

# PHD COURSE IN LIFE AND ENVIRONMENTAL SCIENCES

## Report Form for PhD student annual evaluation (XXXVI and XXXVII cycles)

**Name of PhD student:** Lorena Ancona

**Title of PhD research:** Long and short term evolutionary dynamics of transposable elements in the genome of the endangered Apennine yellow-bellied toad (*Bombina pachypus*)

**Name of PhD supervisor:** Emiliano Trucchi, Marco Barucca

**Research lab name:** Laboratorio di Genomica; Laboratorio di Genetica

**Cycle:**

XXXVI

**PhD Curriculum::**

Biomolecular Sciences

**DISVA instrumentation labs/infrastructure eventually involved in the project:**

Actea Mobile Laboratory

Advanced Instrumentation lab

Aquarium

MassSpec lab

MaSBiC

Simulation/informatics lab

Other. Please, indicate: .....

**ABSTRACT (1000 characters, including spaces):**

Transposable elements (TEs) represent a significant fraction of vertebrate genomes, impacting genome composition, architecture and evolution.

The characterization of TEs in the large genome of *Bombina pachypus* will allow us to explore the genome evolution of this endangered species and unveil the history of TE-host evolutionary dynamics.

We investigated the activity of TEs in somatic and germline tissues, to identify potentially active TE families and tissue-specific dynamics. Our results highlighted a higher expression of TEs in the male gonad, with the Gypsy family showing the highest activity.

Furthermore, the activity of 32 target genes involved in TE host silencing mechanisms was analysed, revealing higher activity of TE regulatory pathways in the male gonad. This result suggests that higher TE activity in the germline is associated with higher TE repressive mechanism activation in the host.

## **Part 1. Scientific case of the PhD Research (2 to 3 pages, including figures)**

### **- BACKGROUND**

Transposable elements (TEs) diversity and abundance are highly variable among vertebrate lineages due to different dynamics acting at both the TEs and host genome levels<sup>1</sup>.

TEs could also be the driver of genome expansion in some organisms, playing a major role in genome size variation. In particular, a positive correlation was detected between the accumulation of specific TE families and species with very large genomes. As an example, salamanders, show a much larger amount of long terminal repeat (LTR) retrotransposons due to a slower rate of DNA loss and poor TE silencing mechanisms<sup>2-3</sup>.

This PhD project is focused on the investigation of transposable elements dynamics in the large genome of the Apennine yellow-bellied toad (*Bombina pachypus*), an endangered anuran species endemic to the Italian peninsula, which experienced a post-glacial range expansion that left a marked signature of southern richness - northern purity on its genetic diversity<sup>4</sup>.

Studying TEs in *Bombina* will allow us to explore both long and short term TE-host evolutionary dynamics and, in general, to contribute to our understanding of large genomes evolution.

### **- SCIENTIFIC AIMS**

- Characterize TE abundance and diversity in the genome of *B. pachypus*.
- Analysing the activity of TEs in somatic and germline tissues, in order to identify potentially active TE families and tissue-specific dynamics.
- Exploring the expression of genes involved in TE silencing mechanisms to control TE expression dynamics.
- Investigate the impact of drift-selection balance on TE dynamics, comparing two populations of *B. pachypus* with markedly different effective population sizes.
- Investigate the relationship between genome size and TE content and diversity in a comparative analysis including species belonging to the 3 orders of amphibians.

### **- WORKPLAN AND RESEARCH ACTIVITIES**

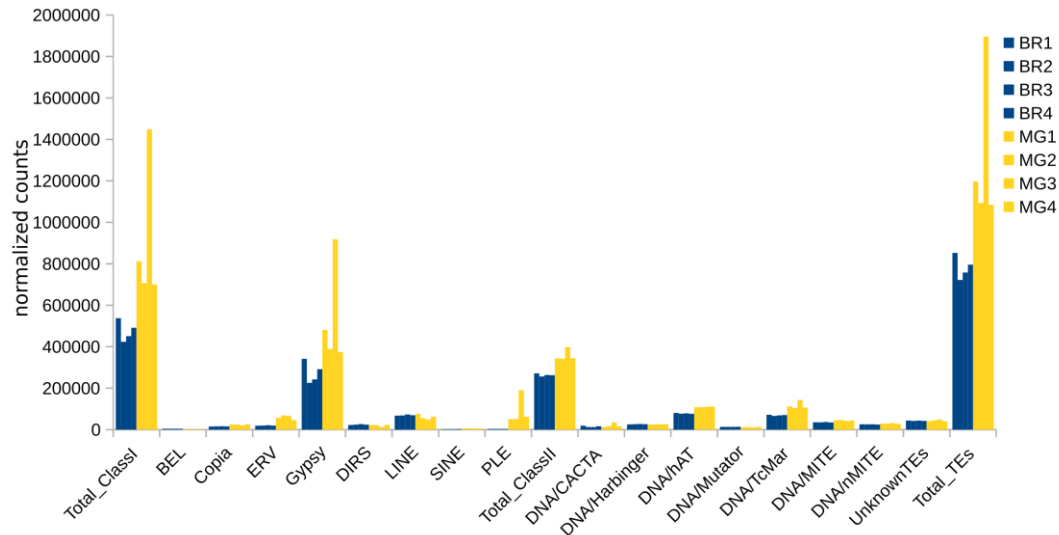
#### **WP 1. Expression analysis of TEs in somatic and germline tissues**

We analysed the activity of TEs in the brain and in the male gonad of four individuals of *B. pachypus* (from a Southern population showing high genetic diversity), to identify potentially active TE families and tissue-specific dynamics. Female gonad tissues are being sequenced to further complement this analysis.

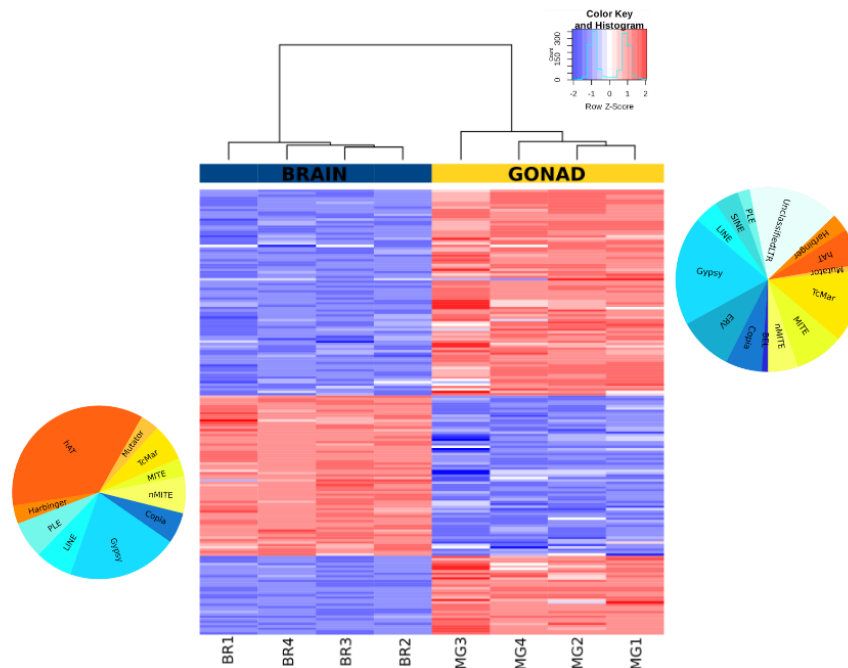
Detection and annotation of TEs in the transcriptome assembly of *B. pachypus* were performed with EDTA<sup>5</sup> and DeepTE<sup>6</sup> to generate a library of TE sequences. Expression levels of the different TE families in the two tissues were estimated with SalmonTE<sup>7</sup> and edgeR<sup>8</sup>, mapping the individual mRNA-seq raw reads on the reference TE library.

On average a higher expression of all TE families was found in the male gonad compared to the brain, a pattern also confirmed by Pasquesi et al.<sup>9</sup> in a study of recent TE expression in 12 vertebrate lineages. Retrotransposons (Total\_ClassI) are the most active class, with Gypsy family showing the highest activity (**Figure 1**).

A total of 186 differential expressed TEs were identified between the two tissues (**Figure 2**): 119 overexpressed in the male gonad, represented mainly by Gypsy and ERV families; and a smaller cluster of 67 overexpressed TEs in the brain, represented mainly by DNA-hAT family and Gypsy family.



**Figure 1. Expression levels of TEs in the transcriptomes of *B.pachypus*:**  
 Total\_ClassI: total of all Retrotransposons (from BEL to PLE families); Total\_ClassII: total of all DNA transposons;  
 BR: brain (blue), MG: male gonad (yellow).



**Figure 2. Differential expression analysis of TEs between the two tissues of *B.pachypus*:**  
 red clusters represent overexpressed TEs ( $\log_{2}FC > 2$ ); pie charts show the abundance of overexpressed TEs.

## WP 2. Expression analysis of TE-regulatory gene pathways

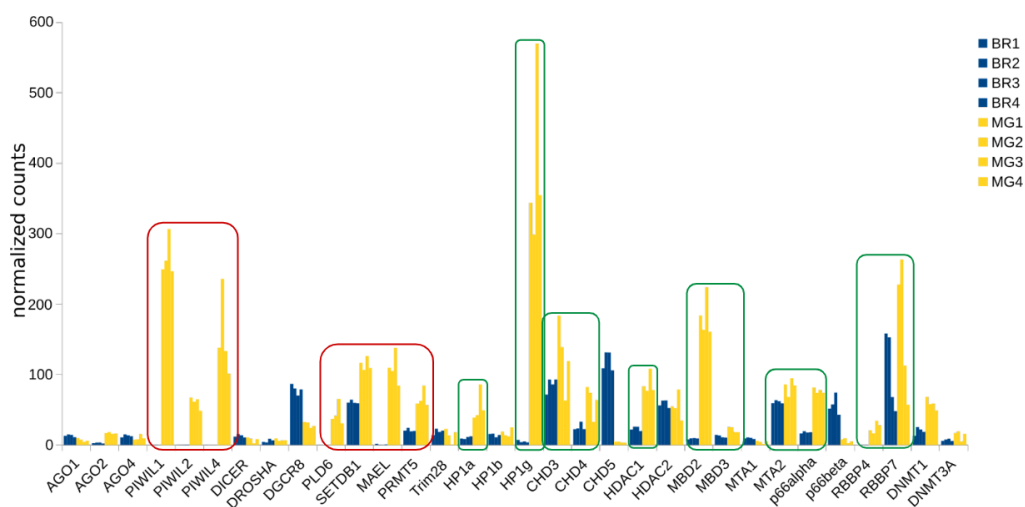
The activity of 32 target genes involved in TE host silencing mechanisms was analysed in the two tissues, to test for tissue-specific strategies to control TE mobilization. Briefly, we searched our annotated brain and male gonad transcriptomes for transcripts which could be orthologous to genes of the Ago and PIWI pathways, genes involved in small RNA biogenesis, and specific transcriptional regulators such as heterochromatin protein, DNA Methyltransferases and the nucleosome remodelling and deacetylase (NuRD) complex.

First, we functionally annotated the transcriptome assembly using different methods (DIAMOND<sup>10</sup>, Blast2GO<sup>11</sup>, UniProt<sup>12</sup>, InterProScan<sup>13</sup> and HMMER<sup>14</sup>), and then we selected, translated and manually

curated the 32 target sequences to identify UTR and CDS regions. Expression levels were estimated with Salmon<sup>15</sup> and edgeR, mapping the individual mRNA-seq raw reads to the reference transcriptome.

We detected higher expression of TE silencing pathways in the male gonad compared to the brain, in particular: PIWI:piRNA pathway and genes involved in primary and secondary piRNA processing (PLD6, SETDB1, MAEL, PRMT5; red boxes in **Figure 3**), genes involved in heterochromatin formation (HP1a-g) and genes encoding proteins of the NuRD complex (CHD3-4, HDAC1, MBD2-3, MTA2, P66a, RBBP4-7; green boxes in **Figure 3**).

These findings suggest that higher TE activity in the germline is also associated with higher TE repressive mechanism activation<sup>16-18</sup>.



**Figure 3. Expression levels of key genes involved in negative regulation of TE activity in the transcriptomes of *B.pachypus*:** BR: brain (blue), MG: male gonad (yellow).

### WP 3. Genomic analysis of TEs

In the coming months, the draft genome assembly of *B. pachypus* will be available and we will soon characterise the abundance and diversity of TEs in this large genome.

Furthermore, we will investigate the impact of selection on TE dynamics by comparing two populations of *B. pachypus* with markedly different effective population sizes: one from the more stable Southern Italy (Calabria) range and the other one from the recently re-colonized northern Apennines (Emilia-Romagna).

Meanwhile, we are performing a comparative analysis across the three orders of amphibians to test whether there is a correlation between genome size and TE content and diversity, and to investigate if the correlation is clade-specific.

At the moment, we have selected 10 anuran species' genomes with different genome sizes (from 1Gb to 10Gb) to which we are applying the same detection and annotation pipeline (FasTE pipeline<sup>19</sup>) in order to obtain a comparable classification.

## - REFERENCES

1. Cibele G, Sotero-Caio, et al. Evolution and Diversity of Transposable Elements in Vertebrate Genomes, *Genome Biology and Evolution*, Volume 9, Issue 1, 2017, Pages 161–177
2. Sun C, Shepard DB, et al. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Gen Biol Evol* 2012;4:168–83
3. Sun C, Lopez Arriaza JR, Mueller RL. Slow DNA loss in the gigantic genomes of salamanders. *Gen Biol Evol* 2012;4:1340–8
4. Canestrelli D, Cimmaruta R, et al. Genetic diversity and phylogeography of the Apennine yellow-bellied toad *Bombina pachypus*, with implications for conservation. *Molecular Ecology* 2006; 15:3741–3754
5. Ou S, Su W, Liao Y, et al. M. B. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, 2019, 20(1), 1–18
6. Yan H, Bombarely A & Li, S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics*, 2020, 36(15), 4269–4275
7. Jeong HH, Yalamanchili HK, et al. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac Symp Biocomput.* 2018;23:168-179. PMID: 29218879
8. Robinson MD, McCarthy DJ, & Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, 26(1), 139-140
9. Pasquesi G.I.M, et al. Vertebrate Lineages Exhibit Diverse Patterns of Transposable Element Regulation and Expression across Tissues. *Genome Biology and Evolution*, 2020, 12(5):506–521
10. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND, *Nature Methods*, 2015, 12, 59-60
11. Conesa A, Götz S, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, Volume 21, Issue 18, Pages 3674–3676
12. The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D480–D489
13. Jones P, Binns D, et al. InterProScan 5: genome-scale protein function classification, *Bioinformatics*, Volume 30, Issue 9, 1 May 2014, Pages 1236–1240
14. Mistry J, Finn RD, et al. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions, *Nucleic Acids Research*, Volume 41, Issue 12, 1 July 2013, Page e121
15. Patro R, Duggal G, et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 2017
16. Carducci F, Carotti E, et al. Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome. *Sci Rep* 11, 14743, 2021
17. Pasquesi GIM, Perry BW, et al. Vertebrate Lineages Exhibit Diverse Patterns of Transposable Element Regulation and Expression across Tissues. *Genome Biology and Evolution*, 2020, 12(5):506–521
18. Reznik B, Cincotta SA, et al. Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells. *Development* 2019, 146(12):dev171157
19. Ellen A. Bell, Christopher L. Butler, et al. Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines. *Mol Ecol Resour.* 2021;00:1–11.

**Part 2. PhD student information on the overall year activity (courses/seminars/schools, mobility periods, participation to conferences)**

***List of attended courses/seminars/schools***

1. Webinar: “Genomic tools for conservation of species: a practitioner 's guide”, 18 November 2021

***List of periods spent abroad***

None

***List of conferences/workshops attended and of contributions eventually presented***

1. Workshop on Population and Speciation Genomics, Cesky Krumlov, Czech Republic, 5-17 June 2022.
2. SIBE congress, UNIVPM, Ancona, 4-7 September 2022.  
Poster presentation: “Ancona L, Carducci F, Biello R, Castrignanò T, Canestrelli D, Barucca M, Trucchi E. Evolutionary dynamics of transposable elements activity and regulation in the endangered Apennine yellow-bellied toad (*Bombina pachypus*)”

**Part 3. PhD student information on publications**

*If not yet published, please indicate the publication status (submitted, accepted, in preparation...)*

***List of publications on international journals***

- J1. Pirri, F., Ometto, L., Fuselli, S., Fernandes, F.A.N., Ancona, L., Perta, N., Di Marino, D., Le Bohec, C., Zane, L., Trucchi, E. Selection-driven adaptation to the extreme Antarctic environment in the Emperor penguin. *Heredity* (2022). <https://doi.org/10.1038/s41437-022-00564-8>
- J2. In preparation: Evolutionary dynamics of transposable elements activity and regulation in the endangered Apennine yellow-bellied toad (*Bombina pachypus*)

***List of publications on conference proceedings***

None

***List of other publications (books, book chapters, patents)***

None

14/10/2022

Student signature



Supervisor signature

