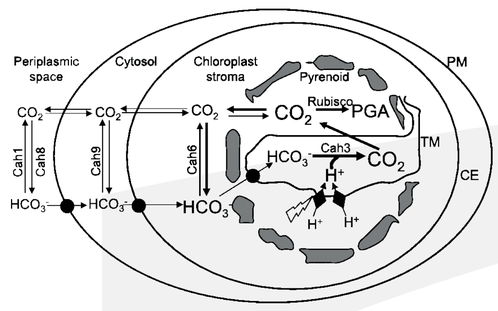
FAIRE-seq data analysis of *Chlamydomonas reinhardtii* under carbon deprivation

David Urbina-Gómez, Flavia Vischi Winck, Bernd Müller-Röber, Diego Riaño-Pachón

Group of Computational and Evolutionary Biology, Biological Sciences Department, Universidad de los Andes

e-mail: da.urbina34@uniandes.edu.co, Vischi@mpimp-golm.mpg.de, bmr@uni-potsdam.de, dm.riano122@uniandes.edu.co

Carbon concentration mechanism

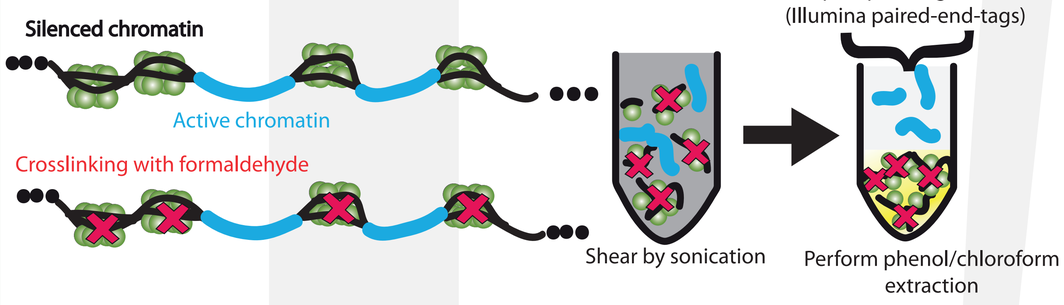


Modified from Moroney and Ynalvez, 2007,Eukaryotic Cell 6: 1251–1259.

The unicellular green algae *Chlamydomonas reinhardtii* can acclimate to limiting and variable concentrations of extracellular inorganic carbon (CO2 or HCO3-) through the activation of the carbon concentration mechanism (CMM). CCM activation is dependent of low carbon concentration and light intensity. This mechanism optimized extracellular inorganic carbon uptake and the increase of its concentration in the chloroplast stroma where the enzime ribulose-1, 5- bisphophate carboxylase oxygenase (Rubisco) is located, so the carbon dioxide fixation is enhanced, producing also the increase in the rate of the photosynthesis reaction **[1]**.

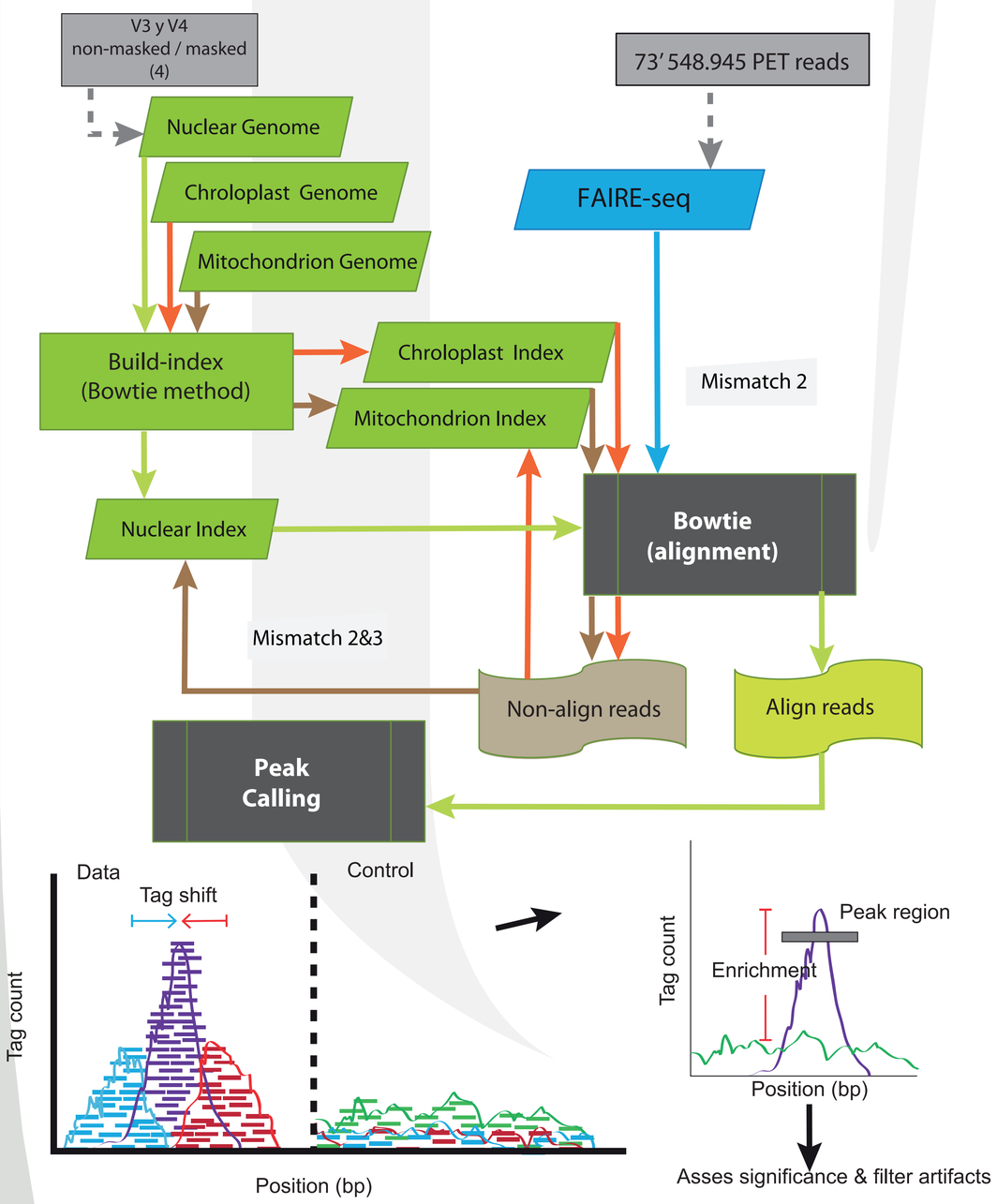
Genetic and genomic studies have allowed the deciphering of diferent CCM-related genes involve in regulation, membrane transport and carbonic anhydrase activity. However, more detailed information about regulation is still needed **[1]**. In this study, we were interested in identifying putative genomic regions that were involved in the active regulation of the transcription processes *in vivo*, under carbon deprivation.

FAIRE assay



Data from a Formaldehyde assisted isolation of regulatory elements, following by high-throughput sequencing (FAIRE-seq). In this assay chromatin is cross-linked using formaldehyde, sonicated and subjected to phenol-chloroform extraction. DNA fragments recovered in the aqueous phase are then sequenced

Alignment strategy and peak identification



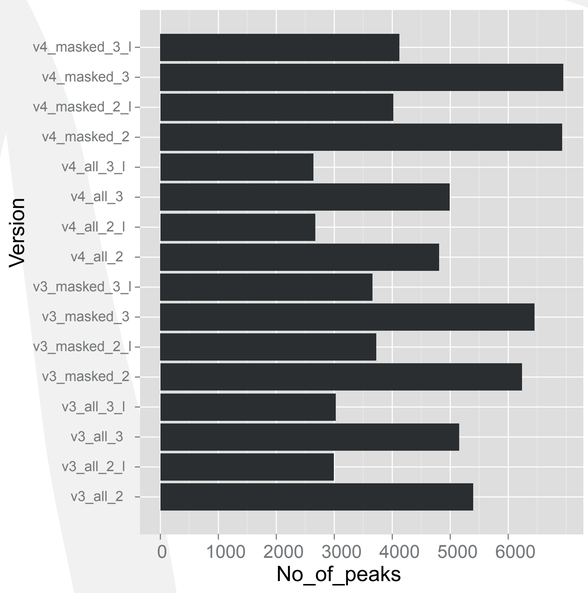
Paired-end Illumina reads of 50 bp obtained in a past study from the FAIRE-seq assay, were mapped using Bowtie **[2]**. First to the chloroplast and mitochondrion sequenced genomes discarding the mapped reads. Unmapped reads from the previous process were mapped to the different versions of the sequenced nuclear genome of *C.reinhardtii*. The resulting mapped reads were used for the identification of enriched regions by the use of two open-source peak calling packages (MACS **[3]** and Fseq). *Results only shown for MACS*

Acknowledegements

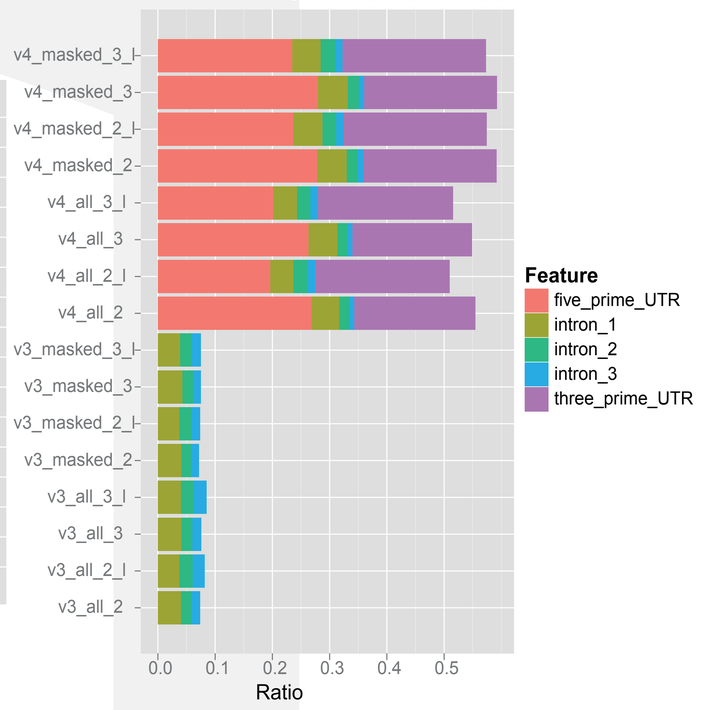
Flavia Vischi Winck and Bernd Müller-Röber for providing the FAIRE data.
For the ISCB Latin America 2012 for the travel fellowship award for conference in Santiago, Chile.
Group of Computational and Evolutionary Biology of the Universidad de los Andes -BCE.
Sciences Faculty of Universidad de los Andes.

We perform a function analysis approach in which we analyzed location of the peaks relative to the annotated genes for each of the different versions obtained; identifying which peaks overlap with specific gene features related with the transcription initiation processes such as UTR regions and first gene introns. The nearest gen start codon distance was also evaluated for each peak. For both analysis, overlapping and nearest start codon, the BEDTools suite was used **[4]**. The related genes were associated from the annotation files (source:JGI and Phytozome) to the Gene-Ontology identifiers and an evaluation of the over-representation of Gene-Ontology categories was carried out by the use of the Cytoscape plugin: BINGO **[5]**.

Number of peaks identified

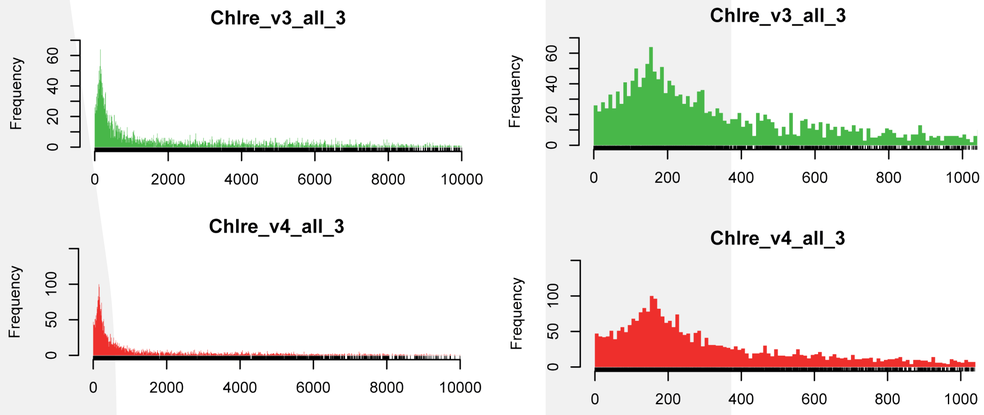


Overlaping of peaks with trascription related fatures



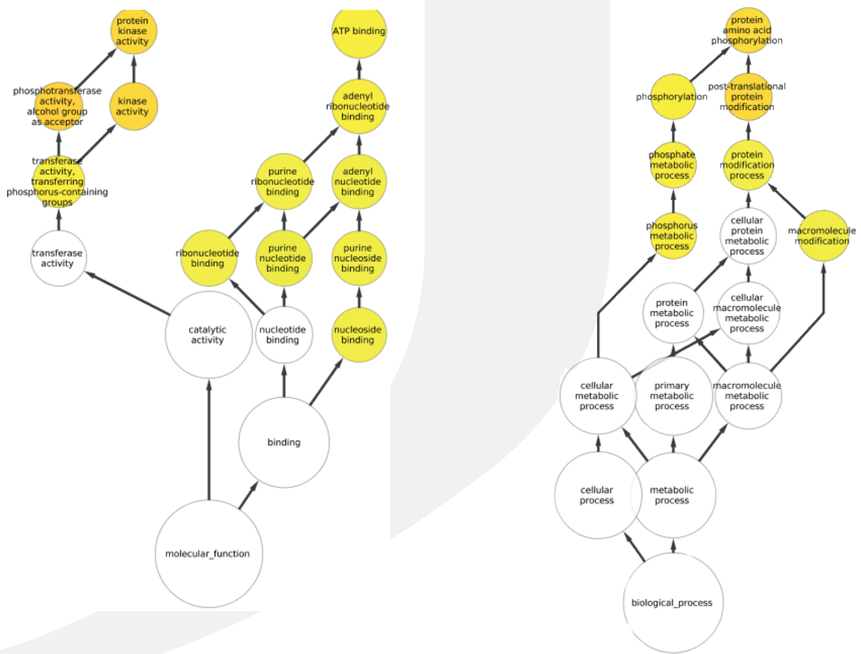
Approximately, 50% of the identified peaks for each of the treatments (different reference genome version, alignment parameters and peak calling package used) overlap with some of the evaluated gen features (UTRs, and first three introns). Regarding the proportion of peaks overlapping with the first three introns features, the amount of peaks overlapping with the first intron is always higher. This is consistent with the model that suggest, that in plants intron sequences near to 5'-end are likely to mediate a change in the transcription machinery which renders it more processive **[6]**.

Nearest start codon distance



We found that the distance of the peaks to the the nearest start codon is frequently less than 500 bp, displaying a summit of the frequency at 180bp. Nevertheless, peaks to a distance of even 10 Mb are found, indicating the presence of open chromatin regions that are likely to contain enhancers.

Over-representation of Gene-Ontology categories



From the analysis with the genes associated to the peaks, over-representation of functions such as, protein kinase activity, atp-binding and scavenger receptor activity were found. Also, genes related with cellular components such as the oxygen evolving complex in the thylakoid membrane were over-represented.

The results allow us to conclude that the genes related to the nucleosome depleted regions, link with functions, specific biological processes and are associated with locations that have been previous described in *C.reinhardtii* under the same conditions **[1,7,8]**, future analysis are needed for a more specific identification of motifs of the putative regulatory elements present in the identified depleted regions.

Bibliography

1. Yamano T, Fukuzawa H: **Carbon-concentrating mechanism in a green alga, Chlamydomonas reinhardtii, revealed by transcriptome analyses**. *Journal of Basic Microbiology* 2009, **49**:42-51.
2. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biol* 2009, **10(3)**:R25.
3. Feng J, Liu T, Zhang Y: **Using MACS to identify peaks from ChIP-Seq data**. *Curr Protoc Bioinformatics* 2011, Chapter 2:Unit 2.14.
4. Quinlan AR, Hall IM: **BEDTools: a fexible suite of utilities for comparing genomic features**. *Bioinformatics 2010*, **26(6)**:841-842.
5. Maere S, Heymans K, Kuiper M: **BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks**. *Bioinformatics* 2005, **21(16)**:3448-3449.
6. Parra G, Bradnam K, Rose AB, Korf I: **Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants**. *Nucleic Acids Res* 2011, **39(13)**:5328-5337.
7. Fukuzawa H, Miura K, Ishizaki K, Kucho Ki, Saito T, Kohinata T, Ohyama K: **Ccm1, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO2 availability**. *Proceedings of the National Academy of Sciences* 2001, 98(9):5347-5352.
8. Xiang Y, Zhang J, Weeks DP: **The Cia5 gene controls formation of the carbon concentrating mechanism in Chlamydomonas reinhardtii**. *Proc Natl Acad Sci* U S A 2001, 98(9):5341-5346.

Nature Precedings : doi:10.1038/npre.2012.7035.1 : Posted 27 Mar 2012