errors, each PCR was repeated.

After analyzing the 48 sequences, we identified 4 alleles: A, B, C and D. The identified alleles cluster together with the sequences of *M. alpestris cyreni*, which is from the

Iberian Peninsula and *M. alpestris reiseri* (Bosnian alpine newt), which species inhabits the Balkans (Fig. 7.). The sequences for comparison were obtained from the NCBI database.

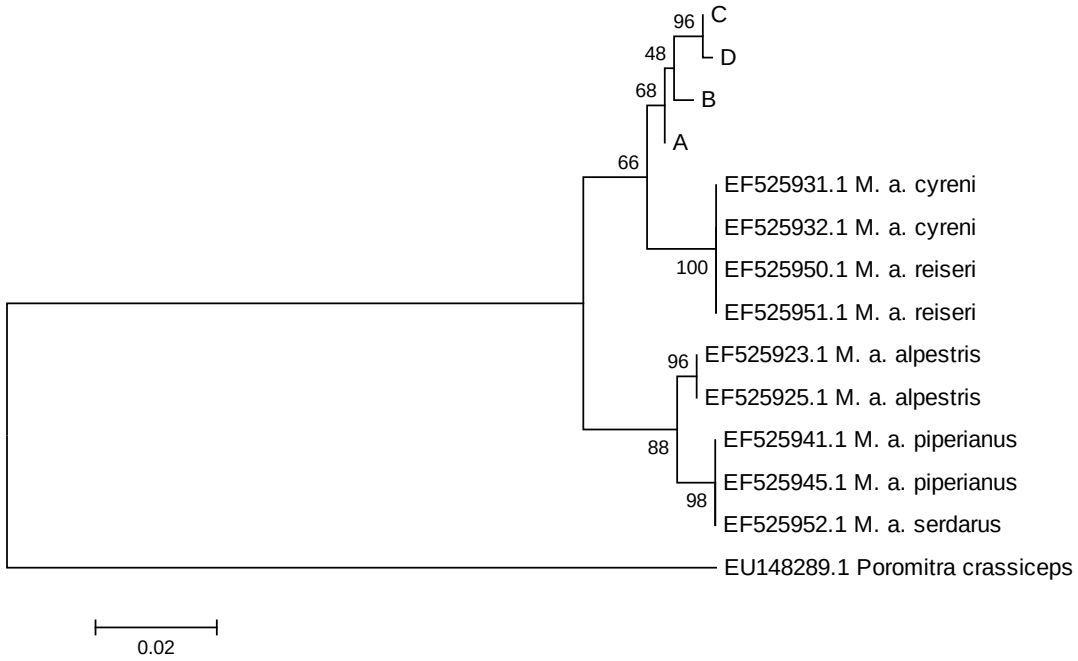

Figure 7: Maximum likelihood phylogenetic tree of the identified alleles and the *M. alpestris* sequences from the NCBI database. Bootstrap values were generated from 1000 replicates. The accession numbers are marked on the left side of the species names.

If we compare the two populations, there is a clear distinction in the allelic composition. All of the 4 identified allele were identified in the population from Mădrigeşti, with a haplotypical diversity of 0.615, while the population from Voineasa is homogenous, having only one allele (Fig. 8a. - b.), fact which involves the founder effect in case of this population. The value of the PhiPT is 0.665 (P=0.001), showing a major difference between the two populations. The AMOVA test shows that the molecular variability regarding the haplotypes is higher between the two populations than inside the two populations (Fig. 9.).

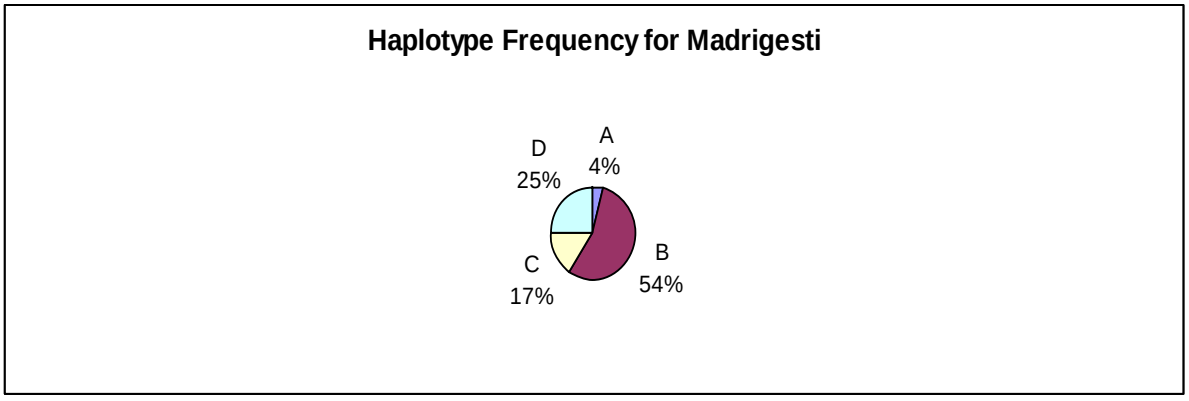

Figure 8a: The frequency of the haplotypes in the populations from Mădrigeşti.

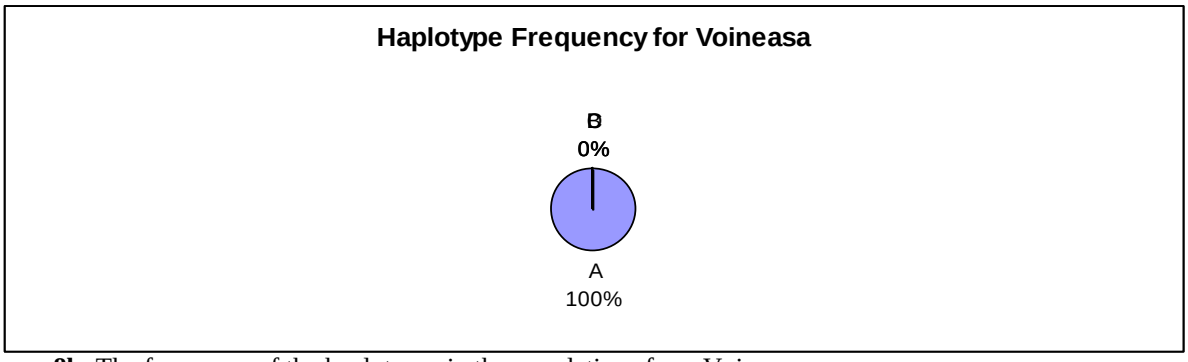

Figure 8b: The frequency of the haplotypes in the populations from Voineasa.

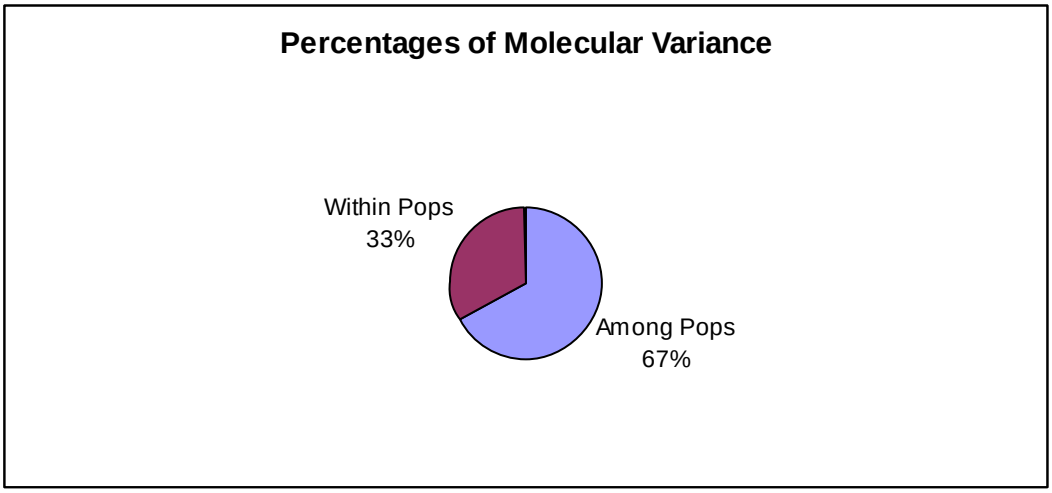

Figure 9: The percentage of molecular variance between and within the populations resulted by the AMOVA test.

The comparison of the two populations clearly supports the hypothesis of cryptic refugia at the Mădrigeşti population, and the subsequent founder effect to the higher altitudes.

7. Phylogeography of the dice snake (*Natrix tessellata*) in South-East Europe

In this study we used the cytochrome oxidase subunit 1 barcoding region to elucidate the phylogenetic relationship between *Natrix tessellata* haplotypes from Eastern Europe and The Balkans. We were able to compare the overlapping between the phylogenetic tree based on the COI sequences and the phylogenetic tree based on the cytochrome b from the study of Guicking et al. (2009). We used in total 15 specimens of *N. tessellata* (Image 2.).

In order to obtain partial COI gene sequence, we used the degenerate primer pair V1F and V1R published by Smith et al (2008).

Out of the 15 specimens we have found 8 haplotypes provided with the following accession numbers: A (JN871603): Ro1 – Ro5, Ser1, Ser2; B (JN871604): Ser3; C (JN871605): Ita; D (JN871606): Gr3, E (JN871607): Ukr1, Ukr2; F (JN871608): Ukr3; G (JN871609): Gr1; H (JN871610): Gr2 (Image 2).

Alignment of the sequences yielded 48 variable nucleotide positions out of 621 nucleotides.

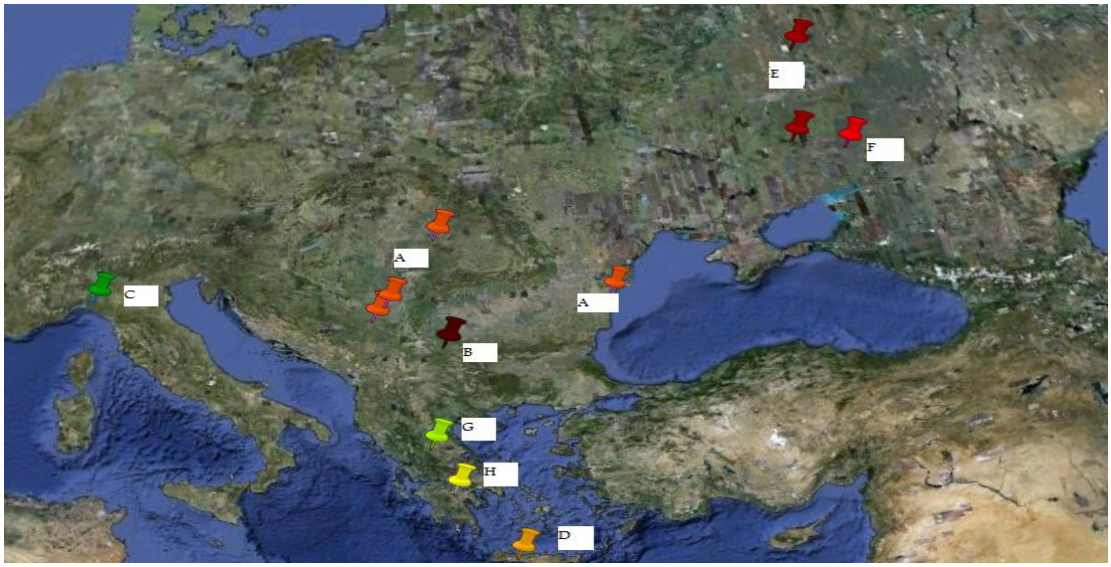

Image 2: Identified haplotypes.

The highest values for the pairwise divergence between sequences were found for the Gr1 and Ukr1-2 sequences (0.056) and in general the highest values were between the Gr1 Gr2 sequences paired with the rest of the sequences suggesting the basal position of the haplotypes from Greece (Table 6.).

Table 6: Estimates of pairwise divergences between sequences based on Kimura 2-parameter model.

Gr1														
Gr2	0.003													
Gr3	0.052	0.050												
Ita	0.052	0.050	0.013											
Ro1	0.054	0.052	0.011	0.002										
Ro2	0.054	0.052	0.011	0.002	0.000									
Ro3	0.054	0.052	0.011	0.002	0.000	0.000								
Ro4	0.054	0.052	0.011	0.002	0.000	0.000	0.000							
Ro5	0.054	0.052	0.011	0.002	0.000	0.000	0.000	0.000						
Ser1	0.054	0.052	0.011	0.002	0.000	0.000	0.000	0.000	0.000					
Ser2	0.054	0.052	0.011	0.002	0.000	0.000	0.000	0.000	0.000	0.000				
Ser3	0.056	0.054	0.013	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002			
Ukr1	0.056	0.054	0.040	0.043	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.047		
Ukr2	0.056	0.054	0.040	0.043	0.045	0.045	0.045	0.045	0.045	0.045	0.045	0.047	0.000	
Ukr3	0.054	0.052	0.038	0.042	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.045	0.002	0.002

The phylogenetic analyses of the *N. tessellata* sequences indicate 3 major clades (Fig. 10.). One group is formed by the Ukrainian sequences and another separate group by the Gr1 and Gr2 sequences which are from the mainland of Greece. It was surprising that spatially distant samples clusters together (Gr3 sample from the island of Crete with the rest of the sequences from Romania, Italy and Serbia), but, actually, our results are in concordance with the findings of Guicking et al. (2009) which were based on the cytochrome b gene. According to that study, the colonization of Greece and Europe occurred in two waves, the Gr1 and Gr2 COI haplotypes are remnants of the first colonization and they survived isolated since the early Pliocene. The rest of Western and Central Europe was colonized after the glacials.

The Ukrainian *N. tessellata* haplotypes form a separate clade from the European dice snake haplotypes and this topology is also in concordance with the findings of Guicking et al. (2009): the north of the Black Sea and the north east European habitats of the dice snake were colonized from the Caucasian refugium. Northernmost and well isolated population of *N. tessellata* from Galichya Gora seems to be result from a very recent and fast colonization event: time of isolation was not long enough for any genetic divergence in the quite variable mitochondrial COI.

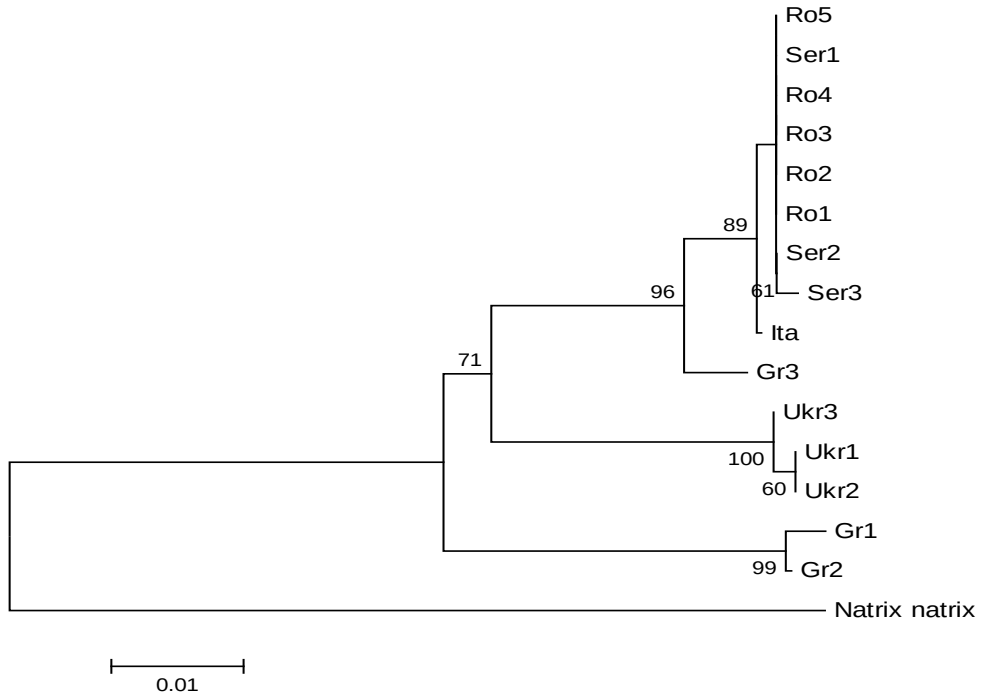

Figure 10: Phylogenetic analysis of the 15 *N. tessellata* COI sequences, based on Kimura 2-parameter model and Maximum likelihood method. The bootstrap consensus tree was inferred from 1000 replicates. *Natrix natrix* ((JN871611) sequence was used for outgroup.

8. Comparison between MHC and COI based phylogeny

Having at disposal the MHC IIB exon 2 sequences and COI sequences from ranid species we can compare the phylogenetic trees yielded by the two types of molecular markers. The Phylogenetic trees were created by using our newly identified sequences together with sequences from the NCBI database. The newly identified sequences for the COI and MHC genes come from the same specimens; this is not the case for the sequences from the NCBI.

The COI sequences group in 3 separate clades. One group is for the European *Rana* species, one group for the New World *Rana* species and one group for the *Pelophylax* species (green frogs). The *Pelophylax* sequences are basal to the brown frog sequences indicating that separation between green frogs and brown frogs predates the separation between the New World and European *Rana* species (Fig. 11.).

The phylogenetic tree based on the MHC sequences seems slightly different. The alleles from the *R. arvalis* cluster together and they are positioned the closest to the *R. temporaria* alleles. The MHC alleles from the New World *Rana* species also cluster together, with one exception of a *R. sylvatica* allele, those groups together with the *R. temporaria* alleles. The trans-species polymorphism can be observed for the New World Rana sequences, the alleles from one species cluster closer to alleles for the other species. This is not the case for the green frog sequences that also group separately from the *Rana* sequences. As in case of the COI sequences the Pelophylax sequences are the basis for the ranid sequences supporting the hypothesis of earlier divergence between green frogs and brown frogs than between brown frogs of the two continents (Fig. 12.).

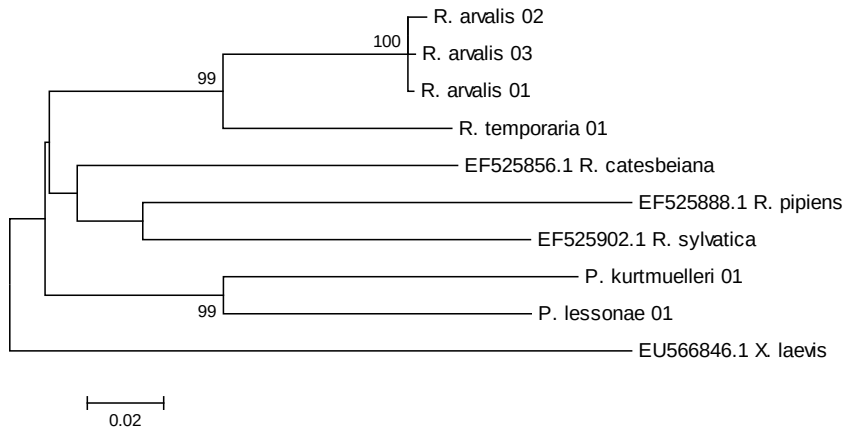

Figure 11: Phylogenetic analysis of the ranid COI sequences, based on Kimura 2-parameter model and Maximum likelihood method. The bootstrap consensus tree was inferred from 1000 replicates.

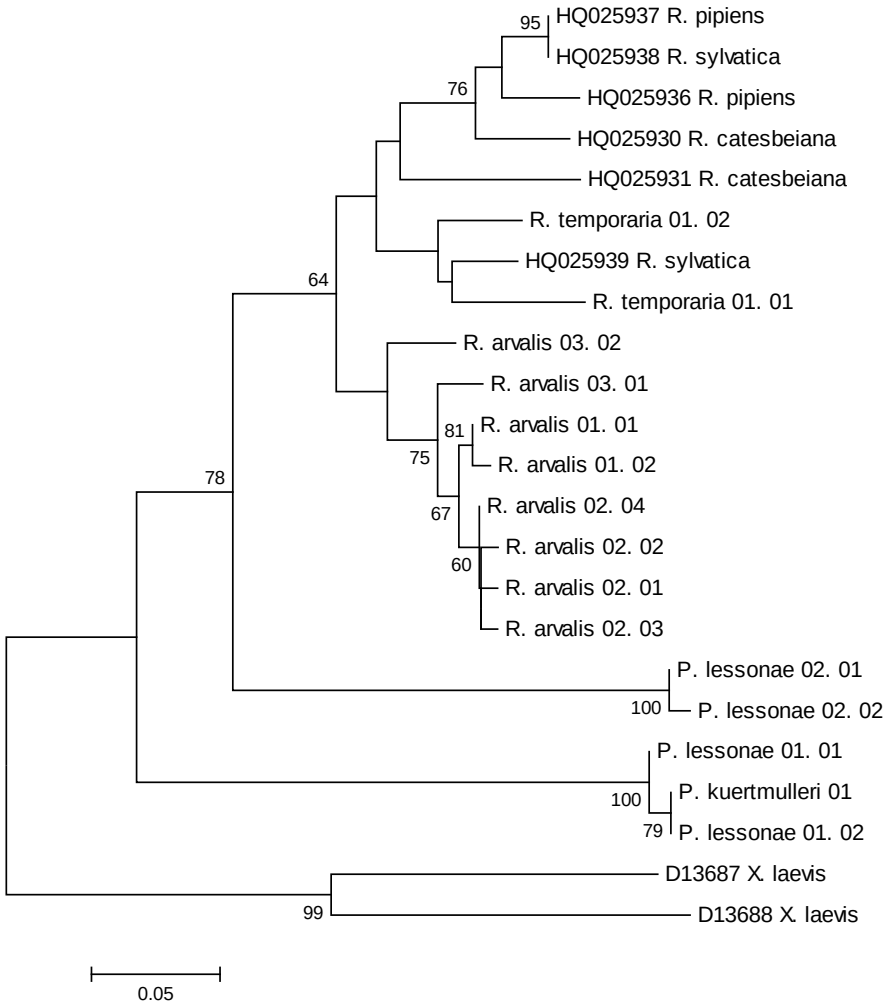

Figure 12: Phylogenetic analysis of the ranid MHC IIB exon 2 partial sequences, based on Kimura 2-parameter model and Maximum likelihood method. The bootstrap consensus tree was inferred from 1000 replicates.

Our assumption is that MHC gene sequences can be more effective in species phylogeny analyses when we have most of the alleles from the target species. When we have more alleles we can differentiate between species specific alleles and common alleles between species. While the species specific alleles help the identification of the species, the common alleles still can indicate those species that share common alleles are closer to each other than species that do not share common alleles. Our study result conform the findings of Berggren et al. (2005), the use of MHC alleles together with the mitochondrial markers allows us to draw more detailed conclusions regarding species phylogeny.

Conclusions

1. The major histocompatibility complex class II B gene partial sequences of the dice snake (*Natrix tessellata*) and the grass snake (*Natrix natrix*) show trans–species polymorphism. There are common alleles among different species.
The grass snake has at least two MHC IIB loci since 4 alleles were identified in 1 specimen and they cluster separately. The 2 MHC loci can be predicted for the dice snake as well, since the two identified alleles cluster in different groups.
The MHC class II B exon 2 region in the snakes has a high variability; in a sequence of 77 bps we found 28 variable nucleotide positions.

2. The major histocompatibility complex class II B gene partial sequences of the *Rana* species also present trans-species polymorphism. Alleles are shared within *Rana* species but not between *Rana* and *Pelophylax* species.
The *Rana arvalis* genome has at least 2 MHC IIB loci, since 4 alleles were isolated from 1 specimen. It can be predicted that the *Pelophylax lessonae* also has at least 2 loci because the amino acid differences between the two groups of alleles is of high value.
For the *Rana arvalis* the rate of nonsynonymous mutation was significantly higher than the rate of synonymous mutation on the antigen binding sites. This substitution pattern is in concordance with the expected evolution pattern of the MHC genes, which predicts a positive selection on the antigen binding sites.

3. The phylogenetic tree of the ranid species based on the cytochrome oxidase Folmer region has a similar topology to the phylogenetic tree based on 16S rRNA sequences. The COI sequences of the European *Rana* species cluster together and separately from the New World *Rana* COI sequences. The *Pelophylax* species form a distinct and basal clade to the *Rana* species. According to the COI sequences the divergence between green frogs (*Pelophylax*) and brown frogs *(Rana*) occurred earlier than the divergence between European and New World *Rana* species.

4. The cytochrome oxidase based phylogenetic study of the agile frog (*Rana dalmatina*) revealed a low genetic variability of the populations in Romania and the Carpathian basin, there were 9 identified alleles.
The Carpathian Mountains do not represent a physical barrier for the gene flow and spread of this species.
It is highly indicated the existence of a cryptic refugium between the Olt River valley and Southern Carpathians. In comparison with the other sampled regions this region showed a high allelic diversity.

5. The comparison of two alpine newt (Mesotriton alpestris) populations from the Apuseni Mountains and the Southern Carpathians revealed a cryptic refugium for the species in the Apuseni Mountains.
The high altitude population form Southern Carpathians is homogenous indicating the founder effect from the lower altitude population, which has 4 alleles.

6. The phylogenetic analysis of the dice snake (*Natrix tessellata*) in East Europe and the Balkans indicates 3 major clades and 8 different haplotypes of cytochrome oxidase subunit I sequence. One group is formed by specimens from Romania, Serbia and Crete, another clade by the samples from Ukraine and the third groups is formed by the sequences from the Balkan Peninsula. Sequences from the Balkan Peninsula (Thessaly) are basal to the other sequences indicating that colonization of Europe occurred in 2 waves and the colonization of Crete and East Europe is result if the from he second wave.

7. The use of MHC alleles (nuclear marker) together with the COI alleles (mitochondrial marker) contribute to a more detailed phylogenetic analysis and systematics of the species, since the two type of markers complete well each other. The COI based phylogenetic tree of the ranid species is similar to the MHC based phylogenetic tree although the two types of markers have different evolutionary forces acting on them.

Selective References

Babik W, Branicki W, Sandera M, Litvinchuk S, Borkin L J, Irwin J T, and Rafiński J. 2004. Mitochondrial phylogeography of the moor frog, Rana arvalis. *Molecular Ecology*, *13*(6): 1469-1480

Berger L, Speare R, Daszak P, Green D E, Cunningham A A, Goggin C L, Slocombe R, Ragan M A, Hyatt A D, McDonald K R, Hines H B, Lips K R, Marantelli G, Parkes H. 1998. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc Natl Acad Sci USA*, 95:9031–9036

Berggren K T, Ellegren H, Hewitt G M and Seddon J M. 2005. Understanding the phylogeographic patterns of European hedgehogs, *Erinaceus concolor* and *E. europaeus* using the MHC. *Heredity*, 95:84-90

Frost D R, Grant T, Faivovich J, Bain R H, Haas A, Haddad C F B, et al. 2006. The amphibian tree of life. *Bull Am Mus Nat Hist*, 297:1–370

Granoff A 1989. Viruses of Amphibia: an historical perspective. Pp 3-12, *Viruses of Lower Vertebrates* Ahne W and Kurstak E eds, Springer-Verlag, Berlin

Guicking D, Joger U, Wink M. 2009. Cryptic diversity in a Eurasian water snake (*Natrix tessellata*, Serpentes: Colubridae): Evidence from mitochondrial sequence data and nuclear ISSR-PCR fingerprinting. *Org Divers Evol*, 9: 201-214

Hauswaldt J S, Stuckas H, Pfautsch S and Tiedemann R. 2007. Molecular characterization of MHC class II in a nonmodel anuran species, the fire-bellied toad *Bombina bombina*. *Immunogenetics*, 59:479-491

Hedrick P W. 2003. The major histocompatibility complex (MHC) in declining populations: an example of adaptive variation. Pp 97-113, *Reproduction Science and Integrated conservation*, Holt W V, Pickard A R, Rodger J C and Wildt D E eds, Cambridge Univ. press, Cambridge, UK

Kelley J, Walter L, Trowsdale J. 2005. Comparative genomics of major histocompatibility complexes. *Immunogenetics,* 56**:**683-695

Klein J, Sato A, Nagl S, O'hUigin C. 1998. Molecular trans-species polymorphism. Annu *Rev Ecol Syst*, 29:1-21

Palumbi S R. 1996. Nucleic acids II: the polymerase chain reaction. Pp 205-247, *Molecular Systematic*, Hillis D M, Moritz C and Mable B K eds, Sinauer & Associates Inc, Sunderland

Penn D J and Ilmonen P. 2005. Major histocompatibility complex. *ELS,* **1-7**

Penn D J. 2002. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology,* 108:1–2
1
Smith M A, Poyarkov N A Jr and Hebert P D N. 2008. DNA BARCODING: CO1 DNA barcoding amphibians: take the chance, meet the challenge. *Mol Ecol Resour*, 8:235-246

Taberlet P, Fumagalli L, Wust-Saucy A G and Cossons J.-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol Eco,* 7:453–464

Teacher A G F, Garner T W J and Nichols R A. 2009. European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, 102(5): 490-496

Teacher A G, Cunningham A A, Garner T W. 2010. Assessing the long-term impact of ranavirus infection in wild common frog populations. *Biol Cons*, 13:514–522

Tong J, Bramson J, Kanduc D, Chow S, Sinha A A and Ranganathan Sh. 2006. Modeling the bound conformation of *Pemphigus Vulgaris*-associated peptides to MHC Class II DR and DQ Alleles. *Immunome Research,* 2:1

Veith M, Kosuch, J, Vences M. 2003. Climatic oscillations triggered post-Messinian speciation of Western Palearctic brown frogs (Amphibia, Ranidae). *Mol Phylogenet Evol,* 26(2): 310-27

Waugh J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays,* 29: 188–197

Wood L R, Griffiths R A, Schley L. 2009. Amphibian chytridiomycosis in Luxembourg. *Bull Soc Nat Luxemb,* 110:109-114

Zeisset I, Beebee J C. 2009. Molecular characterization of major histocompatibility complex class II alleles in the common frog, *Rana temporaria*. *Mol Ecol Resour*, 9:738-745

List of scientific publications

Béla Marosi, Tibor Sós, Ioan V. Ghira, Annette Damert and Octavian Popescu. 2010. Identification of COI partial sequences in two closely related frog species, *Rana dalmatina* and *Rana temporaria*. Herpetologica Romanica. Vol.4. (http://herpetofauna.uv.ro/herprom-vol4.html)

Béla Marosi, Ioan V. Ghira, Tibor Sós and Octavian Popescu. 2011. Identification of partial MHC class II B exon 2 sequences in two closely related snake species: *Natrix tessellata* and *Natrix natrix*. Herpetologica Romanica. In press

Conference participations

Béla Marosi, Octavian Popescu, Annette Damert, Ioan V. Ghira. 2009. MHC (major histocompatibilitycomplex) genek a kockas siklonal - Gene MHC la sarpele de apa. Conferinta X. RODOSZ

Béla Marosi, Karen M. Kiemnec-Tyburczy, Ioan V. Ghira, Tibor Sos and Octavian Popescu. 2011. Identification and characterization of major histocompatibility complex class IIB alleles in three species of European ranid frogs. MCCMB'11 (5^{th} Moscow Conference on Computational and Molecular Biology)