



MASTER THESIS OFFER OF CEPLAS

(FOR COMPUTATIONAL MASTER STUDENTS FROM TECHNICAL UNIVERSITY OF MADRID, UPM)

Title of Master Thesis

Discovery of *cis*-regulatory motifs with epigenetic signature in plants

Description of student's tasks

The student will perform meta-gene analyses testing for enrichment of available *cis*-regulatory position weight matrices across target regions that are controlled by epigenetic repressors of the Polycomb Group pathway. Recent data have indicated that several *cis*-motifs that are crucial for epigenetic repression have a particular epigenetic signature, contributed by motif density, motif combination and a particular distribution across target regions. The project aims to a) improve the approach by implementing a pipeline that allows selecting gene sets of comparable size and structure for statistical comparison and b) discover novel motifs and motif combinations that are predictors of epigenetic gene repression in plants. Methods will be first applied to the model plant *Arabidopsis thaliana* but should be extended to other well annotated land plant species if time permits.

Prerequisites (languages, informatics skills, bioinformatic skills, other knowledge, etc)

The student should be familiar with using and adapting established bioinformatics pipelines that are used in genomics (Bowtie2/MEME suite/Homer suite/Deeptools/various R pipelines). All pipelines will be run as command line tools on the MIPZ compute cluster. Bash scripting and writing plugins e.g. using R/Bash scripting/Python should be done autonomously, solid statistical background is a plus. The group members are mostly experimental biologists with a strong background in epigenetics, but there will be a second Master student from the University of Cologne with a data analysis project. The student can expect ample training in epigenetics, but should be rather self-sufficient in computational approaches.

Training Project

Epigenetic gene regulation allows genes to acquire a memory of their expression state, a requirement for cell-type specification during organism development, but also for an appropriate gene response to environmental cues. For example, epigenetic memories allow many plants to

remember a preceding winter season for a sufficient time to allow rapid flowering in spring. Recent progress, including from our group, established that epigenetic gene control is underpinned by genetic information in the form of cis-regulatory motifs that, through sequence-specific DNA-binding proteins, recruit epigenetic repressors¹⁻⁴. These repressors, which are part of the evolutionary conserved Polycomb Group pathway, create a locally compacted chromatin environment that prevents target gene transcription. Cis-motifs associated with epigenetic repression seem to have particular signatures that distinguish them from cis-motifs associated with conventional transcription factors. For example, these cis-motifs are not enriched at proximal promoters but rather widely distributed across structural genes, they strongly depend on motif combinations and occur in loose repetition across their target regions. This project is aimed to develop a solid pipeline to test cis-motifs for epigenetic signatures and to discover novel motifs associated with epigenetic gene repression in plants.

Task 1 will be the development of a pipeline to select feature based gene-sets for binary statistical comparisons. Features could include gene size, intron/exon structure, epigenetic profiles or any custom feature. The tool should allow permutation, subsampling and ranking for subsequent statistical tests.

Task 2 will be to annotate available position weight matrices (e.g. available DAP-seq data) within gene sets and perform enrichment tests based on absence or presence of epigenetic features such as the H3K27me3 modification or mis regulation of target genes in epigenetic pathway mutants using data from *Arabidopsis thaliana*.

Task 3 will be to perform a co-occurrence analysis of cis-motifs based on available workflows, such as COPS⁵.

Task 4 will be to transfer the analysis to other plant species for which appropriate epigenetic annotations are available, e.g. *A. lyrata*, *Arabis alpina* and the oilseed crop *Brassica napus* within the Brassicaceae family.

Task 5 will be the redaction of a comprehensive project report and that includes a well-documented pipeline.

References

1. Zhou, Y. *et al.* Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in *Arabidopsis*. *Nature genetics* **50**, 638–644; 10.1038/s41588-018-0109-9 (2018).
2. Yuan, W. *et al.* A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in *Arabidopsis*. *Nature genetics* **48**, 1527–1534; 10.1038/ng.3712 (2016).
3. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nature genetics* **49**, 1546–1552; 10.1038/ng.3937 (2017).
4. Qüesta, J. I., Song, J., Geraldo, N., An, H. & Dean, C. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science (New York, N.Y.)* **353**, 485–488; 10.1126/science.aaf7354 (2016).
5. Nati Ha, Maria Polychronidou & Ingrid Lohmann. COPS: Detecting Co-Occurrence and Spatial Arrangement of Transcription Factor Binding Motifs in Genome-Wide Datasets. *PLOS ONE* **7**, e52055; 10.1371/journal.pone.0052055 (2012).

Activities that will be performed in the academic internship/Master Thesis:

The student can dedicate most of her/his time to working on the individual scientific project. In addition, the student will attend regular scientific meetings on a weekly basis: group meeting and journal club, Departmental meeting, Institute seminar (an invited external speaker on various topics in plant sciences). The student will present own progress on a regular basis to the group (around 4-times during the period). Other opportunities for scientific exchange are offered through CEPLAS. Furthermore, the student is encouraged participate in a weekly scientific writing course in English language that is organized at the department level.

Discussions with the supervisor may occur at a daily, informal level. If this easy interaction is impossible due to the pandemic situation, a formal 1h weekly meeting between student and supervisor, will be scheduled. The internship concludes with a ca. 4-week period of writing a final report that includes full documentation of the analysis pipeline on GitHub.

Nº of positions offered:	1
Has the student dealings with underage persons?	NO
Starting date:	February/March 2021 if pandemic situation permits
End date:	October 2021
Weekly hours (only for internship in CEPLAS lab):	40
Working hours (only for internship in CEPLAS lab):	9:00-18:00
Fellowships (if any, NOT REQUIRED): Remuneration (€/month):	800€/month
Academic tutor (UPM/CBGP): <i>(you need a Tutor from UPM Master, not involved in the research activity)</i> Email:	
Department/Research Group of UPM/CBGP Academic tutor:	



CEPLAS Internship/Master Thesis Tutor/Director:	Dr. Franziska Turck
Email CEPLAS tutor:	turck@mpipz.mpg.de
Department CEPLAS tutor:	Plant Developmental Biology
Location of the internship (telecommuting?):	Should be Cologne, but the project could be started from home if the pandemic does not allow otherwise
CEPLAS Institution:	Max Planck Institute for Plant Breeding Research
<i>To be completed by Internship Office ETSIAAB-UPM:</i>	
Number of ECTS (Nº ECTS):	

Send by email to: international.cbgp@upm.es (Pablo Gómez)