Cytonuclear interactions in a species of *Podarcis* lizard

Gabriel Mochales Riaño
Máster universitario en Bioinformática y bioestadística
TFM – Bioinformática y Bioestadística

Cinta Pegueroles Queralt / Catarina Pinho
Javier Luis Cánovas Izquierdo / Marc Maceira
24/06/2020

Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)

A) Creative Commons:

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-NoComercial 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual 3.0 España de Creative Commons

Esta obra está sujeta a una licencia de Reconocimiento 3.0 España de Creative Commons

B) GNU Free Documentation License (GNU FDL)

C) Copyright

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Cytonuclear interactions in a species of Podarcis lizard*
Nombre del autor:	*Gabriel Mochales Riaño*
Nombre del consultor/a:	*Cinta Pegueroles Queralt / Catarina Pinho*
Nombre del PRA:	*Javier Luis Cánovas Izquierdo / Marc Maceira*
Fecha de entrega (mm/aaaa):	24/06/2020
Titulación::	Máster Universitario en Bioinformática y Bioestadística
Área del Trabajo Final:	*TFM – Bioinformática y Bioestadística*
Idioma del trabajo:	*Inglés*
Palabras clave	*Cytonuclear evolution; mtDNA; oxphos*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

La coevolución es el proceso de cambio evolutivo recíproco entre especies que interaccionan. Se piensa que este proceso es clave en diversificaciones biológicas y ha sido relacionado con la especiación. Evolución coordenada entre genomas nucleares y de orgánulos pueden ocurrir debido a cambios recíprocos en las funciones de proteínas que interaccionan e incluso puede jugar un papel clave en la especiación. En este trabajo de fin de máster se usó datos de target-capture sequencing de 122 individuos de dos linajes mitocondriales diferentes para testar la hipótesis de evolución compensatoria mediante la comparación de árboles filogenéticos de tres tipos diferentes de genes: mtDNA, genes nucleares que interaccionan con las mitocondrias y genes nucleares aleatorios. Nuestros resultados confirmaron la presencia de los dos linajes mitocondriales. Además, Tajima's D mostró como las mitocondrias estan bajo una alta presión purificadora. Por el contrario, no se encontraron diferentes entre los genes oxphos y los genes nucleares. Más estudios deberían de realizarse para confirmar o no la hipótesis de la evolución compensatoria en estos lagartos con los datos producidos en este estudio

Abstract (in English, 250 words or less):

Coevolution is the process of reciprocal evolutionary change between interacting species. This process is thought to be a major driver of biological diversification and has been linked to speciation. Coordinated evolution between nuclear and organelle genomes can occur by reciprocal changes in the functional constraints of interacting proteins, and even playing an important role in speciation. In this master's thesis, we used target-capture sequencing data from 122 individuals from two different mitochondrial lineages to test the compensatory evolution hypothesis by comparing the phylogenetic trees of three diferent types of genes: mtDNA, nuclear genes interacting with mitochondrias (oxphos genes) or random nuclear genes. Our results showed the two different mitochondrial lineages. Tajima's D test also showed mitochondria to be under strong purifying selection. However, differences were not observed between the oxphos and the nuclear genes. More studies should be carried out to confirm or not the compensatory evolution hypothesis in this lizard species with the data produced in this study.

Índice

Lista de figuras

1. Introducción

1.1 Contexto y justificación del Trabajo

Coevolution is the process of reciprocal evolutionary change between interacting species (Futuyma, 1998; Thompson, 2001). This process is thought to be a major driver of biological diversification (Thompson, 2005) and has been linked to speciation (Hembry et al. 2014). Multiple examples of coevolution have been reported (e.g. predator-prey relationships in Brodie III et al. 2005; host-parasite relationships in Dybdahl & Lively, 1998). Although coevolution has been mainly focused at the species level, several authors (e.g. Rand et al. 2004) argue that much of nuclear-organelle evolution fits this definition (with genomes rather than species). In fact, coordinated evolution between nuclear and organelle genomes can occur by reciprocal changes in the functional constraints of interacting proteins (Rand et al. 2004), and even playing an important role in speciation (Gershoni et al. 2009).

Mitochondria is responsible for 95% of the eukaryotic cell's energy, having its general structure relatively conserved across vertebrates (Cooper, 2000). Its ca. 13 protein-coding genes encode multiple subunits of four of the five enzyme complexes and play an important role in oxidative phosphorylation (OXPHOS) and other cellular bioenergetic pathways (Taanman, 1999; Lajbner et al., 2018). At the same time, mitochondrial genome (mtDNA) has been widely used as a genetic marker in population genetics to reconstruct phylogenetic relationships or to infer evolutionary history of specific taxa (e.g. Brown, 2002). As a marker, it contains some interesting features that cannot be found in nuclear DNA, as a higher rate of nucleotide substitutions, maternal inheritance and almost no recombination (Moritz et al., 1987). An important remark is that mtDNA has been assumed to evolve under neutrality (e.g. Moritz et al., 1987; Avise et al., 1994). However, pressures shaping mitochondria's evolution are still unclear. Recent studies affirm that mtDNA appears to be anything but a neutral marker and probably undergoes frequent adaptive evolution, e.g., direct selection on the respiratory machinery (Ballard & Whitlock, 2004; Bazin et al., 2006; Lajbner et al., 2018). Moreover, as mtDNA replicates asexually and typically without recombination, mtDNA is vulnerable to the accumulation of deleterious mutations via Muller's ratchet (Muller, 1964), producing a mutational meltdown. In order to avoid that, strong purifying selection, which effectively removes functional mutations on mtDNA, has been observed in several taxa (e.g. Morales et al., 2015; Jacobsen et al., 2016). However, seems that purifying selection cannot be the only cause, as empirical studies indicate that deleterious mutations do accumulate in mtDNA (e.g. van der Sluis et al. 2015). Compensatory coevolution has been proposed to compensate the accumulation of deleterious mutations in the mtDNA without a loss of function of the OXPHOS processes (Rand et al. 2004). From all the genes that are present in an organism, around 1200 genes are transported to the mitochondrion, and are just 150 nuclear genes (from now on oxphos genes) the ones engaging in close functional association with mtDNA and interacting in oxphos processes (Calvo et al. 2016). These processes are involved in cellular respiration and other cellular bioenergetic pathways (Taanman, 1999; Cooper, 2000; Lajbner et al.

2018). This mitonuclear compensatory coevolution hypothesis proposes that oxphos genes interacting with mtDNA will experience a strong selection to evolve novel features, enabling them to reduce or nullify the malfunctions that would be caused by deleterious alleles from mtDNA (Rand et al. 2004; Hill, 2020) and to follow coordinated evolutionary trajectories (Rand, et al. 2004; Lane, 2011).

The Iberian and North African wall lizards (genus *Podarcis*) is a cryptic species complex that has been studied using mtDNA (Harris & Sá-Sousa, 2001; Harris & Sá-Sousa, 2002; Harris et al. 2002; Pinho et al. 2006) and nuclear data(Pinho et al. 2003; Pinho et al. 2004; Pinho et al. 2007; Pinho et al. 2008). Both types of markers have confirmed the existence of multipled highly differentiated lineages, corresponding to currently different accepted species (e.g. *P. bocagei, P. carbonelli* or *P. vaucheri*) whilst, from these studies, the taxonomic situation of other lineages, as *P. liolepis*, were less clear (Pinho et al. 2008). In Busack et al. (2005) proposed to treat the north-eastern Spanish form of *P. hispanica*, currently being *P. liolepis* (Renoult et al. 2010), as a different species. The genetic characterisation of this species has been based on individuals from Barcelona, Girona, Tarragona and the central southern Pyrenees (Harris and Sá-Sousa, 2002), Burgos and Medinaceli (Pinho et al. 2006), Andorra (Harris et al. 2002) and southern France (unpublished data from Renoult et al. 2010). This lineage corresponds to a mitochondrial lineage called "*Podarcis hispanica* type 3" in Pinho et al. (2006) and "Liolepis" in Renoult et al. (2009). In southern populations, in the area of Valencia and surroundings, there is a *Podarcis* linage characterised by a different mitochondrial lineage, named "*Podarcis hispanica* sensu stricto" in Pinho et al. (2006) and "Valencia" in Renoult et al. (2009). According to Pinho et al. (2006), these two mitochondrial lineages diverged several million years ago, but it was not clear if these two lineages correspond to two different evolutionary units (Renoult et al. 2010). Although in Pinho et al. (2008) confirmed that there was an absence of nuclear gene flow between these two lineages, previously in Pinho et al. (2007) individuals from these two lineages were grouped in a single nuclear cluster. Moreover, Renoult et al. (2009) confirmed that nuclear and morphological data were concordant, whilst with the mitochondrial analyses observed that the lineage "*Podarcis hispanica* sensu stricto/Valencia" (Pinho et al. 2006; Renoult et al. 2009) introgressed populations that belong to different evolutionary units, one of them located in Valencia and surroundings, nuclearly similar to *Podarcis liolepis* from the north-eastern of the Iberian Peninsula.

Despite the building up of research suggesting coevolution between the two genomes, the specific role that these coevolutionary forces have in shaping genetic diversity in natural populations, including in promoting speciation, is still obscure. The *Podarcis* model system is appropriate for the study of oxphos genes and mtDNA interactions due to the discordance in geographical and genealogical patterns observed, due to differential lineage sorting (Pinho et al. 2008) and gene flow between the species boundaries (Pinho et al. 2007; Renoult et al. 2009). These discordances provide invaluable opportunities to evaluate whether evolution of oxphos genes are more likely to correlate with mtDNA evolution than any other nuclear gene. In this study, we tested the compensatory evolution hypothesis by comparing the coordinated evolutionary

trajectories in three different datasets: the whole mtDNA, a set of oxphos genes and a third set of nuclear genes not involved in oxphos processes (from now on nuclear genes). To do so, we used targeted capture data of the genomic regions of interest (mitochondrial, oxphos and nuclear genes) for individuals from both mitochondrial lineages to 1) obtain the individual haplotypes, 2) observe signatures of natural selection in both genomes and 3) studying the coordinated evolutionary trajectories of the three datasets within a phylogenetic framework. A higher mutation rate is expected from the oxphos genes when compared with the nuclear genes. Finally, as the studied species presents two different mitochondrial lineages, we assume that oxphos genes will follow a coordinated evolutionary trajectory with the corresponding mtDNA lineage, whilst we expect a random pattern between both lineages in the nuclear genes.

1.2 Objetivos del Trabajo

Objective 1: Process NGS data from Target-Capture Sequencing:

- 1.1 Verify the quality of fastQ and VCF files.
- 1.2 Align the fastQ files to the target genes
- 1.3 Check if target genes were correctly amplified.
- 1.4 Haplotype obtention for data analysis.

Objective 2: Study the coevolution between oxphos and mitochondrial genes:

- 2.1 Observe signatures of natural selection in both genomes.
- 2.2 Analyse the amino acid substitution patterns depending if they take part in genomes interaction or not.

Objective 3: Explore the genetic consequences at the population level of the cytonuclear interactions:

- 3.1 Perform phylogenetic comparison between the three datasets.

1.3 Enfoque y método seguido

Species of study

Podarcis liolepis is one of the species described from the Podarcis hispanica species complex (Renoult et al. 2010). Although there has been a lot of controversy about its distribution, currently can be found in the south of France and north-east of Spain (Fig. 1). This species seems to be nuclear and morphologically uniform (Pinho et al. 2007: Renoult et al. 2009) but with two different mitochondrial lineages (Renoult et al. 2009). According to them, the most likely scenario to explain this discordance is ancient mitochondrial introgression originating from an evolutionary unit absent from the study area.

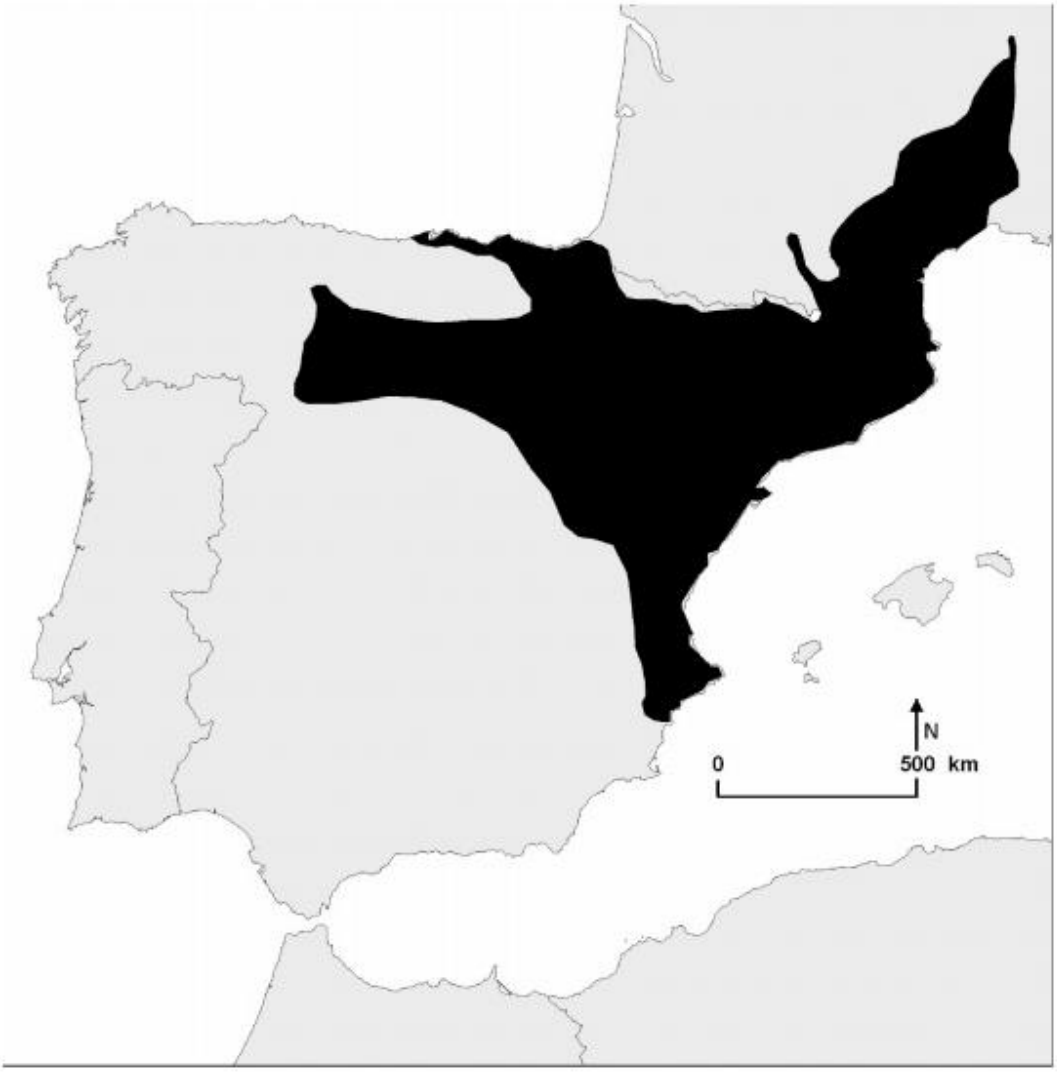

Fig. 1: Distribution of *Podarcis liolepis* from Renoult et al. 2010.

Sequences extraction

High quality DNA was extracted from tail tissue using standard protocols. Sequence data was obtained using Illumina sequencing of genomic regions enriched by hybridization on microarrays according to the method outlined by Hodges et al. (2009). Target-captured sequencing is the proper technique for this study as we are interested in sequence high number of samples of specific genes. Agilent SureSelect array customized was used for the selective enrichment of the target genes (i.e. mtDNA, oxphos and nuclear genes) and individual tagging was carried out (Meyer & Kitcher, 2010). A total of 100 genes were sequenced for both oxphos and nuclear sets. Nuclear genes were chosen to be similar in size and in number of exons with oxphos genes.

A total of 122 individuals (including one individual of *Podarcis muralis* used as an outgroup) from 52 different populations (Fig. 2) were analysed from both mitochondrial lineages. FastQ files were obtained and analysed following GATK best practices (https://gatk.broadinstitute.org/hc/en-us). First of all, both paired files were converted into an unmapped bam file with the *FastqToSam* command, adding a sample name and a read group name into each unmapped file. With the *SortSam* command, unmapped files were sorted by queryname and Illumina adapters were marked with *MarkIlluminaAdapters* function. *SamToFastq* command was used to obtain the final fastq files. After quality check of the Fastq files with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), files were mapped against the target genes with *bwa* and sorted again by queryname with the *SortSam* command. Using the *MergeBamAlignment*, mapped and unmapped files per sample were merged, adding information as the sample name or the read group that was lost during the mapping process. Duplicates were marked with *MarkDuplicates* and each sample file was indexed with *index*. For each sample, the *Haplotypecaller* command was applied and a GVCF file was created. This file is necessary in GATK best practices to get the final VCF. With *CombineGVCFs* all the GVCFs files were combined and the final VCF file was obtained with *GenotypeGVCFs*. VCF file was filtered using vcftools. A minimum coverage was set to 10, mean minimum coverage to 5, maximum mean coverage to 60, maf to 0.05, and maximum missing to 0.7. Phasing for the final VCF file was carried out with the software Beagle and the phased VCF was divided and indexed per sample. Finally, fasta files of the target genes per sample were obtained with *FastaAlternateReferenceMaker*. Each gene was separated in a different fasta file using linux commands and were independently aligned using MAFFT (https://mafft.cbrc.jp/). Genes were concatenated in three different supermatrixs with the program FASconCAT (from the Zoological Research Museum Alexander Koenig): a first one for mitochondrial genes, a second one for oxphos genes and a third one for nuclear genes (See Anexos codes from 1 to 5).

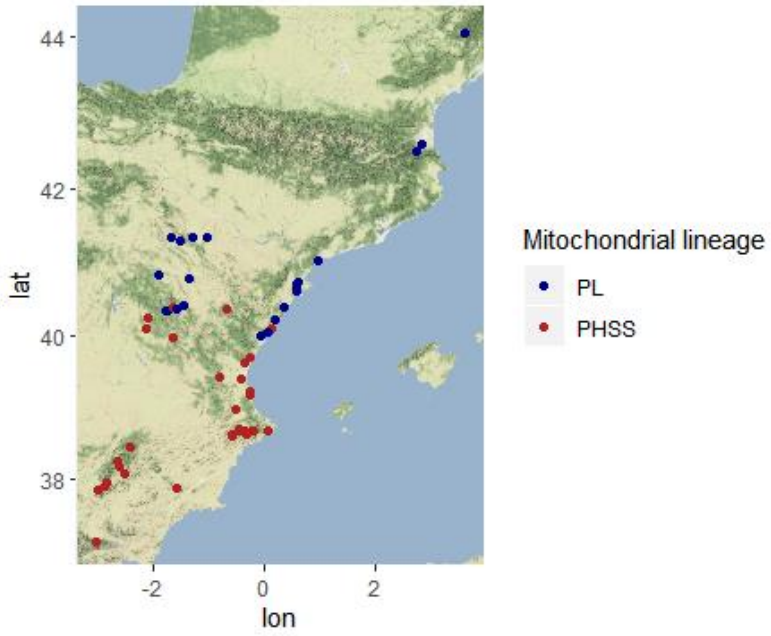

Fig. 2: Map for the 52 sampled sites. Each mitochondrial lineage is depicted in a different colour.

Analyses

To determine from which mitochondrial lineage the sampled individuals were, a phylogenetic tree with CYTB mitochondrial gene was calculated for all the samples. For phylogenetic comparisons, a tree for each dataset containing all the sequenced genes was carried out. All phylogenetic analyses were produced in IQ-TREE (Minh et al. 2020), a phylogenomic inference software implemented in CIPRESS Gateway (Miller et al. 2010). All phylogenetic trees were done with an edge-equal partition model, were tested to choose the best model and ultrafast bootstrap was set to 1000 replicates (Hoang et al. 2018). Moreover, to prove the pattern observed in the phylogenetic trees, a PCA with each of the VCF datasets was calculated using Plink (Anexo, Code 6). Finally, Tajima's D test were computed for some of the genes.

1.4 Planificación del Trabajo

This work is mainly computational, so the access to a cluster is mandatory. All the work was carried out in the Totoro cluster from CIBIO. Each one of the objectives mentioned previously will be a task, which will be explained below:
For the objective 1.1, tasks related with data quality were done for both fastQ and VCF files. Task 1.2 was realized with the function bwa, aligning the fastQ files to the target genes. Task 1.3 was to assess the coverage of the target genes. Task 1.4 was the last one of its group and was about the obtention of the haplotypes. Task 2.1 was applied with the method FUBAR for detecting purifying selection at the genome level and task 2.2 was not done. Finally, task 3.1 was done with the software IQ-tree through CIPRESS. The timing of each task can be observed in Fig. 2.

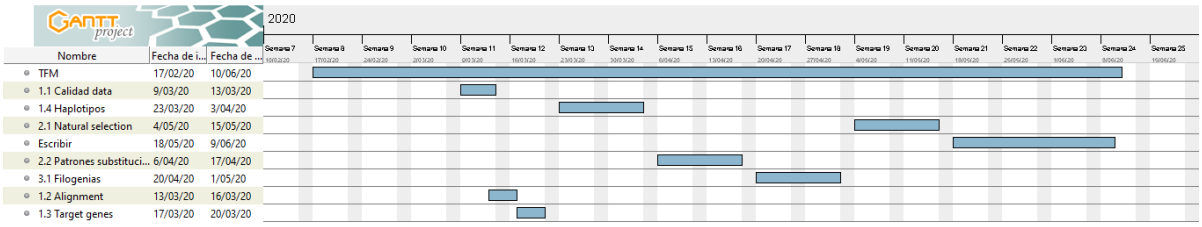

Fig. 3: Timeline of the tasks.

1.5 Breve sumario de productos obtenidos

The biggest product I obtained from this master's thesis is the knowledge I gained working with a server. I started working with shell in a subject during the master program, but after all these months I feel confident in working and interacting with a server and the shell language. A part of that, the most important product I obtained are the data and the pipeline I developed during this study, both for me and the research group. On the other hand, I would mention from the results the different patterns observed in the three different genes sets and

Resultados

Mitochondrial lineages

Phylogenetic analyses using CytB gene resulted in a phylogenetic tree with two mitochondrial clusters, as it was expected (Fig. 3) and a proper classification of the individuals without a mitochondrial lineage was done (Anexos). In Fig. 2 we can see the distribution of both lineages with the updated information, with two areas, one in the coast and a second one in the mainland, where both lineages are present.

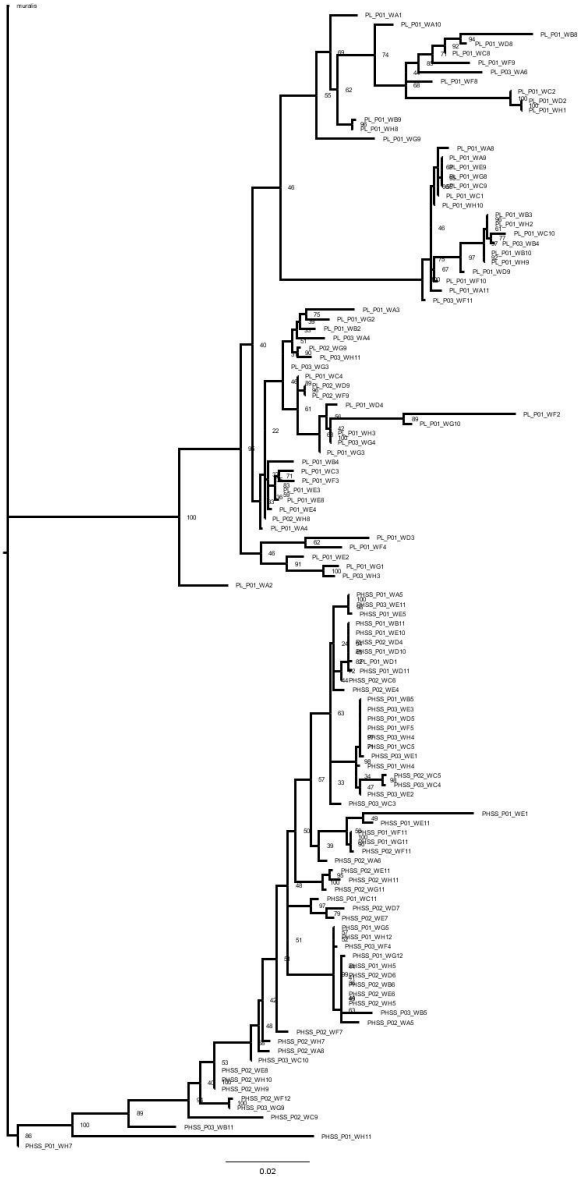

Fig. 4: Phylogenetic tree for the gene CytB. Boostrap values are indicated.

Mitochondrial genes

Phylogenetic analyses for all the 13 protein-coding genes plus the D-loop located in the mtDNA resulted in a phylogenetic tree with both mitochondrial lineages well divided (Fig. 4).

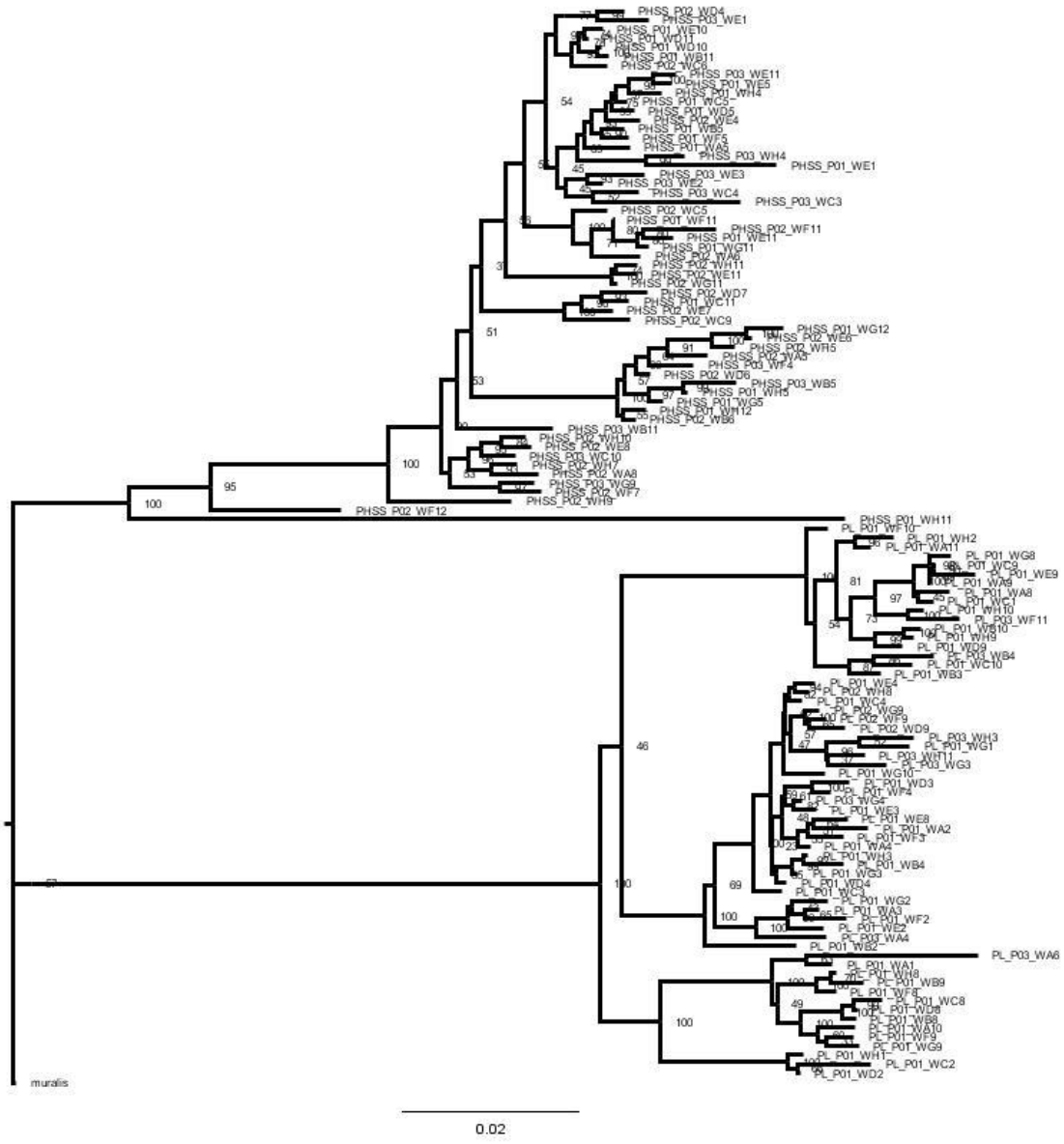

Fig. 5: Phylogenetic tree for all the mitochondrial genes. Bootstrap values are indicated.

Oxphos genes

Phylogenetic analyses for all the oxphos genes gave a phylogenetic tree that mainly separated the two mitochondrial lineages (Fig. 5). However, some individuals from both mitochondrial lineages were found to be between the outgroup and the two main clusters.

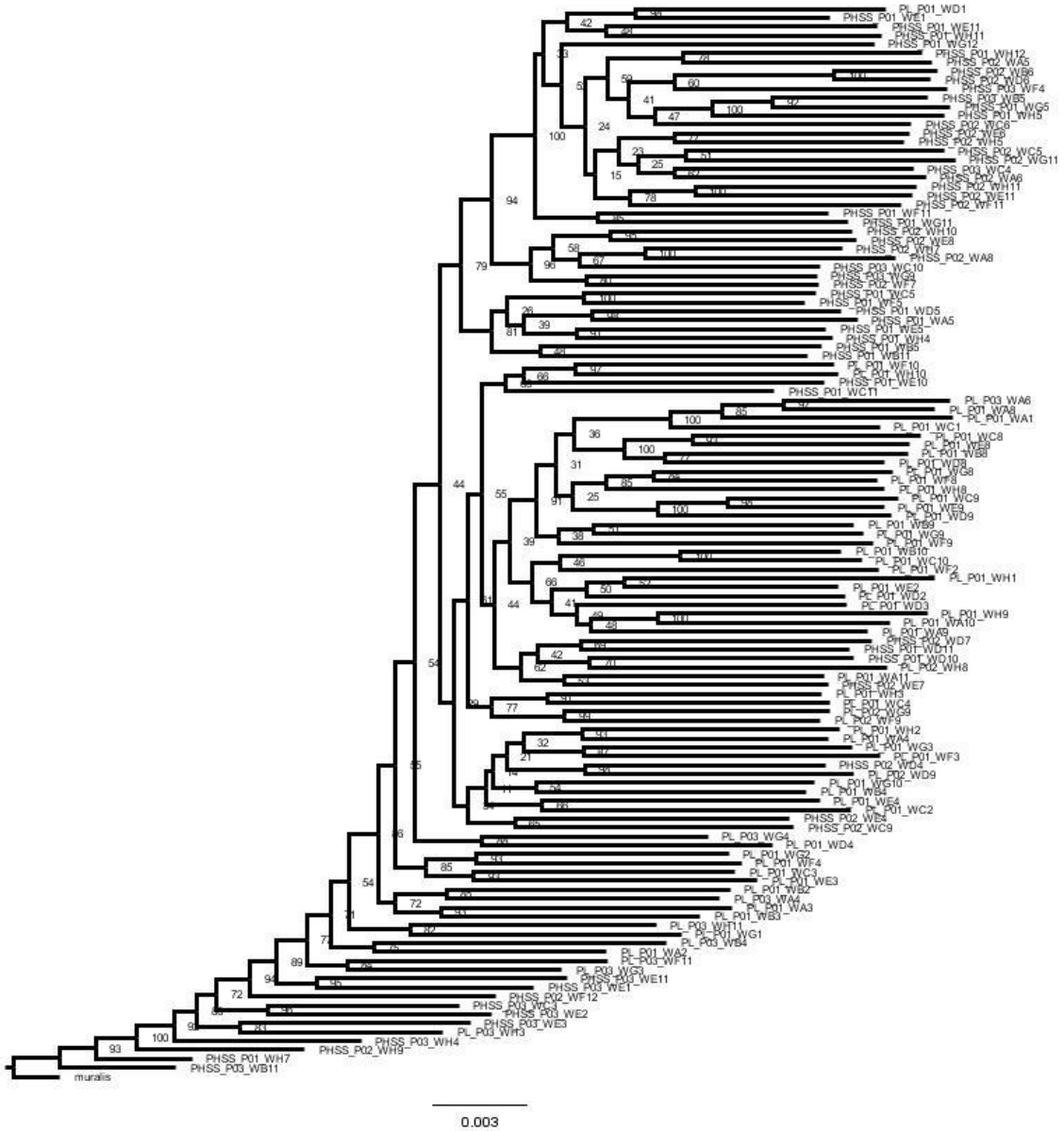

Fig. 6: Phylogenetic tree for the oxphos genes. Bootstrap values are indicated.

Nuclear genes

Phylogenetic analyses for all the nuclear genes gave a phylogenetic tree with two main clusters corresponding to the two mitochondrial lineages (Fig. 6). As in Fig. 5, some individuals from both mitochondrial lineages were found to be between the outgroup and the two main clusters.

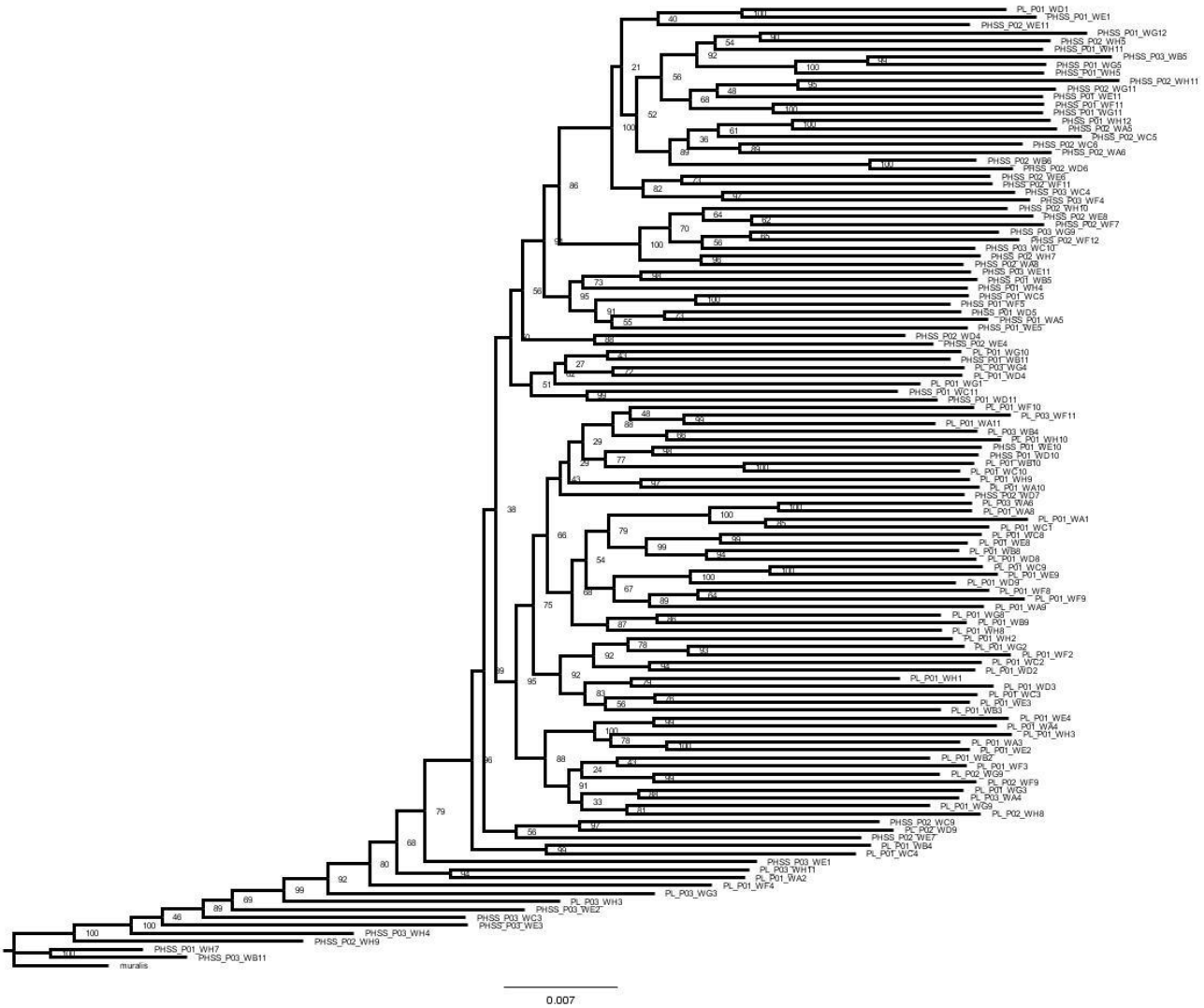

Fig. 7: Phylogenetic tree for the nuclear genes. . Bootstrap values are indicated.

PCA:

Results from the PCA analyses showed similar patterns to the ones observed in the phylogenetic trees. With the mtDNA dataset, both lineages were clearly separated (mitochondrial genes in Fig. X). On the other hand, both oxphos and nuclear genes displayed a similar pattern. Some individuals from both mitochondrial lineages were spatially separated, but at the same time, other individuals were sharing a similar an area in the PCA plot.

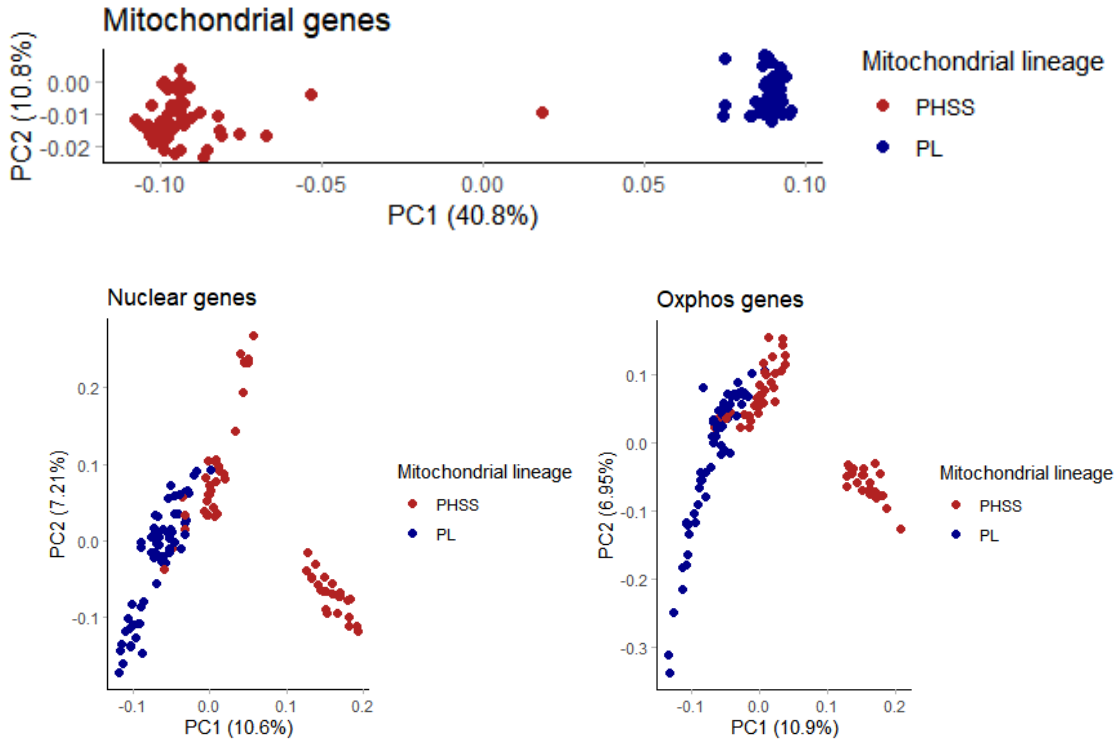

Fig. 8: PCA with SNPs for the three datasets.

Tajima's D test

Results from the Tajima's D test in a small number of mitochondrial genes showed positive values for the complete dataset whilst negative values when both lineages were studied independently (Table X).

Table 1: Tajima's D test for three mitochondrial genes. Significant p-values are in bold.

	Complete dataset		PL lineage		PHSS lineage	
Genes	**Tajima's D**	**p-value**	**Tajima's D**	**p-value**	**Tajima's D**	**p-value**
Atp6	0.28	0.78	-0.32	0.75	-1.04	0.30
Cox1	0.52	0.6	-0.49	0.62	-1.02	0.30
ND1	0.21	0.83	-0.09	0.92	-1.11	0.26

3. Discusión y conclusión

In this master's thesis, the compensatory evolution hypothesis was tested using *Podarcis liolepis* as a model species. The compensatory evolution hypothesis in relation with mitochondrial evolution says that at the DNA sequence level, higher evolutionary rates of mitochondrial genes could drive accelerated evolutionary rates of oxphos genes as compensatory response to the maintenance of function (Hill et al. 2020). For example, this has been observed for protein-coding genes that form oxphos complexes responsible for electron transport (Rand et al. 2004; Willett and Burton 2004; Mishmar et al. 2006;) and ribosomal proteins that must interact with mtDNA-encoded rRNAs (Barreto and Burton 2013; Sloan et al. 2014). Furthermore, there is evidence of coevolution between nuclear-encoded tRNA synthetases and mtDNA-encoded tRNAs (Hoekstra et al. 2013; Meiklejohn et al. 2013).

Our results showed two divergent mitochondrial lineages present in the studied samples (Fig. 4 and 5). This result was expected, as these two mitochondrial lineages were previously described for the localities but helped to properly identify all individuals (Anexos). Phylogenetic results with mtDNA also showed the two mitochondrial lineages observed previously with CytB (Fig. 5). This tree was produced to have a confirmation with more genes of the pattern observed in Fig. 4. A higher support was accomplished using all the mtDNA available, having higher bootstrap values. However, phylogenetic results from both datasets, oxphos and nuclear genes, showed really similar patterns (Figs 6 and 7). These results were not expected as, according to the compensatory evolution hypothesis, just oxphos genes should follow mitochondria's evolution. In fact, this was observed, as oxphos genes were clustering according to the mitochondrial lineages (Fig. 6), but a similar pattern was observed for the nuclear genes (Fig. 7). Nuclear genes should have a lower evolution rate and should cluster randomly. Similar results were present also in the PCA plots, were mtDNA clustered independently and again, a similar pattern was observed between oxphos and nuclear genes (Fig. 8). Interestingly, previous analyses were showing oxphos genes clustering in a similar way than the mtDNA and nuclear genes randomly ordered (Data not shown), suggesting that within the big dataset we are showing in here, different genes could follow different evolutionary pathways. However, the results presented in this master thesis are mostly qualitative and not quantitative. Due to the lack of time, more quantitative analyses were not possible to carry out. Cline analyses or comparisons of synonymous and non-synonymous substitution rates as in Zhang & Broughton (2013) will be done and will give us a better inside of the data obtained in this master's thesis.

Tajima's D test were carried out for the mitochondrial genes, and an interesting pattern was observed. When the complete dataset was analysed (i.e. not separated by lineage), observed Tajima's D values were higher than expected. However, when the dataset was analysed per lineage, Tajima's D values became negative. As studied individuals are from two different lineages, when the whole dataset is analysed, the variability observed by joining both lineages is higher, and we obtain higher values of Tajima. When both lineages are studied separately, this effect is removed, and negative values, indicating strong

purifying selection, are found. These results were expected, as previous studies already reported in other taxa (e.g. Morales et al., 2015; Jacobsen et al., 2016) and confirms that the mitogenome of this species is under strong purifying selection.

Carrying out this master's thesis has been more difficult than expected, due to the general current situation but also for working with servers and NGS data for the first time. Furthermore, the long time spent in obtaining the data jeopardized all the other parts of the master's thesis, especially the analyses part. However, the author of this master's thesis is happy with the achievement of the data and the pipeline that is presented here, as it opens the door to several interesting hypothesis that will be tested in the near future. The obtention of the data itself, more in a bioinformatic's master, is already a good result.

Several methods that were mentioned in the original plan (selection analyses with HyPhy and CodeML) were finally not applied. The internal tutor already advised about this, as she mentioned it in some of the meetings. However, thanks to her useful suggestions, at least the Tajima's D test was applied. Being the final goal of this study to obtain the individual haplotypes, which is a less common approach, and without prior experience on this topic, help was hard to find in the main forums. However, useful comments from both tutors helped to achieve the final data.

Future research is already planned and several analyses, as phylogenetic clines or the study of the mutation rates of the different datasets will be carried out soon. Thanks to the pipeline designed here, we can also obtain data from other species that were sequenced together with the data analysed here.

5. Bibliografía

Avise, J. C. (1994) Molecular markers, natural history, and evolution. New York, NY: Chapman & Hall.

Ballard, J. W. & Whitlock, M. C. (2004) The incomplete natural history of mitochondria. Mol. Ecol..

Bazin, E., Glemin, S. & Galtier N. (2006) Population size does not influence mitochondrial genetic diversity in animals. Science. 312, 570–572.

Brodie III, E.D., Feldman, C.R., Hanifin, C.T., Motychak, J.E., Mulcahy, D.G., Williams, B.L. & Brodie Jr. E.D. (2005). Parallel Arms Races between Garter Snakes and Newts Involving Tetrodotoxin as the Phenotypic Interface of Coevolution. Journal of Chemical Ecology volume **31**, 343–356.

Brown, T. A. Genomes. Oxford: Wiley-Liss; Molecular Phylogenetics. (2002).

Busack, S.D., Lawson, R. & Arjo, W.M. (2005). Mitochondrial DNA, allozymes, morphology and historical biogeography in the Podarcis vaucheri (Lacertidae) species complex. Amphiba–Reptilia, **26**, 239–256.

Calvo, S.E. et al. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins Nucleic Acids Res. **44**, 1251-1257.

Cooper G. M. The Cell: A Molecular Approach. Sunderland (MA): Sinauer Associates; Mitochondria. (Sinauer Associates: Sunderland, MA, 2000).

Dybdahl, M.F. & Lively, C.M. (1998). Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. Evolution, **52**(4), 1057-1066

Futuyma, D. (1998) Evolutionary Biology. Sinauer Associates

Lane, N. (2011). Mitonuclear match: optimizing fitness and fertility over generations drives ageing within generations. BioEssays. **33**, 860–869.

Gershoni, M., Templeton, A.R., and Mishmar, D. (2009). Mitochondrial bioenergetics as a major motive force of speciation. Bioessays 31, 642–650.

Harris, D.J. & Sá-Sousa, P. (2001). Species distinction and relationships of the western Iberian Podarcis lizards (Reptilia, Lacertidae) based on morphology and mitochondrial DNA sequences. Herpetol J. **11**,129-136.

Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C. & Ferrand, N. (2002). Complex biogeographical distribution of genetic variation within Podarcis wall lizards across the Strait of Gibraltar. J Biogeogr. **29**,1257-1262.

Harris, D.J. & Sá-Sousa, P. (2002). Molecular phylogenetics of Iberian wall lizards (Podarcis): is Podarcis hispanica a species complex? Mol Phylogenet Evol. **23**, 75-81.

Hill, G.E. (2020). Mitonuclear Compensatory Coevolution Trends in Genetics. **36**(6), 403-414.

Hembry DH, Yoder JB, Goodman KR (2014) Coevolution and the diversification of life. The American Naturalist, 184, 425–438.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Gordon, D.B., Brizuela, L., McCombie, W.R., and Hannon, G.J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature Protocols 4, 960–974.

Jacobsen, M. W., da Fonseca, R., Bernatchez, L. & Hansen, M. M. (2016). Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (Coregonus ssp.). Mol Phylogenet Evol. **95**, 161-170.

Minh, B.Q., Schmidt, H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**, 1530-1534.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc.

Morales, H. E., Pavlova, A., Joseph, L. & Sunnucks, P. (2015). Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol Ecol. 24(11), 2820-2837.

Moritz, C., Dowling, T. E. & Brown, W. M. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annu Rev Ecol Syst. 18, 269–92 (1987).

Muller, H.J. (1964) The relation of recombination to mutational advance Mutat. Res. **1**, 2-9.

Pinho, C., Harris, D.J., Ferrand, N. (2003). Genetic polymorphism of 11 allozyme loci in populations of wall lizards (Podarcis sp.) from the Iberian Peninsula and North Africa. Biochem Genet. **41**, 343-359.

Pinho C., Ferrand, N., Harris, D.J. (2006). Reexamination of the Iberian and North African Podarcis (Squamata: Lacertidae) phylogeny based on increased mitochondrial DNA sequencing. Mol Phylogenet Evol. **38**, 266-273.

Pinho, C., Harris, D.J., Ferrand, N. (2007). Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (Podarcis, Lacertidae). Biol J Linn Soc. **91**, 121-133.

Pinho C, Ferrand N, Harris DJ: Genetic variation within the Podarcis hispanica species complex: new evidence from protein electrophoretic data. In The Biology of Lacertid Lizards: Evolutionary and Ecological Perspectives Volume.

Pinho, C. Harris, D.J. & Ferrand, N. (2008). Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (Podarcis spp.) are an assemblage of incipient species. BMC Evolutionary Biology. **8,** 63.

Rand, D.M., Haney, R.A., Fry, A.J. (2004). Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology and Evolution. **19**(12), 645-653.

Renoult, J.P., Geniez, P., Bacquet, P., Benoits, L. Crochet, P-A. (2009). Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. **18**, 4298–4315.

Renoult, J.P., Geniez, P., Bacquet, P. Guillaume, C.P. & Crochet, P-A. (2010). Systematics of the Podarcis hispanicus-complex (Sauria, Lacertidae) II: the valid name of the north-eastern Spanish form. Zootaxa **2500**, 58–68.

Taanman J.W. The mitochondrial genome: Structure, transcription, translation and replication. Biochim Biophys, 103-123 (1999).

Thompson, JN (2001). Coevolution. in: Encyclopedia of Life Sciences, London, Nature Publishing Group.

Thompson JN (2005) The Geographic Mosaic of Coevolution. University of Chicago Press, Chicago, Illinois.

van der Sluis, E.O., et al. (2015). Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes Genome Biol. Evol., **7**, 1235-1251

Zhang, F., Broughton, R.E. Mitochondrial–Nuclear Interactions: Compensatory Evolution or Variable Functional Constraint among Vertebrate Oxidative Phosphorylation Genes? (2013). Genome Biology and Evolution. **5**(10), 1781–1791.

6. Anexos

Due to the features of the server, a task was not able to run for more than 24h in the server. For this reason, all the steps are split:

Code 1: Preparing the FastQ files

In GATK, before doing the SNP calling, Illumina adapters must be marked and the read group specified. In this code, sample and read group name are specified in both FASTQ files and converted into a bam:

java -jar ~/softwares/picard/picard.jar FastqToSam FASTQ= 1.fastq.gz FASTQ2=2.fastq.gz OUTPUT=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_unmapped.bam SAMPLE_NAME=HL5_P01_WA01_i5-505_i7-59_S97_L001 READ_GROUP_NAME=HL5_P01_WA01_i5-505_i7-59_S97_L001

The unmapped bam is sorted:
java -jar ~/softwares/picard/picard.jar SortSam I=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_unmapped.bam O=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_unmapped_sorted.bam SORT_ORDER=queryname

Illumina adapters are marked:
java -jar ~/softwares/picard/picard.jar MarkIlluminaAdapters I=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_unmapped_sorted.bam O=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_MarkIlluminaAdapter.bam METRICS=metrics.txt

And the file is converted again to FastQ:
java -jar ~/softwares/picard/picard.jar SamToFastq I=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_MarkIlluminaAdapter.bam FASTQ=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_fastq.fq CLIPPING_ATTRIBUTE=XT CLIPPING_ACTION=2 INTERLEAVE=true NON_PF=true

Code 2: SNP calling

FastQ files are mapped against the target regions (i.e. mitochondrial, oxphos and nuclear genes):
bwa mem -M -t 20 -p target_genes.fasta HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_fastq.fq > HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001_aligned.bam

Data is sorted by queryname:
java -jar ~/softwares/picard/picard.jar SortSam
I=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_aligned.bam O=HL5T3BBXX_POR_100801_P01_WA01_i5-
505_i7-59_S97_L001_aligned_sorted.bam SORT_ORDER=queryname

Aligned files lost some information in the mapping process. For this reason, now we merge them with the unmapped bam file:
java -jar ~/softwares/picard/picard.jar MergeBamAlignment
R=target_genes.fasta
UNMAPPED_BAM=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_unmapped_sorted.bam
ALIGNED_BAM=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_aligned_sorted.bam
O=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_merged.bam CREATE_INDEX=true ADD_MATE_CIGAR=true
CLIP_ADAPTERS=true CLIP_OVERLAPPING_READS=true
INCLUDE_SECONDARY_ALIGNMENTS=tr

And we mark the duplicates:
java - -jar ~/softwares/picard/picard.jar MarkDuplicates
INPUT=HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_merged.bam O=HL5T3BBXX_POR_100801_P01_WA01_i5-
505_i7-59_S97_L001_markduplicates.bam METRICS_FILE=metrics.txt

We create the index file with samtools:
samtools index HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_markduplicates.bam

We create a GVCF file:
java -jar ~/softwares/GATKK/1gatk/gatk-package-4.1.7.0-local.jar
HaplotypeCaller --reference target_genes.fasta --input
HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-
59_S97_L001_markduplicates.bam --output
HL5T3BBXX_POR_100801_P01_WA01_i5-505_i7-59_S97_L001.g.vcf -ERC
GVCF

And we combine all the GVCF files:

java -jar ~/softwares/GATKK/1gatk/gatk-package-4.1.7.0-local.jar
CombineGVCFs --reference target_genes.fasta --variant
/mnt/CIBIO/homes/gabri.mochales/analisis_transect_project/data/6haplotypecal
ler/RAPiD-Genomics_HL5T3BBXX_POR_100801_P03_WH12_i5-507_i7-
56_S480_L002.g.vcf --variant
/mnt/CIBIO/homes/gabri.mochales/analisis_transect_project/data/6haplotypecal
ler/RAPiD-Genomics_HL5T3BBXX_POR_100801_P03_WH11_i5-505_i7-
95_S479_L002.g.vcf --variant
/mnt/CIBIO/homes/gabri.mochales/analisis_transect_project/data/6haplotypecal
ler/RAPiD-Genomics_HL5T3BBXX_POR_100801_P03_WH10_i5-507_i7-
34_S478_L002.g.vcf --output

/mnt/CIBIO/homes/gabri.mochales/analisis_transect_project/data/7combined/combined.g.vcf

Finally, we obtain the SNP calling file:
java -jar ~/softwares/GATKK/1gatk/gatk-package-4.1.7.0-local.jar GenotypeGVCFs -R target_genes.fasta -V combined.g.vcf -O combined.vcf

Code 3: Phasing

The phasing was done with Beagle:
(https://faculty.washington.edu/browning/beagle/b3.html)

java -jar ~/softwares/beagle/beagle.24Mar20.5f5.jar gt=combined.vcf out=phased.gt

Code 4: Fasta files:

As we have one file containing all the samples and the function FastaAlternateReferenceMaker makes the fasta files per gene for just one sample, we need to split the vcf file with this code:

```
for file in *.vcf*; do
  for sample in `bcftools query -l $file`; do
    bcftools view -c1 -Oz -s $sample -o ${file/.vcf*/.$sample.vcf} $file
  done
done
```

With FastaAlternateReferenceMaker function, the final fasta files were obtained for each gene and each sample:

java -jar ~/softwares/GATKK/1gatk/gatk-package-4.1.7.0-local.jar FastaAlternateReferenceMaker --reference target_genes.fasta -O HL5T3BBXX_POR_100801_P03_WH12_i5-507_i7-56_S480_L002.fasta -V /mnt/CIBIO/homes/gabri.mochales/softwares/GATKK/1gatk/analisis/genome/todo/phased/phased_ HL5T3BBXX_POR_100801_P03_WH12_i5-507_i7-56_S480_L002.gt.vcf

Code 5: Gene extraction

For each sample, each gene was independently extracted using linux commands:
awk '/'ND1'/{flag=1;print $0;next}/^>/{flag=0}flag' RAPiD-Genomics_HL5T3BBXX_POR_100801_P0* > MIT/muralis_ncbi/ND1.fasta

And was aligned using MAFFT:
~/softwares/mafft/mafft-linux64/mafft.bat ND1.fasta > aligned/aligned_ND1.fasta

Finally, the perl based program FASconCAT (https://www.zfmk.de/en/research/research-centres-and-groups/fasconcat) was used to create a supermatrix for each one of the three datasets.

Code 6: PCA

```
plink --vcf combined_filtered.vcf.gz --allow-extra-chr --make-bed –double-id --
out pca

plink --bfile  --double-id --allow-extra-chr --set-missing-var-ids @:# \
--make-bed --pca –out pca
```

And in R:

```
library(tidyverse)
library(ggplot2)

pca <- read_table2("pca_ran.eigenvec", col_names = FALSE)
eigenval <- scan("pca_ran.eigenval")

pca <- pca[,-1]

names(pca)[1] <- "ind"

names(pca)[2:ncol(pca)] <- paste0("PC", 1:(ncol(pca)-1))
col <- read.table("col.txt", header =T)


pca <- as.tibble(data.frame(pca, col$lineage))
names(pca)[21] <- "col"
pve <- data.frame(PC = 1:20, pve = eigenval/sum(eigenval)*100)
a <- ggplot(pve, aes(PC, pve)) + geom_bar(stat = "identity")
a + ylab("Percentage variance explained") + theme_light()
b <- ggplot(pca, aes(PC1, PC2, label=ind,col=col.lineage)) + geom_point(size =
2)
b <- b + scale_colour_manual(values = c("orange", "blue", "brown", "yellow",
"violet"))
b <- b + coord_equal() + theme_light()
b + xlab(paste0("PC1 (", signif(pve$pve[1], 3), "%")) +
  ylab(paste0("PC2 (", signif(pve$pve[2], 3), "%"))+
  theme(panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(size = 0.5, linetype = "solid",
                                 colour = "black"))+
  scale_color_manual("Mitochondrial lineage",values=c("PL"="dark blue",
"PHSS"="firebrick"))+ggtitle("Nuclear genes")
```

Individuals properly assigned to the mitochondrial lineages and location.

CODE	MTDNA	lat	lon
BEV.10936	PL	42.578	2.8483
BEV.11057	PL	42.5031	2.7541
BEV.12841	PHSS	39.4291	-0.7815
BEV.12918	PHSS	39.4291	-0.7815
BEV.1801	PL	41.352	-1.036
BEV.1802	PL	41.352	-1.036
BEV.1803	PL	41.352	-1.036
BEV.1807	PL	41.339	-1.269
BEV.1809	PL	41.289	-1.516
BEV.1810	PL	41.289	-1.516
BEV.1812	PL	41.289	-1.516
BEV.1814	PL	41.289	-1.516
BEV.1831	PL	41.289	-1.516
BEV.1833	PL	41.289	-1.516
BEV.1839	PL	41.289	-1.516
BEV.1844	PL	41.289	-1.516
BEV.1849	PHSS	40.236	-2.089
BEV.1850	PL	41.353	-1.654
BEV.1851	PL	40.84	-1.89
BEV.1854	PL	40.84	-1.89
BEV.1855	PL	40.84	-1.89
BEV.1857	PL	40.84	-1.89
BEV.1859	PL	40.795	-1.36
BEV.1860	PL	40.795	-1.36
BEV.1862	PL	40.795	-1.36
BEV.1865	PL	40.408	-1.444
BEV.1866	PL	40.408	-1.444
BEV.1867	PL	40.408	-1.444
BEV.1872	PL	40.372	-1.574
BEV.1876	PL	40.3414	-1.7317
BEV.1877	PL	40.335	-1.7594
BEV.1877	PL	40.3354	-1.7594
BEV.1878	PL	40.3354	-1.7594
BEV.1879	PL	40.3354	-1.7594
BEV.1880	PL	40.3354	-1.7594
BEV.1881	PL	40.3354	-1.7594
BEV.1882	PL	40.3354	-1.7594
BEV.1884	PHSS	40.236	-2.089
BEV.1886	PHSS	40.236	-2.089
BEV.1887	PHSS	40.0901	-2.1276
BEV.1888	PHSS	40.0901	-2.1276
BEV.1889	PHSS	40.0901	-2.1276
BEV.1890	PHSS	40.0901	-2.1276
BEV.1892	PHSS	40.0901	-2.1276
BEV.1893	PHSS	40.0901	-2.1276
BEV.1894	PHSS	40.0901	-2.1276
BEV.1895	PHSS	40.0901	-2.1276

BEV.7020	PHSS	38.665	0.088
BEV.7021	PHSS	38.665	0.088
BEV.7022	PHSS	38.665	0.088
BEV.8354	PL	38.665	0.088
BEV.8357	PHSS	38.665	0.088
BEV.8360	PL	44.046	3.63
BEV.9815	PL	41.0311	0.9666
BEV.9816	PL	41.0311	0.9666
BEV.9817	PL	41.0311	0.9666
BEV.9819	PL	41.0337	0.968
BEV.9823	PL	40.6212	0.5985
BEV.9824	PL	40.6212	0.5985
BEV.9825	PL	40.6212	0.5985
BEV.9826	PL	40.6716	0.5887
BEV.9827	PL	40.6716	0.5887
BEV.9830	PL	40.7093	0.5877
BEV.9831	PL	40.7093	0.5877
BEV.9832	PL	40.7093	0.5877
BEV.9833	PL	40.7417	0.6122
BEV.9834	PL	40.7417	0.6122
BEV.9836	PL	40.3911	0.3759
BEV.9837	PL	40.3911	0.3759
BEV.9838	PL	40.2237	0.2062
BEV.9839	PL	40.2237	0.2062
BEV.9841	PHSS	40.0962	0.1346
BEV.9842	PHSS	40.0962	0.1346
BEV.9844	PL	40.0566	0.0883
BEV.9845	PL	40.0581	0.0856
BEV.9846	PL	40.0581	0.0856
BEV.9847	PL	40.0031	-0.0709
BEV.9848	PL	40.003	-0.0709
BEV.9851	PHSS	39.7054	-0.2598
BEV.9853	PHSS	39.6144	-0.3502
BEV.9856	PHSS	39.4112	-0.3954
BEV.9859	PHSS	39.2262	-0.2574
BEV.9860	PHSS	39.2195	-0.2505
BEV.9861	PHSS	39.2195	-0.2505
BEV.9863	PHSS	39.1907	-0.2438
BEV.9872	PHSS	38.6119	-0.5797
BEV.9874	PHSS	38.6255	-0.5711
BEV.9879	PHSS	38.6372	-0.3166
BEV.9881	PHSS	38.6444	-0.3194
BEV.9883	PHSS	38.6444	-0.3194
BEV.9885	PHSS	38.6716	-0.3308
BEV.9887	PHSS	38.6716	-0.3308
BEV.9889	PHSS	38.6738	-0.2303
BEV.9891	PHSS	38.676	-0.1991
BEV.9893	PHSS	38.676	-0.1991
BEV.9895	PHSS	38.6925	-0.4433
BEV.9899	PHSS	38.6865	-0.4587

1.11	PL	40.372	-1.578	
1.12	PL	40.372	-1.578	
1.14	PHSS	40.38	-1.65	
1.16	PL	40.372	-1.578	
10.53	PHSS	37.892	-2.865	
9.14	PHSS	39.985	-1.629	
9.15	PHSS	39.985	-1.629	
9.16	PHSS	39.985	-1.629	
9.24	PHSS	39.985	-1.629	
9.26	PHSS	39.985	-1.629	
9.38	PHSS	38.986	-0.52	
9.41	PHSS	38.986	-0.52	
9.45	PHSS	38.986	-0.52	
9.46	PHSS	38.986	-0.52	
DB11239	PHSS	37.113	-3.033	
DB13474	PL	40.369	-0.671	
DB13475	PHSS	40.369	-0.671	
DB13476	PHSS	40.369	-0.671	
DB1735	PHSS	37.839	-2.975	
DB1834	PHSS	38.445	-2.411	
DB1853	PHSS	38.246	-2.628	
DB1895	PHSS	38.174	-2.601	
DB3167	PHSS	37.944	-2.832	
DB3858	PHSS	37.87	-1.572	
DB9360	PHSS	38.065	-2.499	