Reviewer’s report

Title: Origin of the nuclear proteome on the basis of pre-existing nuclear localization signals in prokaryotic proteins

Version: 0 Date: 25 Feb 2020

Reviewer: Sergey Melnikov

Reviewer's report:

Please indicate how interesting you found the manuscript (whilst Biology Direct does not base publication decisions on interest levels, your feedback is useful for internal purposes):

An article of importance in its field

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?
6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to Biology Direct’s open peer review policy. I understand that my name will be included on my report to the authors and, if the authors elect to proceed to publication, my named report along with the authors' responses will be included within the published article. I agree for my
Please provide a brief overview of your review, stating plainly your opinion of the manuscript’s overall validity, significance and originality.

I reviewed this manuscript in detail when it was submitted to Molecular Biology and Evolution. I recommended the authors to make numerous changes, and they addressed every single of my comments. I therefore have no reason to criticize this work any further. This study is important to the field as it shows that the nuclear localization signals in modern eukaryotic proteins could simply emerge from DNA-/RNA-binding domains of cellular proteins, because having a DNA- or RNA-binding domain is frequently sufficient for a protein to be recognized as a nucleus resident. This is an important finding and I encourage you to publish this work as is.

In this concise and thought-provoking manuscript, Olga Lisitsyna et al. investigate a central evolutionary enigma: the origin of the cell nucleus. The authors convincingly show that, in most instances, all that a protein needs to enter the cell nucleus is a DNA-binding domain. For instance, in their experiments with prokaryotic proteins, they show that – even in the absence of predicted NLS sequences – some DNA-binding prokaryotic proteins are actively transported into the cell nucleus (Fig. 1). This experiment, along with their analysis of NLS overlaps with DNA-binding domains in protein structures, suggests that NLSs have initially evolved from (and within) DNA-binding domains of chromatin-binding proteins – the conclusion that makes the perfect sense from the point of evolutionary contingency. Furthermore, in their supplementary data, the authors have collected a wonderful review of the experimentally identified and predicted nuclear localization signals. This information alone will be very useful for other scientists working in the field of the origin of eukaryotes and origin of the nucleus.

Please detail any minor comments for the authors attention (spelling, typographical errors, grammatical errors, stylistic suggestions etc.) so that, once addressed, the authors may remove them from the review.
My only suggestion to the authors is to divide their data set of NLSs into two groups – experimentally-defined vs in silico predicted: when they describe their statistics on the % of NLSs overlap with RNA/DNA-binding domains, it seems useful to me to provide it first for the experimentally-defined NLSs (as a more reliable data), and then complement these numbers with additional data for in silico-identified NLSs.

**Reviewer confidential comments to Editor**

Please use this only for comments that relate to ethical or policy issues. Do not use it to repeat all or part of the comments in your review for the authors. These comments will not be included in the report passed to the authors.

Thank you for letting me review this work and for waiting for an extra day to receive my review. I like this short manuscript a lot and I think it fully deserves seeing the light of day in your magazine.