Title of the PhD project:

Study of piRNA clusters: *loci* involved in genome stability

PhD Supervisor

Name, first name: Teysset Laure & Clément Carré
Email : laure.teysset@upmc.fr; clement.carre@gmail.com
Phone : 01 44 27 34 39; 07 60 56 97 03
Title/Employer : Institut de Biologie Paris Seine - LBD
Address : 9 quai Saint Bernard case courrier 24 - 75252 Paris Cedex 05
Laboratory : Laboratoire de Biologie du Développement UMR7622
Title of the team : Transgenerational Epigenetics & small RNA Biology (TErBio)
Team leader (*if different*) *:*
Doctoral School : ED515 - Complexité du Vivant

Overview of the scientific projects of the team

We study the biology of small RNAs involved in gene expression regulation and Transposable Elements (TE). We are also interested in RNAs post-transcriptional modifications that play a role in RNA biogenesis and fate. In metazoans, 3 types of small RNAs have been described, whose specificity differs according to the Argonaute protein with which they interact: microRNA (miRNA), small interfering RNA (siRNA) and PIWI interacting RNAs (piRNA). Transposition of TEs provokes deleterious mutations. In *Drosophila*, TE expressions are controlled by piRNAs in the gonads. Our projects aim to understand the dynamics of piRNA clusters in heterochromatic loci from which piRNAs are maturated and, using functional screens, to characterize genes involved in piRNA-dependent repression. Some of those genes have human orthologue with potential therapeutic targets that will help to better understand pathologies related to the invalidation of those pathways.

Main publications since January 1er, 2016

- Molla-Herman A*, Angelova M, **Carré C**, Antoniewski* and Jean-René Huynh: tRNA Fragments Populations Analysis in Mutants Affecting tRNAs Processing and tRNA Methylation. *Front Genet* (2020) Oct 9;11:518949. doi: 10.3389/fgene.2020.518949.

- Angelova M, Dimitrova D,…Antoniewski C, **Teysset L**,… and **Clément Carré*** (2020) tRNA 2'-O-methylation modulates small RNA silencing and life span in Drosophila *Nucl Acids Res*. 2020 Jan 28;48(4):2050-2072

- Casier K, Delmarre V,...Viodé E, Vaury C, Ronsseray S, Brasset E, **Teysset Laure*** and Boivin Antoine* (2019). Environmentally induced epigenetic conversion of a piRNA cluster. *eLIFE*. 2019 ;8:e39842

- van den Beek M, da Silva B,..., **Carré C***, Antoniewski C*. Dual-layer transposon repression in heads of Drosophila melanogaster. RNA. 2018 Dec;24(12):1749-1760.

- Genenncher, B., Durdevic, Z.,...., Da Silva, B., Legrand, C., **Carre, C**., Lyko, F. & Schaefer, M. Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats (2018) *Cell Rep* 22, 1861-1874.

- Amna Asif-Laidin, Valérie Delmarre,..., Stéphane Ronsseray, **Laure Teysset***. Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in Drosophila. DNA Research, Volume 24, Issue 5, (2017), p. 459–472.

PhD Co-Supervisor

Name, first name : Antoniewski, Christophe
Email : christophe.antoniewski@sorbonne-universite.fr
Phone : 06 68 60 51 50
Title/Employer : Directeur de Recherche CNRS
Address : Plateforme Analyse Bioinformatique ARTbio, IBPS, Sorbonne-Université, Campus Jussieu, 7 quai Saint-Bernard, 75005 Paris
Laboratory : IBPS, FR3631
Title of the team : ARTbio
Team leader *(if different) :*
Doctoral School : ED-515 Complexité du Vivant

Overview of the scientific projects of the team

The primary goal of the ARTbio team is accompanying biologists in their computational analyses. We have built our expertise from studies of small RNA sequencing datasets and the development of the computational framework Galaxy. Today, we are focusing on the development of statistical tools for 1) single cell RNAseq analysis of highly heterogeneous tumors 2) genomic variant detection in leukemias and 3) deconvolution of nanopore signals for direct identification of modified nucleotides in sequencing datasets. We have 3-years experience in an innovative bioinformatics training program called *companionship*. ARTbio is member of the CONECT-AML research consortium, affiliated to the Institut Français de Bioinformatique where it is in charge of the national task-force-covid19 and to the European network ELIXIR, and member of the Galaxy devteam.

Main publications since January 1er, 2016

- Angelova M, Dimitrova D, …, **Antoniewski C**, Teysset L,… and Clément Carré (2020) tRNA 2'-O-methylation modulates small RNA silencing and life span in Drosophila Nucl Acids Res. 2020 Jan 28;48(4):2050-2072. doi:10.1093/nar/gkaa002

- Tekman M, Batut B, Ostrovsky A, **Antoniewski C**, Clements D, Ramirez F, et al. (2020) A single-cell RNA-sequencing training and analysis suite using the Galaxy framework. GigaScience. doi:10.1093/gigascience/giaa102

- Saurty-Seerunghen MS, Bellenger L, El-Habr EA, Delaunay V, Garnier D, Chneiweiss H, **Antoniewski C**, Morvan-Dubois G and Junier M.-P. (2019) Capture at the single cell level of metabolic modules distinguishing aggressive and indolent glioblastoma cells. Acta Neuropathol Commun. 7: 155.

- van den Beek M, da Silva B, Pouch J, Ali Chaouche MEA, Carré C and **Antoniewski C**. (2018) Dual-layer transposon repression in heads of Drosophila melanogaster. RNA. 24: 1749–1760.

- Carissimo G, van den Beek M, Vernick KD, **Antoniewski C.** (2017) Metavisitor, a Suite of Galaxy Tools for Simple and Rapid Detection and Discovery of Viruses in Deep Sequence Data. PLoS One. 12: e0168397.

Doctoral Project

Title: Study of piRNA clusters: loci involved in genome stability

Abstract :

Transposable elements (TE) are a source of deleterious mutations. A first defense that comes into play is based on small RNAs, called piRNAs, which target TEs by sequence complementarity and block their transposition. This defense mechanism is often compared to an adaptive immune system. piRNAs come from discrete loci, the piRNA clusters. Composed of intertwined and inactive TEs, these clusters represent the library of mobile sequences that cells must repress in order to maintain genome integrity. The main goal of this *iBio* PhD project is to understand the dynamics of activation of these piRNA clusters and subsequent production of piRNAs that control TE expression.

RNA chemical modifications (mostly methylation) on the piRNA precursors transcripts will be also investigated in this project. Indeed those RNA modifications are probably involved in the correct funneling of the piRNA precursors transcripts to the piRNA processing factory localized in the cytoplasm. Thus, the PhD project required both biochemistry and chemistry knowledge. In addition, *Drosophila* genetics and bioinformatics sequences analysis and statistics will be an asset for the future recruited *iBio* PhD student.

Context and objective :

Transposable Elements (TE) are mobile DNA sequences present in all genomes. Their insertion into the genome can provoke gene mutation. Thus, specific repressive mechanisms have been set up to protect the genome against uncontrolled TE activity. One of these mechanisms involves particular heterochromatic loci, called piRNA clusters, highly enriched in TE fragments. These loci are expressed in the germline. Their transcripts are exported into the cytoplasm where they are processed into small non-coding RNAs associated with PIWI proteins belonging to the argonaute family of proteins. These RNAs have been called PIWI interacting RNAs or piRNAs. These PIWI/piRNA complexes specifically target active TE transcripts by sequence complementarity to repress transposition *via* RNA interference mechanisms. To date, about thirty protein factors have been identified as being involved in piRNA biogenesis. However, the mechanisms allowing the recognition of these primary transcripts are still unknown.

Our team is interested in the biology of piRNA clusters in the *Drosophila* germline. We are studying more specifically two germline piRNA clusters. One of them is a locus consisting of 7 tandem repeated transgenes, called BX2, which we have been shown to exist in two different epigenetic states: unable to produce piRNAs ($BX2^{OFF}$) or activated for the production of transgenic piRNAs ($BX2^{ON}$). The other one is cluster 1A, which is one of the 140 piRNA clusters identified in the genome of *D. melanogaster*. Cluster 1A is localized at the subtelomeric region of the X chromosome. This locus has the unique feature among all piRNA clusters of being present only in some laboratory strains, allowing controlled genetic experiments (Asif-Laidin et al., 2017). The aim of this PhD project is to better understand the genesis and activation of piRNA clusters, using BX2 and cluster 1A as piRNA cluster models.

In a previous study, we showed that maternal inheritance of transgenic piRNAs is crucial for the activation and maintenance of a locus into a piRNA-producing $BX2^{ON}$ locus (de Vanssay et al., 2012).

The first aim of this *iBio* PhD project will be to set up the tools to understand the determinism of this maternal inheritance. For this part, the student will have to prepare immunoprecipitations from ovaries of $BX2^{ON}$ lines using antibodies specific to the three PIWI proteins (Aub, Piwi and Ago3). These PIWI-piRNA complexes will then be microinjected separately or in combination into embryos of the $BX2^{OFF}$ lineage. Using a functional test developed in the laboratory (de Vanssay et al., 2012), the epigenetic status of this locus will then be tested. This study will provide a better understanding of the mechanistics and determinism of maternal inheritance.

The second aim of this *iBio* PhD project is to understand the dynamics of activation and production of cluster 1A. In a study currently underway in the laboratory, we noticed that cluster 1A does not seem to be transcribed homogeneously. This result was surprising because it was assumed that piRNA clusters were transcribed as a single transcription unit. In order to understand this observation, the PhD student will sequence and analyze a transcriptome using a direct RNA sequencing approach from Oxford Nanopore Technology (ONT). This ONT approach has the advantage of providing long sequence fragments (long reads) adapted to the sequencing of repeated regions. In addition, using this direct RNA sequencing approach, RNA sequences are read directly without cDNA preparation. This analysis will provide structural information of transcripts from cluster1A, such as their size, initiation sites, splicing region, termination site, etc... It will also allow us to specify how piRNA clusters are transcribed and how their transcripts are identified as precursor sequences to be funneled to the piRNA production factory in the cytoplasm.

The third aim of this *iBio* PhD project will be to test whether RNA post-transcriptional modification such as ribose 2'O-méthylation (Nm), cytosine methylation (m^5C) and adenine methylation (m^6A) are required on primary transcripts of BX2 and 1A piRNA clusters to funnel those transcripts to the piRNA factory in the cytoplasm. Importantly, in the two piRNA clusters cited above, we already have identified that the activation state is not associated with a particular accumulation of transcripts. Indeed in both states (piRNA production ON or OFF), those piRNAs clusters are transcribed at the same level. We hypothesized that the recruitment of factors allowing the addressing of piRNA precursors to the piRNA biogenesis pathway could be achieved through post-transcriptional modifications of these transcripts, allowing the recruitment of particular factors (readers of those RNA modifications). Recently, the team has identified two *Drosophila* 2'-O methyltransferases (Nm MTase) and two m^5C methyltransferases necessary for the repression of certain germline TEs (Angelova & Dimitrova et al., 2020 and Genenncher et al., 2018).

We have already built and characterized genetic tools (piRNA sensors for BX2 and 1A clusters: de Vanssay et al., 2012; Asif-Laidin et al., *in prep*) and mutant lines (Nm, m^5C, m^6A, etc...MTases: Angelova & Dimitrova et al., 2020 and Genenncher et al., 2018) necessary to tackle this question. In addition, we will take advantage of the ONTs direct RNA long reads generated in aim 2 to look for modifications on the precursor transcripts of piRNA clusters using dedicated bioinformatics algorithms (Ueda et al., 2020).

References

Angelova MT, Dimitrova DG, Da Silva B, Marchand V, Jacquier C, Achour C, Mira B, Goyenvalle C, Bourguignon-Igel V, Shehzada S, Khouider S, Lence T, Guerineau V, Roignant JY, **Antoniewski C**, **Teysset L**, Bregeon D, Motorin Y, Schaefer MR and **Carré C***. 2020. tRNA 2'-O-methylation by a duo of TRM7/FTSJ1 proteins modulates small RNA silencing in Drosophila. Nucleic Acids Research, 48:2050-2072.

Asif-Laidin A, Delmarre V, Laurentie J, Miller WJ, Ronsseray S and **Teysset L***. 2017. Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in Drosophila. DNA Res. 27.

de Vanssay A, Bougé AL, Boivin A, Hermant C, **Teysset L**, Delmarre V, **Antoniewski C***, Ronsseray S*. 2012. Paramutation in Drosophila linked to birth of a piRNA producing locus. Nature. 490:112-5.

Genenncher, B., Durdevic, Z., Hanna, K., Zinkl, D., Mobin, M. B., Senturk, N., Da Silva, B., Legrand, C., **Carré, C.**, Lyko, F. & Schaefer, M. Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. Cell Rep 22, 1861-1874.

Hiroki Ueda. *nanoDoc*: RNA modification detection using Nanopore raw reads with Deep One Class Classification https://doi.org/10.1101/2020.09.13.295089

Justification of suitability for i-Bio: The PhD project raises two main questions that represent a breakthrough in the piRNA biology field as well as technological advances in the epitranscriptomics that combine both, chemical approaches as well as bioinformatics analysis and development of bioinformatics pipelines necessary for the ONT approaches that we started to develop recently. This project will be possible in a reasonable amount of time (3 years) only if there is a strong collaboration between two types of expertise: 1- *Drosophila* genetics, epigenetics and epitranscriptomics that *Transgenerational Epigenetics and small RNA Biology* (*TErBio*) team leaded by Clément Carré and Laure Teysset at Sorbonne University - IBPS - possesses, but also 2- chemistry based RNA modification detection and advance statistical and bioinformatics analysis that will be fulfilled by Christophe Antoniewski at the *ARTbio* bioinformatic analysis platform located at IBPS. As detailed above, this project will benefit from the bioinformatics expertise of Christophe Antoniewski on RNA modification detection (recently developed at *ARTbio*), expertise that is not to our knowledge yet developed elsewhere at Sorbonne University Alliance. The progress of the PhD project would benefit from the co-direction between our 2 teams and the fact that the student will perform (sequencing), develop (computer technologies including machine learning software) and analyze (statistics associated to RNA sequences analysis) the experiments in the two teams, in three area of expertise: Biology, Chemistry and Bioinformatics development and analysis. Finally, Clément Carré is a COST *Epitran* French representative, which currently counts 26 member states and 60 management members. The COST *Epitran* will help to attract good and specialized PhD candidates on our *iBio* PhD project.

Role of each supervisor / skills provided: *Drosophila* Genetics and Biochemistry experiments will be led by Laure Teysset and Clement Carré in the *Transgenerational Epigenetics and small RNA Biology* team. Our team developed dedicated tools to identify and characterize the biology of the two piRNA clusters (BX2 and cluster 1A) as well as detection of RNA modifications (Nm and m^5C) on piRNA precursors transcripts of these clusters. These tools described in the above references tagged in red will be used in this PhD project. The production of direct RNA sequencing using ONT approaches will be performed directly in our lab with a dedicated Minion *Mk1C* sequencer (already available in the team). The bioinformatic analysis part (ONT sequencing analysis) of the PhD project will be led by Dr. Antoniewski with whom we started an ONT sequencing analysis project in 2020. Christophe Antoniewski is in charge of the *ARTbio* bioinformatics analysis platform (IBPS) which has strong expertise in the field of small non-coding RNA and corresponding RNA sequencing analysis including nanopore technologies. Based on our strong preliminary results and past collaborations, our consortium is among the pioneers in this field.

Profil of the desired student: The successful candidate will specifically investigate the formation and activation of piRNA clusters in already available piRNAs clusters sensors. His/her project is based on recent discoveries (see red references above) from the lab indicating that maternally deposited piRNAs (in dedicated argonautes proteins) and RNA modifications are involved in the properties and functions of the piRNA pathway. The successful candidate should have a strong background in genetics, epigenetics, molecular biology and biochemistry, as well as skills in computer technologies, statistics and good knowledge of nucleic acid chemistry. Our labs have a long time experience on PhD mentorship for several years (average one PhD student/ year dispatch in 4 HDR lab members). Our labs also work regularly with students from all around the world (Bulgaria, Pakistan, France, Algeria, Germany), thus the lab common language is both English and French.