THE EMBO JOURNAL

Manuscript EMBO-2017-96392

The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*

Michiel Matthijs, Michele Fabris, Toshihiro Obata, Imogen Foubert, Jose Franco-Zorrilla, Roberto Solano, Alisdair Fernie, Wim Vyverman and Alain Goossens

Corresponding author: Alain Goossens, VIB-Ghent University

Review timeline:

Submission date:	05 January 2017
Editorial Decision:	23 January 2017
Revision received:	03 March 2017
Editorial Decision:	10 March 2017
Revision received:	14 March 2017
Accepted:	15 March 2017

Editor: Andrea Leibfried

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 23 January 2017

Thank you for submitting your manuscript for consideration by the EMBO Journal. As your manuscript had been under consideration elsewhere before, it has now been seen by an arbitrating referee, who had access to the initial concerns raised, as well as to your point-by-point response to them. I enclose the comments of this referee on the current version of your manuscript below.

As you will see, the arbitrating referee endorses publication of a further revised version of your manuscript in The EMBO Journal. The referee suggests extending the circadian experiments via a meta-analysis (point 1), and we strongly encourage this. The referee furthermore notes that more information on bZIP14 is needed (point 2) and that the quality of the discussion and of the figures needs to be improved (points 3 and 4).

REFEREE REPORTS

Referee #1:

I have been through the revised manuscript and the rebuttal letter. Overall my evaluation is that the replies to the referee's comments are adequate and that the advance reported is sufficient to justify it being published in EMBO J, although I recommend some further improvements (see below). In my opinion it is good for the journal to encourage studies on less orthodox experimental systems, and although the quality of the work is not equivalent to what is seen for more conventional organisms I

believe that it is of sufficiently high quality for marine micro-eukaryotes, for which far fewer resources are available. I furthermore appreciated the wide range of techniques used, and the new yeast one-hybrid analyses provide important support for the main conclusions of the paper, even though the authors were unable to demonstrate transcriptional activation. I also believe that the new phylogenetic analyses add further value to the paper and make the work of interest beyond the diatom community.

Specific recommendations for improvement:
1. I do not see much value in the new circadian experiments. The data is quite limited and is insufficient to conclude that there is circadian control because of the lack of comprehensive experiments in free-running, extended light and extended dark conditions. I therefore recommend to remove this section, unless more data can be mined from a very recent and extensive analysis of diel cycling of gene expression in Phaeodactylum published in PLoS Genetics (see http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006490). In my opinion it will be preferable to strengthen this section of the manuscript by focusing on diel expression patterns and pulling in relevant and informative data from this new expression dataset, but if there is nothing new to report then the whole section should be removed.

2. More information should be provided about the presence of the two-domain bZIP14 gene beyond the stramenopiles, eg, in SAR group organisms, and in marine prasinophytes such as Ostreococcus. Fig 7c is not informative and should be removed because the experiments on these three different organisms were surely done in very different conditions.

3. The Discussion is poor and lacks depth. The significance of the findings in ecological and biotechnological contexts should be evoked, as should the experiments that didn't work, such as the search for a knockout mutant, the inability to demonstrate transcriptional activation, and why they couldn't generate an antibody.

4. Some of the supplementary figures are very poor (eg, S2, S3) or are illegible (S8, X axes). There is no excuse for providing such poor figures, especially when hoping to publish the work in a top journal.

1st Revision - authors' response 03 March 2017

Response to Referee #1:

I have been through the revised manuscript and the rebuttal letter. Overall my evaluation is that the replies to the referee's comments are adequate and that the advance reported is sufficient to justify it being published in EMBO J, although I recommend some further improvements (see below). In my opinion it is good for the journal to encourage studies on less orthodox experimental systems, and although the quality of the work is not equivalent to what is seen for more conventional organisms I believe that it is of sufficiently high quality for marine micro-eukaryotes, for which far fewer resources are available. I furthermore appreciated the wide range of techniques used, and the new yeast one-hybrid analyses provide important support for the main conclusions of the paper, even though the authors were unable to demonstrate transcriptional activation. I also believe that the new phylogenetic analyses add further value to the paper and make the work of interest beyond the diatom community.

Specific recommendations for improvement:
1. I do not see much value in the new circadian experiments. The data is quite limited and is insufficient to conclude that there is circadian control because of the lack of comprehensive experiments in free-running, extended light and extended dark conditions. I therefore recommend to

remove this section, unless more data can be mined from a very recent and extensive analysis of diel cycling of gene expression in Phaeodactylum published in PLoS Genetics (see http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006490). In my opinion it will be preferable to strengthen this section of the manuscript by focusing on diel expression patterns and pulling in relevant and informative data from this new expression dataset, but if there is nothing new to report then the whole section should be removed.

We have extended the circadian experiments via a 'meta-analysis' of all relevant published transcriptome data, in three different diatom species including the experiment indicated by the referee. This meta-analysis corroborates the findings from our dataset and, additionally, the same circadian gene expression patterns were found to be present in all three species examined. Thereby this meta-analysis strengthens the evolutionary conserved role of the *bZIP14* transcription factor. As such, we believe that it is advantageous to highlight the similarities between nitrogen starved cells and cells during dusk. While circadian rhythms are known to affect nitrogen metabolism this is the first instance linking the two on a regulatory level in a diatom.

2. More information should be provided about the presence of the two-domain bZIP14 gene beyond the stramenopiles, eg, in SAR group organisms, and in marine prasinophytes such as Ostreococcus. Fig 7c is not informative and should be removed because the experiments on these three different organisms were surely done in very different conditions.

The requested additional phylogenetic/evolutionary analysis has been carried out. More emphasis has been placed on the 'non-diatoms' within the SAR group such as the dinoflagellates, prasinophytes and golden algae. The original phylogenetic search had already included members of the green algae but no orthologs could be identified in these organisms, which has been clarified in the revised manuscript text.

Because the transcriptome analysis of the two other diatom species shown in Fig. 7c was carried out by the same research group, we thought the presentation of the data may nonetheless be relevant, therefore we have moved this panel to the Appendix data set (new Appendix Fig. S12)

3. The Discussion is poor and lacks depth. The significance of the findings in ecological and biotechnological contexts should be evoked, as should the experiments that didn't work, such as the search for a knockout mutant, the inability to demonstrate transcriptional activation, and why they couldn't generate an antibody.

The discussion has been edited and expanded, as suggested by the referee.

4. Some of the supplementary figures are very poor (eg, S2, S3) or are illegible (S8, X axes). There is no excuse for providing such poor figures, especially when hoping to publish the work in a top journal.

We apologise for this. All figures have been verified and remade where necessary according to the guidelines and standards requested by The EMBO Journal. Additionally we consulted the editor for approval of the implemented changes.

2nd Editorial Decision 10 March 2017

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by the arbitrating referee again whose comments are enclosed. As you will see, the referee is broadly in favor of publication, pending satisfactory minor revision.

I would thus like to ask you to address the remaining concerns and to provide a final version of the manuscript.

The following editorial points should be addressed as well:
- number of replicates is not indicated everywhere, please add this information
- part of the method section heavily resembles the one from your previous paper published in plant physiology, 2016. Please add the reference or change the wording slightly.

REFEREE REPORT

Referee #1:

I am satisfied with the author's revisions and am happy to see that the manuscript has been further improved. In my opinion, there are just a few issues to be resolved before the paper can be published:

1. The Fragilariopsis genome has now been published and can thus be cited (Mock et al. Nature (2017))
2. Fig 1B is cited before Fig. 1A so the panels should be inverted
3. Fig 3C is cited before Fig. 3B so the panels should be inverted
4. Table 1 seems to be overly long to be included in the main text and in my opinion would be better placed in the Supplementary Information
5. There seems to be some confusion in defining what are circadian and diel expression patterns, so I advise the authors to check the manuscript thoroughly for accuracy in defining the different expression patterns

2nd Revision - authors' response 14 March 2017

Response to the editor
The following editorial points should be addressed as well:
- number of replicates is not indicated everywhere, please add this information

The missing information has been appended both in the Figure legends and the Materials & Methods section.

- part of the method section heavily resembles the one from your previous paper published in plant physiology, 2016. Please add the reference or change the wording slightly.

We apologise for this. Repetitive passages in the sections 'Expression profiling' and 'Molecular cloning' have been omitted and replaced by a reference to our paper.

Response to Referee #1:
I am satisfied with the author's revisions and am happy to see that the manuscript has been further improved. In my opinion, there are just a few issues to be resolved before the paper can be published:
1. The Fragilariopsis genome has now been published and can thus be cited (Mock et al. Nature (2017))

The corresponding reference has been included in the introduction section.

2. Fig 1B is cited before Fig. 1A so the panels should be inverted

The figure has been edited accordingly.

3. Fig 3C is cited before Fig. 3B so the panels should be inverted

We prefer the current order given that 3A and 3B are clusters derived from the same RNA-Seq experiment and 3C represents another data set derived from qPCR on a different sample set.

4. Table 1 seems to be overly long to be included in the main text and in my opinion would be better placed in the Supplementary Information

The indicated table has been moved to the Appendix (new Table S1).

5. There seems to be some confusion in defining what are circadian and diel expression patterns, so I advise the authors to check the manuscript thoroughly for accuracy in defining the different expression patterns

We apologise for this confusion. The corresponding sections have been revised and edited accordingly. We have consistently used the term 'diurnal', given the currently accepted definitions of diurnal/circadian, indicated hereafter: "A diurnal rhythm is any output that is synchronized to Earth's 24-h day. It may be endogenously generated or it may simply be a response to environmental cues. A circadian rhythm is an endogenously generated rhythm with a period close to 24 h. There are three specific criteria that must be satisfied for something to be called a circadian rhythm. The rhythm must continue under constant conditions (i.e., with no environmental time cues) with a period close to 24 h. The rhythm must be able to be phase reset by environmental cues, so that it can be synchronized to the 24-h day. The rhythm must be temperature compensated, meaning the period depends on weakly on temperature within a normal biological temperature range. A diurnal rhythm may or may not be a circadian rhythm."

3rd Editorial Decision 15 March 2017

Thank you for sending the revised version of your manuscript. I appreciate the introduced changes, and I am happy to accept your manuscript for publication in the EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Alain Goossens
Journal Submitted to: EMBO Journal
Manuscript Number: EMBOJ-2016-96392

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Sample sizes were estimated based on previous small scale experiments to ensure that treatment effects would not be masked by biological variation
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	N/A
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	N/A
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Three independent replicates were run for all treatments of the RNA-seq datasets, all other samples were done in triplicate with the exception of the lipid analysis which was done with a single replicate as it was merely done to confirm the expected pattern, Metabolite analysis was performed with six replicates
For animal studies, include a statement about randomization even if no randomization was used.	N/A
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	For the cluster analysis the initial stages were performed with randomised identifiers to select interesting gene candidates. All other data used was quantitative and not subject to interpretation
4.b. For animal studies, include a statement about blinding even if no blinding was done	N/A
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	For the RNA-seq analysis the default parameters were used present in the Tophat package
Is there an estimate of variation within each group of data?	Yes, with the exception of the lipid quantification
Is the variance similar between the groups that are being statistically compared?	Yes, F-tests were performed prior to the T-test

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	No antibodies used
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Phaeodactylum tricornutum, CCAP 1055/1 from SAMS (UK)

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	No animals used
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	N/A
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	N/A

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	N/A
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	N/A
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	N/A
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	N/A
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	N/A
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	N/A
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	N/A

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	The generated RNA-seq data is available from the SRA website with accession number: ERP013403
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	N/A
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	N/A
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	Yes
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	N/A

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	No