Cite as: C. Wilson, D. Kern, *Science* 10.1126/science.abd0364 (2020).

Response to Comment on "Ancient origins of allosteric activation in a Ser-Thr kinase"

Christopher Wilson and Dorothee Kern*

Department of Biochemistry and Howard Hughes Medical Institute, Brandeis University, Waltham, MA 02453, USA.

*Corresponding author. Email: dkern@brandeis.edu

Park *et al.* question one out of seven findings from Hadzipasic *et al.*: whether TPX2 allosterically regulates the oldest Aurora. We had already addressed the two concerns raised—sparse sequence sampling and not forcing the gene to the species tree—before publication. Moreover, we believe their ancestral sequence reconstruction would be consistent with a nonallosteric common ancestor, and we show large sequence differences caused by species tree–enforced gene trees.

The key findings in Hadzipasic *et al.* (*1*) are that (i) autophosphorylation is the ancient allosteric regulation for Aurora kinases; (ii) a gradual increase in allosteric activation took place during the holozoan evolution; (iii) an allosteric network in Aurora exists that, when mutated, alters allosteric activity; (iv) allosteric activation by TPX2 is entirely encoded in the kinase; (v) the interface between Aurora and TPX2 is co-conserved; (vi) evolution of specificity in signaling happens on binding affinity; and (vii) the oldest ancestral Aurora is not allosterically activated by TPX2.

The comment by Park *et al.* (*2*) questions only the seventh finding, on the basis of differences in the computation of ancestral sequence reconstruction (ASR). Notably, even though the ASR calculations differ, we believe the outcome is consistent with, rather than contradicting, the finding. The two concerns raised are (i) the small number of modern sequences used in the ASR calculations and (ii) the mismatch between the gene tree and the species tree. We had these same concerns and extensively investigated both points before publication:

1. The reason for the original sparse sampling was the use of BALi-Phy (*3*) to coestimate phylogeny and alignment using Bayesian statistics, a commonly used method in the field. After the completion of several years of experiments, we tested our concerns about the smaller number of modern sequences used by repeating the computational part of ASR with more than 500 sequences using a maximum likelihood (ML) approach, very similar to (*2*). By then, newer software was available, and the field, including the Thornton lab, had demonstrated that ML trees are in several cases comparable in quality to Bayesian trees. Therefore, we created a ML tree with more than 500 collected sequences from Uniport with an identity cutoff of 99%. We aligned the obtained sequences with PRANK (*4*). We used IQ-TREE (*5*) to simultaneously search for a best-fit model and constructed a ML tree with rapid bootstrapping and SH-aLRT to estimate node supports. Notably, the resulting tree contained the same general topology as in our Bayes-based tree (Fig. 1) with robust support at the relevant nodes. We did not include this additional analysis in the supplementary materials because of the agreement with the original tree and ancestors and because all experiments were performed on the original ancestors. We shared this information, including Fig. 1, with Park *et al.* before they submitted their Comment to *Science*.

2. Park *et al.*'s concern that the gene tree does not match the species tree is well known and controversially discussed in the field. We cite and state (*1*): "Discrepancies between gene trees and species trees are common…" (*6*–*9*). We also had tried to force the gene tree to the species tree. Unlike other cases where S-aware trees have been used to produce ancestors, the difference between the gene and species tree is large for Aurora; consequently, we found that forcing the gene tree onto the species tree resulted in large penalties. The best-fitting tree still differed significantly from the species tree and resembled the maximum a posteriori tree. Fungi were placed distant from the animals because the gene sequences differ more from animals than the plant sequences do. We disagree with a general approach of forcing gene trees to species trees, as done in figure 2 of (*2*).

Park *et al.*'s calculations alone highlight an issue with forcing the gene tree to the species tree: The sequences for their ancestor 1, which is the one questioned by the authors (AncEukarya), are very different between their ML tree and an ML tree that is forced onto the species tree (Fig. 2). They differ by 25% (yellow), significantly more than our nonallosteric Anc2 and allosteric Anc3 (10%). This comparison be-

tween the ASR from their ML and maximum congruence constraint (MCC) tree challenges their logic of restraining the gene tree topology to match a species tree for the Aurora system and their conclusion that both approaches give very similar results.

The question is whether the differences in ASR calculations matter for the outcome. We believe Park *et al.*'s ASR calculations would be consistent with our conclusion that TPX2 does not allosterically activate the oldest ancestor. These are the essential points for interpreting computational results without experimentally measuring the properties of those ancestors:

1. Their oldest ancestor 1, which the authors speculate will be allosterically regulated by TPX2 binding, is very likely not regulated by TPX2: In Fig. 2, AncAur_Eukarya_ML (F) has only 6 of the 15 required residues in the allosteric network (dark green); all others either have the incorrect residue (8 residues, yellow and gray) or too low posterior probabilities (1 residue, light green). We had experimentally determined that the majority of these 15 residues are necessary for allosteric TPX2 activation [figures 4 and S11 of (*1*)]. The "nonallosteric state" and "other state" [figure 2 of (*2*)] are all "nonallosteric" because they have different amino acids than the correct one. From our experimental data (*1*), an Aurora ancestor with a subset of only six correct residues in this allosteric network would not show allosteric activation.

2. Their ancestor 2 agrees with our findings that this node has all the allosteric network residues because this is the last common ancestor for Holozoa [ancestor 2 in figure 2F of (*2*)]. Their ancestor 3 node has no experimental counterpart in our study.

3. Park's ancestors 4 and 5 are in the Fungi kingdom, for which no allosteric activation is expected because fungal AurA are not activated by TPX2 binding (they lost the AurA binding motif of TPX2). Therefore, the loss of TPX2 activation in the Fungi kingdom is another point on which we agree with Park *et al.*, and we interpret this as the reason why fungi diverge further in the Aur gene tree.

In summary, we believe the outcome of Park *et al.*'s calculations is consistent with our key findings that the oldest regulation is autophosphorylation, and allosteric regulation by TPX2 was dialed in (*1*). The significant differences in overall sequence and critical positions in their ancestor sequences derived from a gene ML tree and a gene ML tree with congruence constraint to the species tree (Fig. 2) would, in our view, argue against constraining the gene tree to the species tree in this system. However, this question is under active investigation in the field and can only be answered by comparing both computational approaches in combination with experimental interrogation of the derived ancestral sequences. The Comment authors have convincingly performed such experimental tests for their systems (*10*–*13*). Such tests would be informative for the Aur ancestors suggested in Park *et al.* Allosteric enzyme activation is a difficult task to resurrect compared to only ligand binding. Apart from this methodological disagreement, the experimental results in (*1*) stand independent on the detailed phylogenetic tree: allosteric activation by either autophosphorylation or TPX2 binding, gradual dialing in of that second activation along the evolutionary trajectory in holozoans, identification of the allosteric network spanning a large portion of the kinase, co-conservation of the interface between Aurora and TPX2, and evolution of selectivity in activation by the correct activator via binding affinity.

REFERENCES

1. A. Hadzipasic, C. Wilson, V. Nguyen, N. Kern, C. Kim, W. Pitsawong, J. Villali, Y. Zheng, D. Kern, Ancient origins of allosteric activation in a Ser-Thr kinase. *Science* **367**, 912–917 (2020). doi:10.1126/science.aay9959 Medline
2. Y. Park, J. E. J. Patton, G. K. A. Hochberg, J. W. Thornton, Comment on "Ancient origins of allosteric activation in a Ser-Thr kinase". *Science* **370**, eabc8301 (2020).
3. M. A. Suchard, B. D. Redelings, BAli-Phy: Simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**, 2047–2048 (2006). doi:10.1093/bioinformatics/btl175 Medline
4. A. Löytynoja, Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014). Medline
5. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi:10.1093/molbev/msu300 Medline
6. R. Nichols, Gene trees and species trees are not the same. *Trends Ecol. Evol.* **16**, 358–364 (2001). doi:10.1016/S0169-5347(01)02203-0 Medline
7. W. P. Maddison, Gene Trees in Species Trees. *Syst. Biol.* **46**, 523–536 (1997). doi:10.1093/sysbio/46.3.523
8. N. Takahata, Gene genealogy in three related populations: Consistency probability between gene and population trees. *Genetics* **122**, 957–966 (1989). Medline
9. P. Pamilo, M. Nei, Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583 (1988). Medline
10. J. W. Thornton, E. Need, D. Crews, Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling. *Science* **301**, 1714–1717 (2003). doi:10.1126/science.1086185 Medline
11. G. C. Finnigan, V. Hanson-Smith, T. H. Stevens, J. W. Thornton, Evolution of increased complexity in a molecular machine. *Nature* **481**, 360–364 (2012). doi:10.1038/nature10724 Medline
12. M. J. Harms, G. N. Eick, D. Goswami, J. K. Colucci, P. R. Griffin, E. A. Ortlund, J. W. Thornton, Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11475–11480 (2013). doi:10.1073/pnas.1303930110 Medline
13. A. N. McKeown, J. T. Bridgham, D. W. Anderson, M. N. Murphy, E. A. Ortlund, J. W. Thornton, Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **159**, 58–68 (2014). doi:10.1016/j.cell.2014.09.003 Medline

ACKNOWLEDGMENTS

D.K. is a Howard Hughes Medical Institute Investigator.

1 September 2020; accepted 5 November 2020
Published online 20 November 2020
10.1126/science.abd0364

Fig. 1. Improved sequence sampling results in reconstruction similar to that in (*1*). Aurora homologs' ML tree using more than 500 sequences, with ultrafast bootstraps (left number) and SH-aLRT statistic (right number) shown at the relevant nodes.

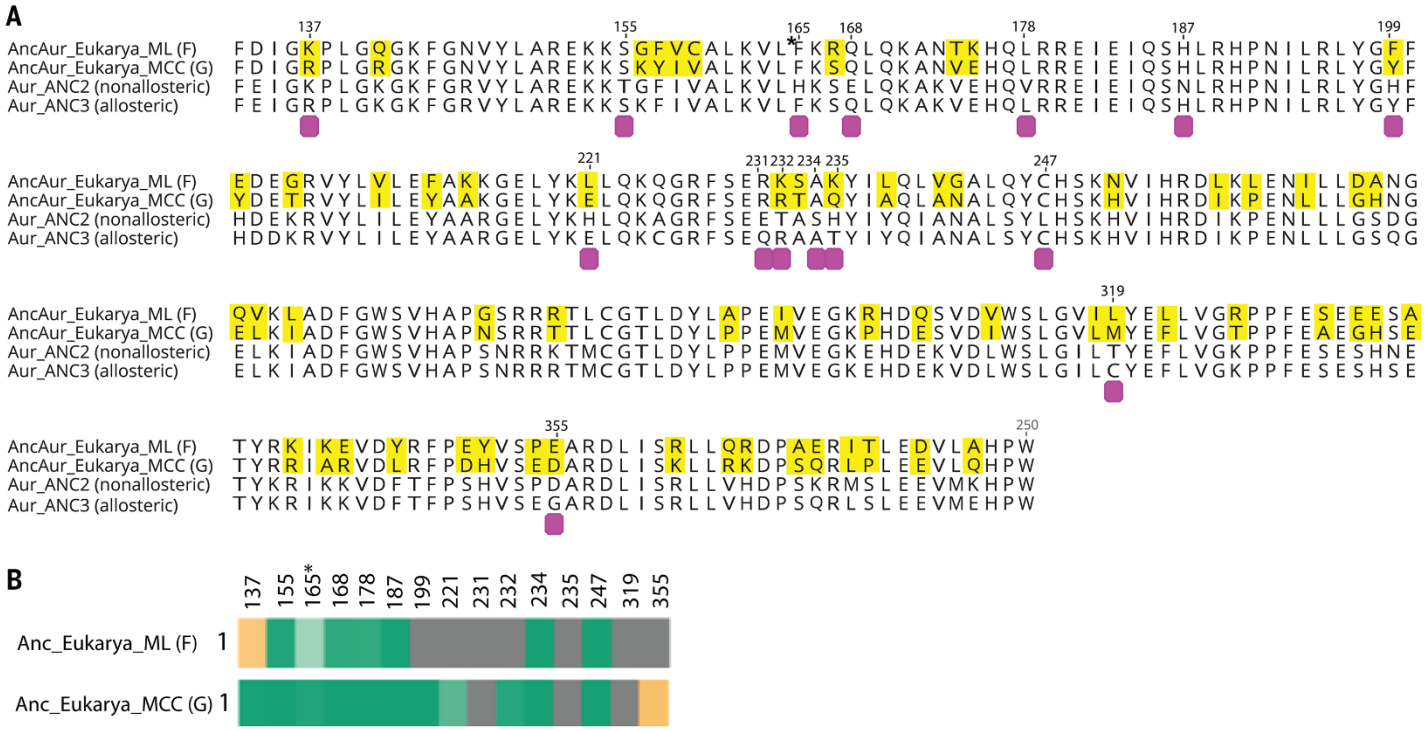

Fig. 2. Ancestral sequences from Park *et al.* argue against a species enforced gene tree and suggest agreement with our finding in (*1*) of incomplete allosteric network in oldest ancestor. (**A**) Comparison of the sequences of the oldest ancestor 1 in the ML tree [AncEukarya_ML (F)] and MCC tree [AncEukarya_MCC (G)] showing that the sequences differ by 25% when constraining the gene tree to the species tree (highlighted in yellow) (*2*), a finding that argues against a species enforced gene tree. Also shown are the two critical ancestors in (*1*), Aur_ANC2 and Aur_ANC3, which were shown experimentally to be nonallosteric and allosteric, respectively, in (*1*), differing by only 10%. Of the 15 positions of the allosteric network (indicated by pink symbols below), only six have the correct amino acid in the network in Anc_-Eukarya_ML (F). (**B**) The same 15 positions as in (A), represented as in figure 2, F and G, of (*2*) (posterior below 0.5 indicated by an asterisk). Accordingly, the oldest ancestor produced by the ML tree with expanded sampling is expected to be nonallosteric, consistent with our findings.

Response to Comment on "Ancient origins of allosteric activation in a Ser-Thr kinase"

Christopher Wilson and Dorothee Kern

Science **370** (6519), eabd0364.
DOI: 10.1126/science.abd0364